

Consensus Analysis of Signal Peptide Peptidase and Homologous Human Aspartic Proteases Reveals Opposite Topology of Catalytic Domains Compared with Presenilins*

Received for publication, July 13, 2004, and in revised form, September 20, 2004
Published, JBC Papers in Press, September 21, 2004, DOI 10.1074/jbc.M407898200

Elena Friedmann[‡], Marius K. Lemberg^{‡§}, Andreas Weihofen^{‡¶}, Kumlesh K. Dev^{||}, Uwe Dengler^{**},
Giorgio Rovelli^{||}, and Bruno Martoglio^{‡ ‡‡}

From the [‡]Institute of Biochemistry, Swiss Federal Institute of Technology (ETH), ETH-Hoenggerberg, 8093 Zurich and ^{||}Nervous System Research and ^{**}Functional Genomics, Novartis Pharma AG, Novartis Institutes for Biomedical Research, 4002 Basel, Switzerland

The human genome encodes seven intramembrane-cleaving GXGD aspartic proteases. These are the two presenilins that activate signaling molecules and are implicated in Alzheimer's disease, signal peptide peptidase (SPP), required for immune surveillance, and four SPP-like candidate proteases (SPPLs), of unknown function. Here we describe a comparative analysis of the topologies of SPP and its human homologues, SPPL2a, -2b, -2c, and -3. We demonstrate that their N-terminal extensions are located in the extracellular space and, except for SPPL3, are modified with *N*-glycans. Whereas SPPL2a, -2b, and -2c contain a signal sequence, SPP and SPPL3 contain a type I signal anchor sequence for initiation of protein translocation and membrane insertion. The hydrophilic loops joining the transmembrane regions, which contain the catalytic residues, are facing the exoplasm. The C termini of all these proteins are exposed toward the cytosol. Taken together, our study demonstrates that SPP and its homologues are all of the same principal structure with a catalytic domain embedded in the membrane in opposite orientation to that of presenilins. Other than presenilins, SPPL2a, -2b, -2c, and -3 are therefore predicted to cleave type II-oriented substrate peptides like the prototypic protease SPP.

Intramembrane-cleaving proteases cut proteins in transmembrane regions. They liberate fragments from dormant, membrane-bound precursor proteins, which typically exert downstream functions such as cell signaling and regulation (1, 2), immune surveillance (3), and intercellular communication (4, 5). They are also thought to contribute to the development and propagation of pathological conditions such as Alzheimer's disease (6) and hepatitis C virus infection (7). Three families of proteases that promote intramembrane cleavage are known at present. These are a group of metalloproteases represented by the human site-2 protease (S2P)¹ (8), a group of serine pro-

teases represented by *Drosophila melanogaster* rhomboid-1 (9), and a group of aspartic proteases with the human presenilins (PSs; PS1 and PS2) (10) and signal peptide peptidase (SPP) (11) as prototypic members.

PSs and SPP belong to the family of GXGD aspartic proteases (12, 13). Although the sequence homology is limited, sequences of PSs and SPP can be aligned almost throughout the entire length (14). They are membrane proteins with multiple predicted transmembrane regions, two of which contain the active site motifs YD and GXGD, a unique trait specific for this protease family. These structural similarities make a case for a common catalytic mechanism. Indeed, a number of protease inhibitors, including aspartic protease transition state analogues, targeted PSs and SPP and inhibited both activities (15). Despite common features, there are major differences between PSs and SPP. PSs undergo endoproteolysis for activation and appear to act as proteolytic subunits of a multiprotein complex called γ -secretase (16–18). For SPP activity, in contrast, there is no indication for a requirement neither for endoproteolytic activation nor for additional components (11). One of the most striking differences, however, is the apparent opposite orientation of the catalytic domains within the plane of the membrane (11). Transmembrane regions containing the catalytic aspartates have opposite orientation in PSs when compared with corresponding transmembrane regions in SPP. Because PSs are known to catalyze intramembrane cleavage of several type I anchored membrane proteins (*e.g.* β -amyloid precursor protein and Notch-1) (19), whereas SPP cuts type II-oriented transmembrane peptides (*e.g.* signal sequences, hepatitis C virus core protein precursor), the orientation of the respective catalytic site seems predictive for the orientation of the relevant substrates.

Besides the two PSs and SPP, the human genome contains four additional candidate GXGD proteases, which have been recognized by data base searching (11, 14, 20). These proteins, referred to as PSH (for PS homologues) (14), IMPAS (for intramembrane proteases) (20), or SPPLs (for SPP-like proteases) (11), have no identified function. Like the PSs and SPP, they are proteins with multiple predicted transmembrane regions. SPP and SPPLs share high homology particularly in the C-terminal half of the molecules, which contain the catalytic motifs and the conserved sequence QPALLY (11, 14). Because of this structural homology, it is likely that these proteins have

* This work was supported by grants from the Centre of Neuroscience Zurich, the Swiss National Science Foundation, and the National Competence Centre for Research on Neuronal Plasticity and Repair (to B. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[§] Present address: MRC Laboratory of Molecular Biology, Hills Rd., Cambridge CB2 2QH, UK.

[¶] Present address: Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115.

^{‡‡} To whom correspondence should be addressed. Tel.: 41-1-6326347; Fax: 41-1-632-1269; E-mail: bruno.martoglio@bc.biol.ethz.ch.

¹ The abbreviations used are: S2P, site-2 protease; ER, endoplasmic

reticulum; PBS, phosphate-buffered saline, PNGase F, *N*-glycosidase F; PS, presenilin; SPP, signal peptide peptidase; SPPL, SPP-like protease; TFPP, type IV prepilin peptidase; HA, hemagglutinin; Bicine, *N,N*-bis(2-hydroxyethyl)glycine.

the same enzymatic activity, namely that of an intramembrane-cleaving aspartic protease. The divergence observed in the N-terminal portions may be a determinant of their individual function.

In the present study we have investigated the topologies of SPP and the homologous human SPPLs. After cloning the respective cDNAs, we expressed the proteins in a cell-free *in vitro* system and in living cells, and we determined the topological function of the most N-terminal hydrophobic region and the localization of the N- and C-terminal portions. Furthermore, we determined the localization of the hydrophilic loops that connect the two transmembrane regions containing the catalytic aspartate residues. The topology of the latter is crucial for the future search for substrates of the SPPLs, because experimental data on PSs and SPP suggest that the topology of the catalytic domain correlates with the membrane orientation of the respective substrates.

EXPERIMENTAL PROCEDURES

Cloning SPPLs and Plasmid Constructs—Total RNA was isolated from SH-SY5Y and HEK293 cells using GlassMax kit (Invitrogen) and reverse-transcribed with Superscript II (Invitrogen). On the bases of available sequence information (for accession numbers see Fig. 1A), SPPL2a and SPPL3 cDNAs were amplified using primer pairs, 5'-TTAAGTCTAGCCAGTCCGATGGGGCCGC-3' (forward) and 5'-GCA-GCTCCCGATTCTGGAAGTATAATCTAGAGC-3' (reverse), and 5'-GCTCGAGATGGCGGAGCAGACCTACTCGTG-3' (forward) and 5'-G-CAGCTCCCGATTCTGGAAGTATAATCTAGAGC-3' (reverse), respectively. SPPL2b cDNA was amplified from a human dorsal root ganglion cDNA library (Invitrogen) and SPPL2c cDNA from a brain cDNA library (K. Kaupmann, Novartis) using the primer sets, 5'-GAATTCATGGCGGCAGCGGTGGCGGC-3' (forward) and 5'-CAGCCTGGCGCC-TCGGCCTAATCTAGAC-3' (reverse), and 5'-GGCTCGAGATGGCGTG-CCTGGGCTTCTCC-3' (forward) and 5'-GCAGCTCCCGATTCTGGAAGTATAATCTAGAGC-3' (reverse), respectively. The amplified products were cloned into pCR4-TOPO (Invitrogen) and subcloned into the mammalian expression vector pSV-Sport1 (Invitrogen). Mutations and HA tags were introduced by overlap extension PCR. All constructs were verified by sequencing.

Microarray Analysis—Human tissues, total RNA samples, and mRNA samples were obtained from Clonetics Biosciences, Inc., Clontech, AllCells, LLC, Clonetics/BioWhittaker, AMS Biotechnology, and the University of California, San Diego. The quality of all samples was determined with an Agilent Bioanalyzer. Microarray analysis was performed as described (21). In brief, 5 μ g of total RNA was used to synthesize cDNA that was then used as a template to generate biotinylated cRNA. cRNA was fragmented and added to Affymetrix HG-U133A chips and the custom human chip GNF1b from the Genomics Institute of the Novartis Research Foundation, according to the standard protocol outlined in the Gene Chip Expression Analysis Technical Manual of Affymetrix. After sample hybridization, microarrays were washed and scanned with an Agilent laser scanner. Affymetrix Microarray Suite version 5.0 (MAS5) was used to generate .CHP files. Expression values provided are the average from duplicate measures generated by taking two male and two female samples.

In Vitro Transcription, Translation, and Translocation—To prepare mRNA coding for "x+100" constructs, the respective coding region was amplified with PCR using pSV-Sport1 plasmids as templates, DynAzyme Polymerase (Finnzymes), SP6 primer, and a reverse primer, starting with 5'-N₆CTAN₂₀-3' to introduce a TAG stop codon at the desired position (22). To prepare template for SPPL3^{N₆} x+100, a forward primer containing the SP6 promoter and Kozak sequence 5'-CGTATTTAGGTGACACTATAGAATACCATG... followed by 9 bases coding for a consensus site for N-glycosylation (Asn-X-Ser) and 20 bases overlapping the template was used. PCR products were transcribed *in vitro* with SP6 RNA polymerase at 42 °C in the presence of 500 μ M m7G(5')ppp(5')G RNA CAP structure analogue (New England Biolabs). mRNAs were translated in 25 μ l of rabbit reticulocyte lysate (Promega) containing [³⁵S]methionine and [³⁵S]cysteine (Promix; Amersham Biosciences) and, where indicated, 2 eq of nuclease-treated rough microsomes prepared from dog pancreas and 30 μ M N-benzyloxy-L-thromethylamide to prevent N-glycosylation (23). Samples were incubated for 30 min at 30 °C. Microsomes were extracted with 500 mM KOAc and subjected to SDS-PAGE (22).

Expression in Tissue Culture Cells and Deglycosylation—HeLa cells

were plated at 50% confluence in 6-well plates in minimum Eagle's medium containing 10% fetal calf serum and transfected the following day using FuGENE (Roche Applied Science). In brief, 1 μ g of plasmid and 3 μ l of the reagent were used for each transfection, and cells were incubated for 9 h with medium containing DNA complexes. Thereafter, the cells were washed twice with PBS and lysed in PBS containing 0.5% SDS, 50 mM dithiothreitol, and Complete protease inhibitor mixture (Roche Applied Science). For deglycosylation, lysates were treated with N-glycosidase F (PNGase F, New England Biolabs) at 37 °C for 3 h according to the manufacturer's protocol but without denaturing.

Selective Permeabilization and Indirect Immunofluorescence—HeLa cells were plated on 12-mm coverslips in 6-well plates the day before transfection. 9 h after transfection, cells were washed twice with PBS, treated for 20 min with 2% paraformaldehyde at room temperature, and washed with PBS. Cells were next selectively permeabilized by incubating for 15 min at 4 °C in 5 μ g/ml digitonin, 0.3 M sucrose, 0.1 M KCl, 2.5 mM MgCl₂, 1 mM EDTA, and 10 mM HEPES-KOH, pH 6.8. Thereafter, cells were washed three times with PBS and thereafter incubated with blocking solution (4% fetal calf serum in PBS) at room temperature for 30 min. Control cells were treated the same way, but 0.2% saponin was added to the blocking solution. For immunofluorescence, coverslips were washed three times with PBS and incubated with primary mouse monoclonal antibody against the HA epitope (YPYDVPDYA; Covance) and rabbit polyclonal antibody against ERP57 or the cytosolic tail of calnexin (A. Helenius, ETHZ) in blocking solution for 1 h at room temperature. After washing three times with PBS, the cells were incubated for 1 h at room temperature with secondary antibodies (goat anti-rabbit IgG Alexa Fluor488 and goat anti-mouse IgG Alexa Fluor594, Molecular Probes) in blocking solution. After washing three times with PBS, coverslips were rinsed with water and mounted with IMMUMOUNT (Thermo-Shandon). Samples were visualized using a Zeiss Axiovert 100M fluorescence microscope.

SDS-PAGE and Western Blotting—Proteins were analyzed by SDS-PAGE using Tris-glycine acrylamide gels (12% T, 2.7% C for Figs. 2B, 3, and 4, SPP and SPPL3; 10% T, 2.7% C for Fig. 4B, SPPL2a, -2b) and a Tris-Bicine urea acrylamide gel (6% T, 3–6% linear gradient C for Fig. 4B, SPPL2c) (22). ³⁵S-Labeled proteins were visualized by a Phosphor-Imager (STORM 860, Amersham Biosciences). For Western blot analysis, proteins were separated by SDS-PAGE and transferred to polyvinylidene difluoride membranes, which were blocked with 5% milk powder in 20 mM Tris-Cl, pH 7.6, 140 mM NaCl, 0.1% Tween 20. Membranes were probed with monoclonal anti-HA antibody (Covance), and bound antibody was detected by enhanced chemiluminescence (Amersham Biosciences).

RESULTS

Defining Subfamilies and Sequence Analysis of SPP and SPPLs—SPP, the presenilins PS1 and PS2, and the bacterial type IV prepilin peptidases (TFPPs) are members of the GXGD family of aspartic proteases (12, 13). Searching the data bank revealed a number of additional candidate GXGD proteases present in the genome of organisms of all kingdoms; many are hypothetical proteins with unassigned function. Phylogenetic tree analysis clusters these sequences into different subfamilies (Fig. 1A). SPP and its human homologues appear to be members of three different subfamilies, one containing SPP, one containing SPPL3, and a third family containing three closely related members, SPPL2a, -2b, and -2c. These proteins seem only distantly related to the PSs and TFPPs, suggesting that SPP and its homologues may have structural features and substrate specificities that distinguish them from the PSs.

In order to experimentally analyze the relationships of SPP and SPPLs, we cloned the SPPL cDNAs from human cell lines and confirmed the sequences already deposited in the data base (see accession numbers Fig. 1A). By using microarray analysis, we then determined the appearance of SPP and SPPL mRNAs in different human tissues. The resulting profiles indicated the presence of mRNAs coding for SPP, SPPL2a, -2c, and SPPL3 in all major human adult tissues implying a wide distribution of the respective proteins (Fig. 1B). For SPPL2b mRNA, the resulting profile was more restrictive suggesting that SPPL2b is more selectively expressed.

We next investigated the sequences of the human proteins

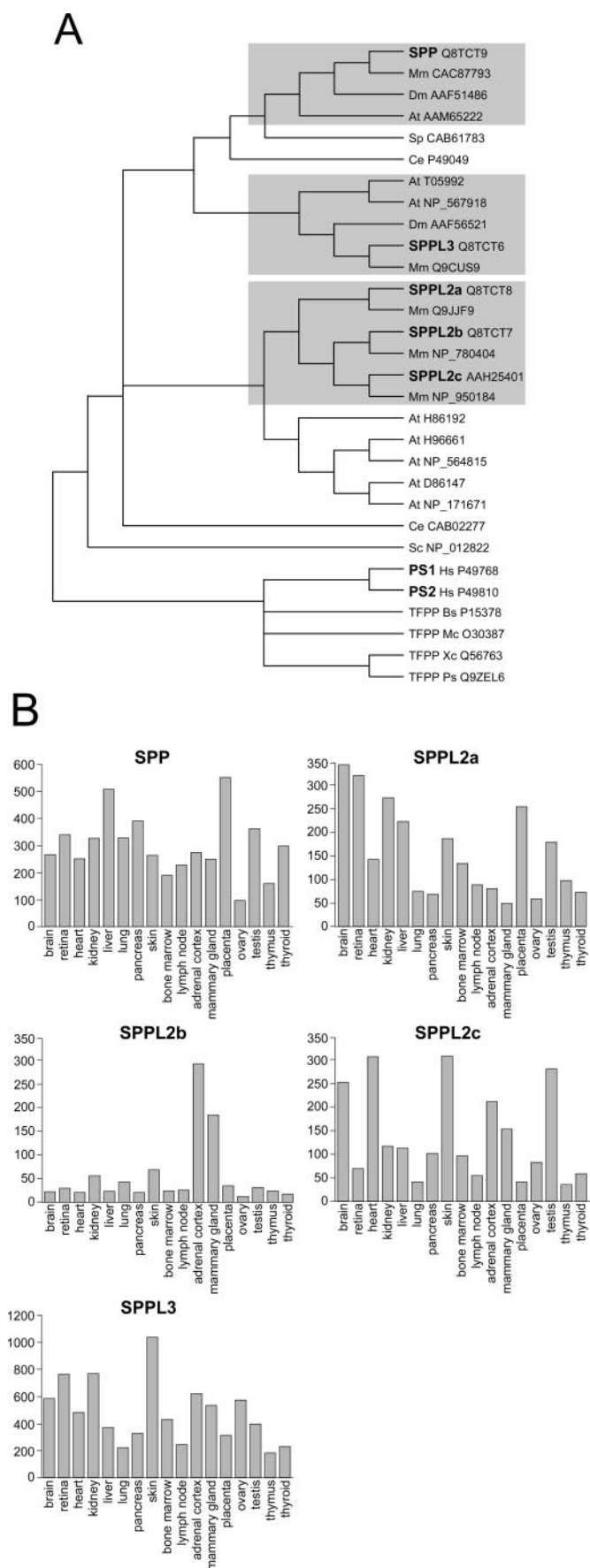


FIG. 1. **Relation of SPP and SPPLs.** A, phylogenetic tree of GXGD aspartic proteases. Sequences of indicated proteins were analyzed by the MacVector™ version 7.2 software (method, UPGMA; bootstrap; tie breaking = Systematic (1000 reps); distance, uncorrected ("p"); gaps

TABLE I

Consensus alignment of predicted transmembrane regions

Sequences of human SPP, SPPL3, SPPL2a, -2b, and -2c were analyzed by the sequence analysis programs SignalP version 2.0 (www.cbs.dtu.dk/services/SignalP-2.0/), TMHMM version 2.0 (www.cbs.dtu.dk/services/TMHMM-2.0/), HMMTOP version 2.0 (www.enzim.hu/hmmtop/index.html), PSORT II (psort.nibb.ac.jp/form2.html), and TMpred (www.ch.embnet.org/software/TMPRED_form.html). Predicted signal sequences (Sig) and transmembrane regions (I–X) are indicated by the position of the respective flanking amino acid residues. Putative transmembrane regions containing the catalytic motifs YD (YD in SPPL2c) and GXGD are highlighted in gray.

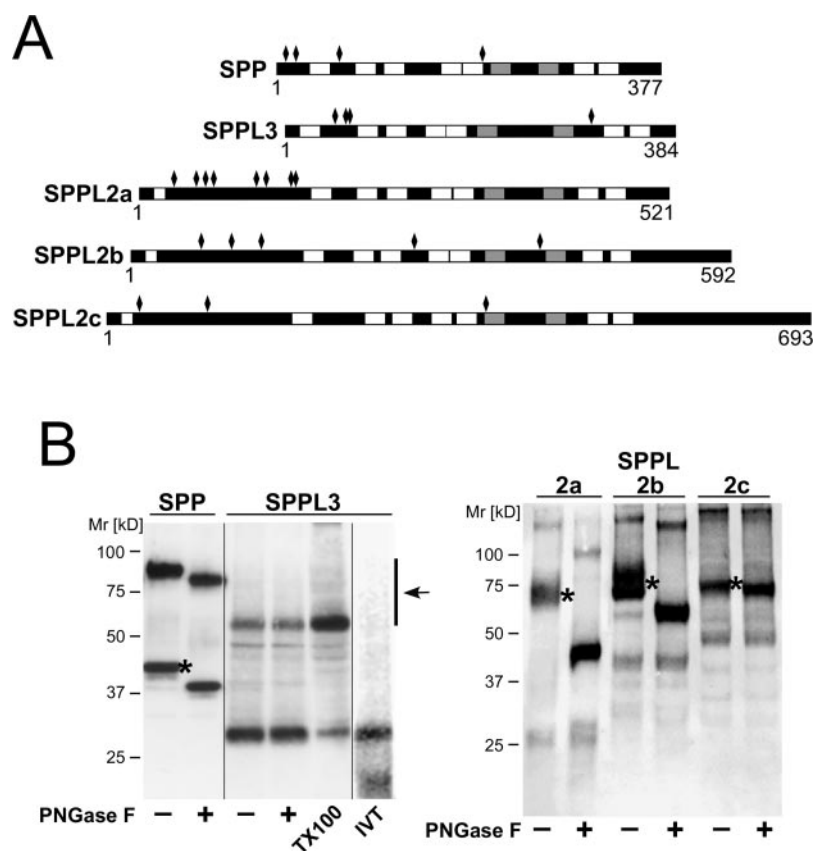
	Sig	I	II	III	IV	V	VI (YD ¹)	VII (GXGD)	VIII	IX	X
SPP											
SignalP	no										
TMHMM	–	32–54	75–97	107–129	–	–	210–232	258–277	290–312	317–336	–
HMMTOP	–	32–51	78–96	105–124	162–178	–	209–231	258–277	290–309	318–336	–
PSORT	–	32–48	80–96	101–117	162–178	–	208–224	261–277	295–311	319–335	–
TMpred	–	34–52	76–93	103–125	160–178	182–202	202–225	259–277	290–309	317–336	–
SPPL3											
SignalP	no										
TMHMM	–	15–37	73–90	94–116	137–159	163–181	188–210	261–283	312–334	338–360	–
HMMTOP	–	15–33	73–91	96–115	140–159	164–181	188–206	265–283	314–333	342–360	–
PSORT	–	17–33	74–90	99–115	144–160	165–181	190–206	267–283	317–333	342–358	–
TMpred	–	12–33	73–91	100–118	144–164	–	191–211	264–283	314–332	342–360	–
SPPL2a											
SignalP	1–25										
TMHMM	–	170–192	218–240	244–266	287–306	311–328	340–362	400–422	435–457	462–481	–
HMMTOP	–	171–193	216–237	244–268	289–306	313–329	340–361	392–414	435–457	464–481	–
PSORT	–	171–187	221–237	244–260	289–305	313–329	341–357	392–408	442–458	464–480	–
TMpred	10–30	170–188	223–241	243–260	289–306	314–334	342–362	400–423	438–458	463–479	–
SPPL2b											
SignalP	1–25										
TMHMM	12–34	173–195	222–244	249–271	298–315	320–337	350–369	414–431	444–466	–	–
HMMTOP	7–24	175–192	223–240	247–266	297–314	321–337	350–369	414–431	444–466	473–491	–
PSORT	–	175–191	224–240	246–262	298–314	321–337	349–365	414–430	451–467	474–490	–
TMpred	1–27	175–191	224–244	246–268	297–314	322–342	351–370	414–432	447–467	473–491	495–514
SPPL2c											
SignalP	1–21										
TMHMM	5–27	185–207	251–273	280–302	317–339	–	375–397	436–458	471–493	–	–
HMMTOP	6–30	188–205	255–273	280–304	317–340	347–364	377–396	441–458	471–494	501–518	–
PSORT	–	187–203	257–273	277–293	324–340	–	377–393	441–457	478–494	500–516	625–641
TMpred	2–19	187–203	256–274	277–296	320–339	349–365	378–397	441–458	474–497	500–518	–

SPP, SPPL2a, -2b, and -2c, and SPPL3 by the sequence analysis programs SignalP version 2.0 (www.cbs.dtu.dk/services/SignalP-2.0/), TMHMM version 2.0 (www.cbs.dtu.dk/services/TMHMM-2.0/), HMMTOP version 2.0 (www.enzim.hu/hmmtop/index.html), PSORT II (psort.nibb.ac.jp/form2.html), and TMpred (www.ch.embnet.org/software/TMPRED_form.html) (Table I). The resulting consensus of membrane protein topology (24) of SPP and SPPLs predicted a common arrangement of hydrophobic stretches, nine predicted transmembrane regions two of which contain the conserved catalytic motifs, YD and GXGD (Table I and Fig. 2A). Furthermore, the sequences of SPPL2a, -2b, -2c, and 3 contain close to their N terminus an additional hydrophobic region for which prediction programs propose the function of an endoplasmic reticulum (ER) targeting signal (25). In all three proteins, an extended hydrophilic region containing consensus sites for N-glycosylation follows the putative signal sequence (Fig. 2A). In contrast, neither an N-terminal signal sequence nor an extended hydrophilic stretch in the N-terminal part of the protein is predicted for SPP and SPPL3 (Table I and Fig. 2A).

Expression of SPP and SPPLs in Tissue Culture Cells—To

distributed proportionally). Proteins present in the human genome are highlighted in **boldface**. *Hs*, *Homo sapiens*; *Mm*, *Mus musculus*; *Dm*, *D. melanogaster*; *At*, *Arabidopsis thaliana*; *Sp*, *Schizosaccharomyces pombe*; *Ce*, *Caenorhabditis elegans*; *Sc*, *Saccharomyces cerevisiae*; *Bs*, *Bacillus subtilis*; *Ps*, *Pseudomonas stutzeri*; *Xc*, *Xanthomonas campestris*; and *Mc*, *Myxococcus xanthus*. *B*, expression profiles of mRNA coding for SPP and SPPLs. Total mRNA isolated from human tissue was probed by microarray analysis. Values are expressed as the average \pm S.D. from duplicate measures generated by taking two male and two female tissues per pool. Numbers in the y axis indicate relative levels of fluorescence intensity (arbitrary units).

FIG. 2. Expression of SPP and SPPLs in tissue culture cells. **A**, schematic illustration of the arrangement of transmembrane regions (white and gray bars) in SPP and SPPLs. Numbers refer to first and last residue of the respective protein; gray bars indicate transmembrane regions containing the catalytic motifs (see also Table I), and diamonds indicate consensus sites for *N*-glycosylation. **B**, HA-tagged proteins were expressed in HeLa cells and analyzed by SDS-PAGE and Western blotting using an HA-specific antibody. Where indicated, samples were treated with PNGase F to remove *N*-glycans. One sample of SPPL3-expressing cells was solubilized under mild conditions with Triton X-100 (TX100). *In vitro* translated (IVT) shows an autoradiography of ³⁵S-labeled SPPL3 produced by cell-free *in vitro* translation. Asterisks indicate products modified with *N*-glycans; the arrow indicates the region showing potentially dimeric forms of SPP and SPPL3, respectively.



experimentally address the topology of SPP and SPPLs, we first tested expression of cloned proteins in tissue culture cells. We transiently transfected HeLa cells with expression vectors encoding HA-tagged proteins under the control of the SV40 promoter. From the SPPL2a, -2b, and -2c, we expressed proteins containing a C-terminal HA tag, and from the SPP we expressed a construct containing an HA tag directly ahead of the C-terminal KKXX ER retrieval signal (26). Expression of proteins was analyzed by Western blotting. Because all investigated proteins contain multiple consensus sites for *N*-glycosylation (Fig. 2A), we also included treatment with *N*-glycosidase F (PNGase F). Sensitivity toward PNGase F is indicative of the acquisition of *N*-glycans and hence exposure of the respective region toward the exoplasm.

Expression of SPP resulted in two products, one of ~45 kDa corresponding to monomeric SPP (Fig. 2B, left panel, asterisk) and one of ~90 kDa. The latter product most likely represents the dimeric form of SPP which, according to a recent report by Nyborg *et al.* (27), shows up on SDS-PAGE upon mild solubilization in sample buffer. Both products were sensitive to treatment with PNGase F, which reduced the apparent size of the products to ~38 and ~80 kDa, respectively. This shift is consistent with the removal of two *N*-glycan moieties from the N-terminal domain. Similar results were obtained for SPPL2a, -2b, and -2c. For these three proteins we obtained products of about 75 kDa (Fig. 2B, right panel, asterisks), which were all sensitive to treatment with PNGase F (for the analysis of SPPL2c samples, see also the optimized gel system shown in Fig. 4B). The observed reductions in size are in agreement with the removal of the expected number of oligosaccharides from the N-terminal domain, up to 8 for SPPL2a, 3 for SPPL2b, and 1 or 2 for SPPL2c (the most N-terminal one is very close to the N terminus where glycosylation is typically inefficient; see *e.g.* Fig. 3B, SPPL3^{N_g}, *x*+100).

Expression of SPPL3 did not yield a glycoprotein as revealed

by the lack of sensitivity toward PNGase F (Fig. 2B, left panel). The product showed unexpected high electrophoretic mobility (~30 kDa) but nevertheless corresponded to the full-length protein as demonstrated by comparison with *in vitro* translated SPPL3 (Fig. 2B, left panel, IVT). Also for SPPL3 a potentially dimeric form migrating at ~60 kDa was observed, particularly when cells were lysed under mild conditions (Triton X-100; no heating; Fig. 2B, left panel, TX100). Taken together, these results indicate that the N-terminal hydrophilic domains of SPP and SPPLs are exposed toward the exoplasm and, with the exception of the one of SPPL3, modified with *N*-glycans.

Topological Function of Most N-terminal Hydrophobic Regions—We next analyzed the localization of the mature N termini and the topology of the first hydrophobic regions in more detail by using a cell-free *in vitro* translation/translocation system. The most N-terminal hydrophobic region of a multi-spanning membrane protein usually mediates targeting to the ER membrane and induces integration into the lipid bilayer. Such a “first” hydrophobic region may function as sole signal sequence that is cleaved off from the pre-protein once integration of the polypeptide chain into the membrane has been initiated (25). The N terminus of the mature protein then becomes located outside the cell in the exoplasm. Alternatively, the first hydrophobic region may function as a signal anchor sequence that also promotes targeting and membrane insertion, but is not removed, and eventually functions as a membrane anchor. Signal anchor sequences can integrate into the lipid bilayer either in a type I orientation, *i.e.* the N-terminal portion is translocated through the bilayer and becomes exposed toward the exoplasm, whereas the C-terminal portion remains in the cytoplasm or in a type II orientation, *i.e.* the N terminus remains in the cytosol whereas the C-terminal portion becomes translocated (25).

For SPP, we have previously reported that its first hydrophobic region functions as a type I signal anchor sequence (11).

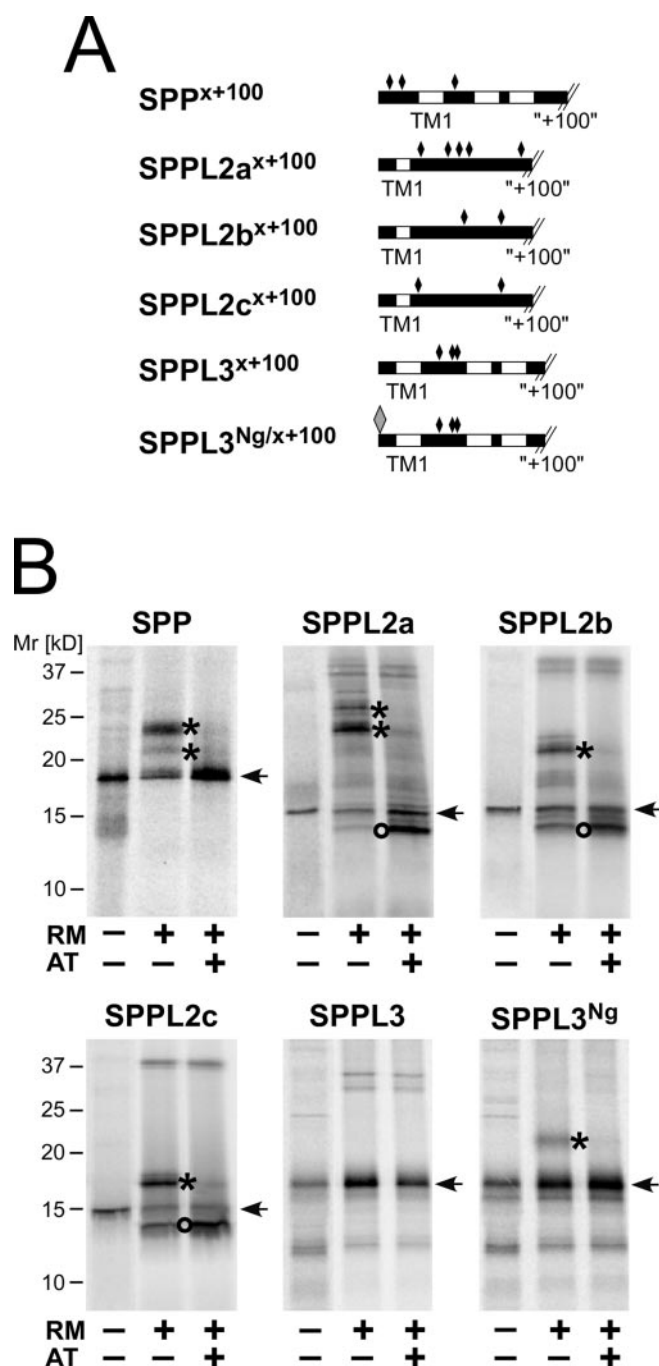


FIG. 3. Localization of N-terminal extensions investigated *in vitro*. *A*, schematic illustration of constructs used for the *in vitro* experiments. For each protein, the N-terminal portion comprising the hydrophilic N-terminal extension, the first hydrophobic region (see Table I) plus the downstream 100 residues (“+100”) were synthesized by cell-free *in vitro* translation. In SPP and SPPL3, the downstream region also contains transmembrane regions II and III. White bars indicate transmembrane regions; diamonds indicate consensus sites for *N*-glycosylation. SPPL3^{Ng} refers to an SPPL3 mutant containing an additional consensus site for *N*-glycosylation (gray diamond) at the N terminus. *B*, synthesis of *x*+100 constructs in a cell-free *in vitro* translation system in the presence of ER-derived rough microsomes (RM) for protein translocation and acceptor tripeptide (AT, *N*-benzoyl-Asn-Leu-Thr-methylamide) to inhibit *N*-glycosylation. Arrows indicate the respective *x*+100 translation product; asterisks indicate proteins modified with one or more *N*-glycans; circles indicate nonglycosylated products from which a signal sequence had been cleaved.

Translocation of the N-terminal hydrophilic extension of the SPP is indicated by the acquisition of *N*-glycans at the two consensus sites for *N*-glycosylation present in this region (Fig.

3, *A* and *B*, asterisks versus arrow). To test whether the first N-terminal regions of the SPPLs are similarly integrated into membranes as the one of SPP, we made use of the classic cell-free *in vitro* translation/translocation system comprising reticulocyte lysate for protein synthesis and ER-derived rough microsomes for membrane insertion (28). In the presence of [³⁵S]methionine, we translated mRNA of each SPPL coding for the N-terminal hydrophilic extension, the first hydrophobic region, plus the downstream 100 residues (Fig. 3*A*, *x*+100 constructs). Sequence analysis programs predict the function of a signal sequence for the first hydrophobic region of SPPL2a, -2b, and -2c, and consensus sites for *N*-glycosylation within the downstream 100 residues. Indeed, when the respective “*x*+100” constructs were translated in the *in vitro* assay, products of higher molecular weight appeared in the presence of microsomes when compared with samples where microsomes were not added (Fig. 3*B*, asterisks versus arrows). These higher molecular weight products were not produced when the acceptor tripeptide glycosylation inhibitor *N*-benzoyl-Asn-Leu-Thr-methylamide (29) was added to the reactions. The major products in these latter samples, however, were smaller than the ones produced in samples without microsomes, indicating the removal of a signal sequence (Fig. 3*B*, circles). These data confirm the prediction, namely that SPPL2a, -2b, and -2c contain an N-terminal signal sequence, and demonstrate that N-terminal domains of mature SPPL2s are glycosylated, and therefore facing the exoplasm.

Analysis of the sequence of SPPL3 predicts the function of a signal anchor sequence for the first hydrophobic region, like for SPP, and consensus sites for *N*-glycosylation within the downstream 100 residues but none in the N-terminal hydrophilic extension (Fig. 3*A*). When the SPPL3 *x*+100 construct was translated in the presence of microsomes, the translation product had equal electrophoretic mobility as the one produced in the absence of microsomes (Fig. 3*B*). Also in the presence of the glycosylation inhibitor, the apparent molecular weight was unchanged. This indicates that the first hydrophobic region of SPPL3 functions as a type I signal anchor sequence, and the downstream hydrophilic loop does not become glycosylated and hence should be facing the cytoplasm. To confirm the topology of this first transmembrane region, we inserted a consensus site for *N*-glycosylation into the short hydrophilic N-terminal extension (Fig. 3*A*, SPPL3^{Ng}). When translated in the presence of microsomes, SPPL3^{Ng} *x*+100 was glycosylated, albeit inefficiently, indicating translocation into the microsomes (Fig. 3*B*).

Topology of Catalytic Loops—In the next set of experiments, we wanted to determine the localization of the “catalytic loop,” *i.e.* the hydrophilic portion, which links the two transmembrane regions containing the catalytic aspartate residues (Table I; hydrophobic regions VI and VII). In SPP, this loop is predicted to be exposed toward the exoplasm, whereas in the related presenilins the respective regions are located in the cytosol (30–33). To test whether the catalytic loops of SPP and SPPLs are either exposed toward the exoplasm or the cytosol, we introduced a consensus site for *N*-glycosylation, Asn-X-Ser, into the hydrophilic loop region (Fig. 4*A*), and we expressed HA-tagged mutant proteins in HeLa cells. Because oligosaccharyltransferase can only transfer *N*-glycans to sites more than 12–14 residues away from transmembrane regions (34, 35), we generated an *N*-glycosylation site in the central part of the loop for SPPL3 (PL_{227/8}AS), which contains a larger loop. For the other proteins, which contain only a short catalytic loop, we inserted a glycosylation site plus a few additional residues to slightly expand the loop (for SPP, SGSGPAENASAHGAQSP after Phe²⁴⁴; for SPPL2a, NASEFRH after Val³⁸⁵; for SPPL2b, NASEFRH after Val³⁹⁶; and for SPPL2c, NAS-

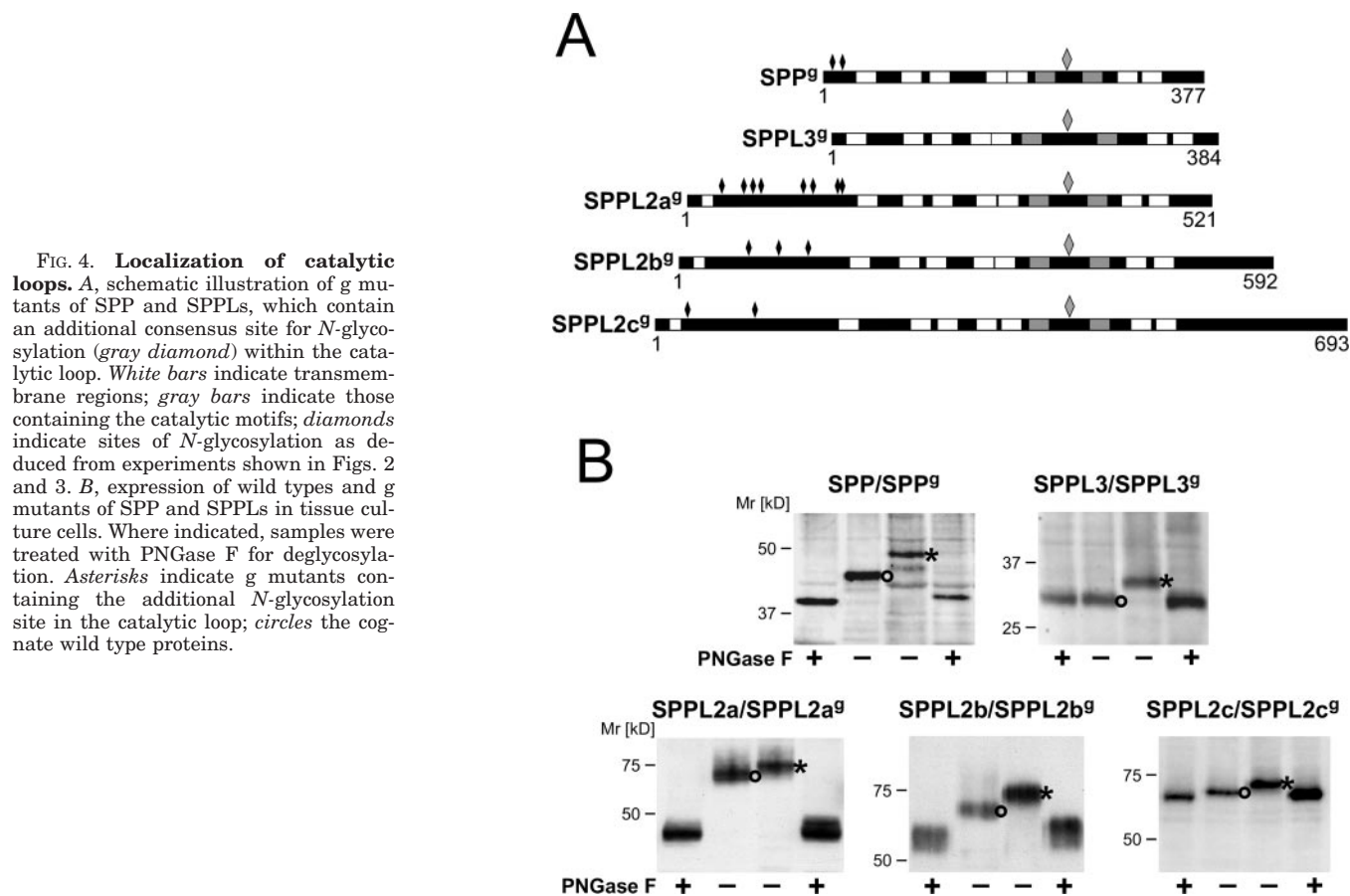


FIG. 4. Localization of catalytic loops. **A**, schematic illustration of *g* mutants of SPP and SPPLs, which contain an additional consensus site for *N*-glycosylation (gray diamond) within the catalytic loop. White bars indicate transmembrane regions; gray bars indicate those containing the catalytic motifs; diamonds indicate sites of *N*-glycosylation as deduced from experiments shown in Figs. 2 and 3. **B**, expression of wild types and *g* mutants of SPP and SPPLs in tissue culture cells. Where indicated, samples were treated with PNGase F for deglycosylation. Asterisks indicate *g* mutants containing the additional *N*-glycosylation site in the catalytic loop; circles the cognate wild type proteins.

EFRH after Ser⁴¹⁶). When expressed in HeLa cells, all the mutant proteins (“*g*” mutants) with the additional *N*-glycosylation site showed an increased electrophoretic mobility compared with the wild type protein counterparts. This increase is consistent with the acquisition of one additional *N*-glycan (Fig. 4B, asterisks versus circles). These results demonstrate that the catalytic loops of the *g* mutants of SPP and SPPLs are all facing the exoplasm.

Localization of C Termini—Finally, we determined the localization of the C termini of SPP and SPPLs by using selective permeabilization of cellular membranes with digitonin. This method makes use of differences in the cholesterol content of membranes and allows permeabilization of the plasma membrane, which is rich in cholesterol, whereas particularly ER and Golgi membranes, which contain little cholesterol, are unaffected (36). Therefore, in permeabilized cells, cytosolic domains of membrane proteins are accessible to antibodies for immunofluorescence, for example, whereas luminal domains are hidden. We expressed C-terminally HA-tagged SPP and SPPLs in HeLa cells, selectively permeabilized the cells with digitonin, and for the purpose of control, permeabilized all cellular membranes with saponin. Samples were analyzed by immunofluorescence with anti-HA antibodies as well as an antibody specific for the ER luminal protein ERP57 (37).

SPP contains a KKXX ER retrieval signal, which can only function when exposed toward the cytosol. Not surprisingly however, immunofluorescence after selective permeabilization of the plasma membrane revealed accessibility for the C-terminal extension of the SPP, whereas the ER luminal protein ERP57 was not stained (Fig. 5). The control ER luminal protein was protected by the intact ER membrane and only accessible after permeabilization of all membranes with saponin. Accordingly, the C terminus of SPP is, as expected, exposed toward the cytosol, like the cytosolic tail of the control protein calnexin

(38), from which we obtained the same staining pattern after selective permeabilization (Fig. 5). Similar results were obtained for the SPPLs. In all cases, the C-terminal HA tags were accessible after selective permeabilization of the plasma membrane and could be visualized by immunofluorescence (Fig. 5), whereas the control ER luminal protein ERP57 was protected by the intact ER membrane and only accessible after permeabilization with saponin (not shown). These results indicate that the C-terminal tails of SPP and SPPLs are all exposed in the cytosol. Taken together with the results described above, we conclude that SPP and SPPLs are candidate aspartic proteases sharing the same principal structure and, like SPP, are prone to promote intramembrane proteolysis of transmembrane regions, which span the membrane in type II orientation.

DISCUSSION

A good topology model is a necessary prerequisite for experimental studies on the structure-function relationship of membrane proteins. Here we have characterized the key structural elements of the intramembrane-cleaving protease SPP and its four human homologues SPPL2a, -2b, -2c, and -3. Starting with a consensus prediction of membrane protein topology, we have experimentally determined *in vitro* and in living cells the location of the N termini and topological role of most N-terminal hydrophobic regions, the orientation of catalytic domains, and the location of the C termini. We found that SPP and SPPLs share the same principal topology and appear to be designed to cleave a type II-oriented transmembrane region of substrate proteins. This is in contrast to the structurally related prenilins, whose catalytic domains have the opposite orientation within the membrane and, accordingly, are intended to cleave type I-oriented transmembrane regions. Our findings provide the elemental information for the search of candidate sub-

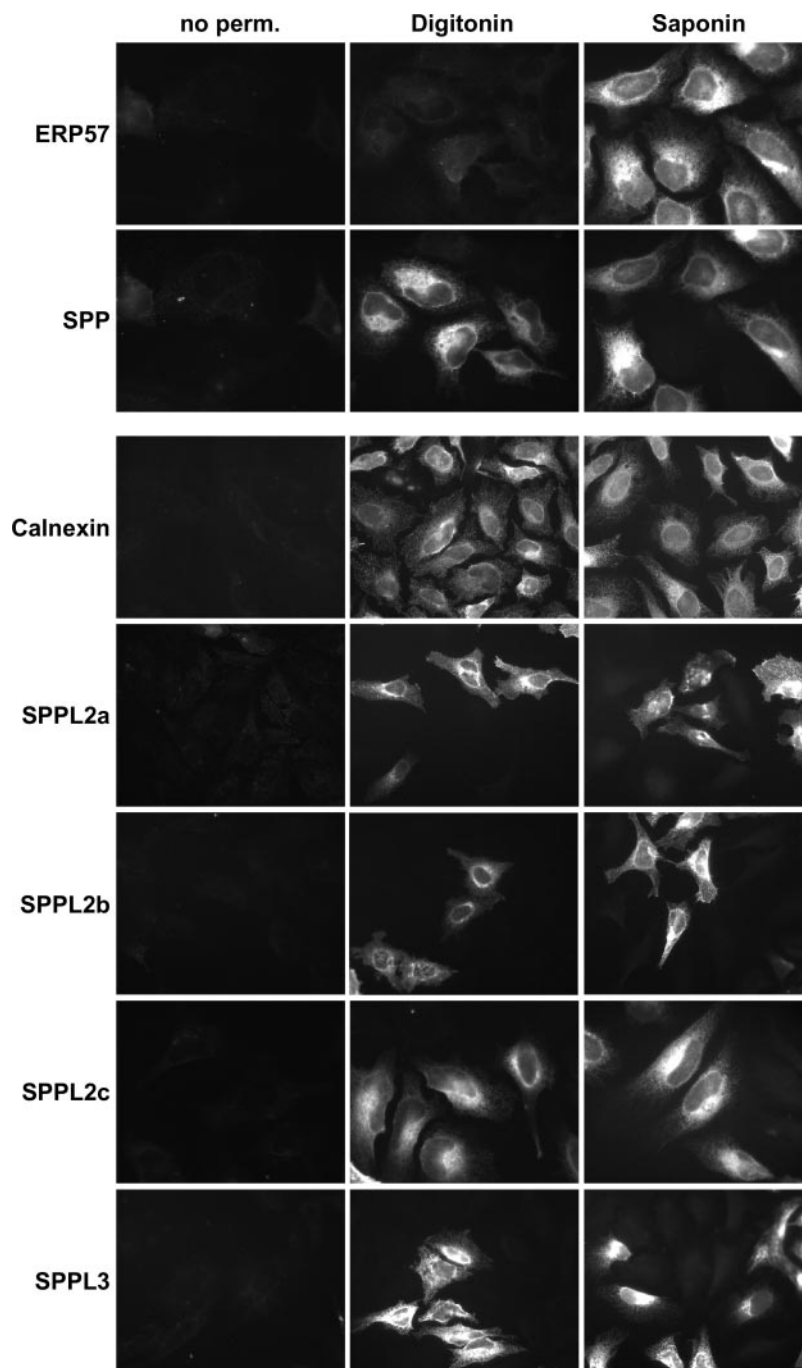


FIG. 5. Localization of C-terminal tails. Immunofluorescence after selective permeabilization (*perm*) of the plasma membrane. Cells expressing HA-tagged SPP and SPPLs, respectively, were permeabilized with either digitonin or saponin and probed with an HA-specific antibody as well as an ERP57-specific antibody (shown for experiment with SPP only). For controls, cells were permeabilized with either digitonin or saponin and probed with an antibody specific for the cytosolic tail of calnexin.

strates and future exploration of cellular functions of these membrane proteases.

During biosynthesis, a multispanning membrane protein is thought to acquire its correct topology by properly orienting the first transmembrane region and then alternate insertion of subsequent ones in sequentially opposite orientation (39, 40). Correct prediction and experimental analysis of the topology of the first transmembrane region are therefore crucial for the overall topology prediction of a multispanning membrane protein. Topology prediction programs use model formalisms to forecast the architecture of an integral membrane protein. To a variable degree, they all consider common features of integral membrane proteins, *e.g.* that transmembrane regions are generally hydrophobic, helical, and oriented in the lipid bilayer according to the “positive inside rule,” which is, however, followed less in eukaryotes (41). Best current topology prediction methods such as TMHMM and HMMTOP, which both use a

hidden Markov model formalism, predicted correct topology for up to 79% of all proteins, for which experimental structural information is available (42). The reliability of topology prediction can be significantly improved when several methods are combined (24). Consensus topology predictions of eukaryotic membrane proteins is, however, complicated by the fact that many of these proteins have N-terminal signal peptides, which can undermine the reliability of prediction. Errors in topology assignment of the most N-terminal transmembrane region can, for example, lead to an incorrectly predicted orientation for the subsequent transmembrane regions. Also, a hydrophobic core of a soluble domain may be misinterpreted as a transmembrane region, and hydrophobic segments either containing charged residues or comprising only a short region can be overlooked. Therefore, it is crucial to verify the predicted topology model by complementary experimental approaches.

Limits in the accuracy of topology predictions, as compared

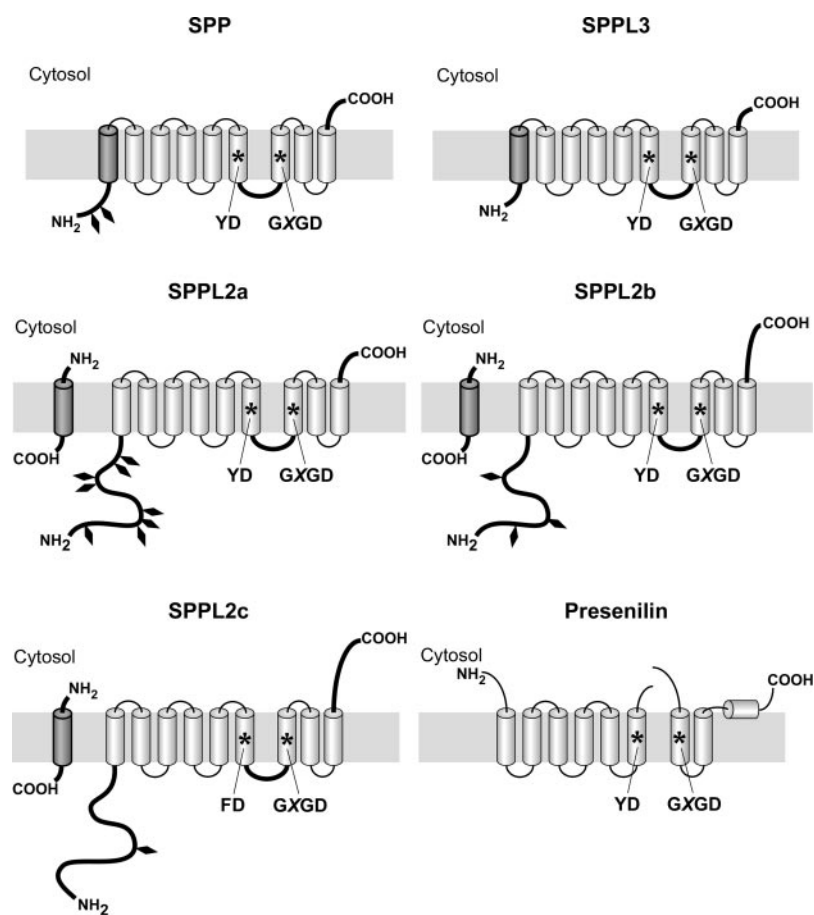


FIG. 6. **Topology model of SPP and SPPLs.** Gray barrels indicate predicted transmembrane regions. Black lines and dark gray barrels highlight regions experimentally investigated in this study. Diamonds show sites modified with *N*-glycans, and asterisks indicate the position of active site motifs. For comparison, a model of presenilin is also shown (47).

with data substantiated by experiments, became apparent in the present study. The programs TMMHMM and HMMTOP, for example, did not recognize the signal sequence of SPPL2a (Table I), but the “outside” localization for the large N-terminal domain was nevertheless correctly predicted. For SPPL2c, TMMHMM predicted type I orientation for the first hydrophobic region (the experimentally determined signal sequence), and as a consequence the orientation of the subsequent transmembrane regions was incorrectly assigned. Similarly, Tmpred predicted a type II signal anchor sequence at the N termini of SPP and SPPL3, whereas our experiments revealed a type I signal anchor sequence. As a result, the topology prediction for the downstream catalytic loops and the C termini were opposite the empirically determined location. On the other end, the topology of SPP and SPPLs was correctly assigned when the predictions were complemented with our experimental data; once the topology of the most N-terminal transmembrane regions and the location of the C termini were firmly established, prediction on the topology of the catalytic loops became consistent with the gathered data. Taken together, our results support a structural model where SPP and its related candidate proteases, SPPL2a, -2b, -2c, and -3, have nine transmembrane regions with the hydrophilic N-terminal domain and catalytic loop exposed toward the exoplasm, whereas the C terminus is located in the cytosol (Fig. 6).

Intramembrane-cleaving proteases contain multiple hydrophobic regions that are considered to assemble into a proteolytic domain within the plane of the membrane. We hypothesize that this peculiar type of proteases, like all membrane proteins, have a well defined orientation in the membrane and are likely to cleave transmembrane regions of only one given topology, *i.e.* either of type I or type II orientation but not both. Experimental evidence supporting this model can be found in

studies comparing the intramembrane-cleaving metalloproteases S2P (43) with its homologue SpoIVFB (44), and the aspartic proteases SPP (11) with the related PS1 (30–33). The hydrophobic regions containing the respective catalytic residues are predicted to have opposite transmembrane orientation in S2P and SPP, when compared with those of SpoIVFB and PS1, respectively. In accordance with the apparently opposite orientation of catalytic domains, the scissile transmembrane regions of known respective substrates have opposite orientation too; those cleaved by S2P and SPP have type II orientation, whereas those processed by SpoIVFB and PS1 have type I orientation. Because the topology of the catalytic domain of an intramembrane-cleaving protease seems to be the key in determining the transmembrane orientation of its substrates, our results predict that SPP2a, -2b, -2c, and -3, like SPP, cleave substrate proteins within a type II-oriented transmembrane region.

In summary, our experiments revealed structural features of SPP that are common to its four human homologues, SPPL2a, -2b, -2c, and -3. With information about their potential substrate preferences, we can now selectively search for candidate substrates of the SPPLs. These can be either type II membrane proteins or accessible type II-oriented transmembrane regions of multispinning membrane proteins (45). Candidate substrates of SPPLs may depend on similar requirements for intramembrane cleavage as the ones of SPP; shedding of a membrane-anchored ectodomain of the precursor may produce the immediate substrate for the intramembrane-cleaving protease, and helix-destabilizing residues within the transmembrane region may guarantee efficient processing (46). Based on such information and the data presented here, a systematic exploration of appropriate membrane proteins will most likely reveal a number of additional signaling molecules and regulatory factors that make use of intramembrane proteolysis for their activation.

Acknowledgments—We thank A. I. Su and J. R. Walker for microarray data; K. Kaupmann for the brain cDNA library; and A. Helenius for anti-calnexin and anti-ERP57 antibodies. We also thank D. Stauffer and M. Harrer for technical assistance.

REFERENCES

1. Brown, M. S., and Goldstein, J. L. (1997) *Cell* **89**, 331–340
2. Mumm, J. S., and Kopan, R. (2000) *Dev. Biol.* **228**, 151–165
3. Lemberg, M. K., Bland, F. A., Weihofen, A., Braud, V. M., and Martoglio, B. (2001) *J. Immunol.* **167**, 6441–6446
4. An, F. Y., Sulavik, M. C., and Clewell, D. B. (1999) *J. Bacteriol.* **181**, 5915–5921
5. Lee, J. R., Urban, S., Garvey, C. F., and Freeman, M. (2001) *Cell* **107**, 161–171
6. Selkoe, D. J. (2001) *Physiol. Rev.* **81**, 741–766
7. McLauchlan, J., Lemberg, M. K., Hope, G., and Martoglio, B. (2002) *EMBO J.* **21**, 3980–3988
8. Rawson, R. B., Zelenski, N. G., Nijhawan, D., Ye, J., Sakai, J., Hasan, M. T., Chang, T. Y., Brown, M. S., and Goldstein, J. L. (1997) *Mol. Cell* **1**, 47–57
9. Urban, S., Lee, J. R., and Freeman, M. (2001) *Cell* **107**, 173–182
10. Wolfe, M. S., Xia, W., Ostaszewski, B. L., Diehl, T. S., Kimberly, W. T., and Selkoe, D. J. (1999) *Nature* **398**, 513–517
11. Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K., and Martoglio, B. (2002) *Science* **296**, 2215–2218
12. Haass, C., and Steiner, H. (2002) *Trends Cell Biol.* **12**, 556–562
13. Martoglio, B., and Golde, T. E. (2003) *Hum. Mol. Genet.* **12**, (suppl. 2) R201–R206
14. Ponting, C. P., Hutton, M., Nyborg, A., Baker, M., Jansen, K., and Golde, T. E. (2002) *Hum. Mol. Genet.* **11**, 1037–1044
15. Weihofen, A., Lemberg, M. K., Friedmann, E., Rueeger, H., Schmitz, A., Paganetti, P., Rovelli, G., and Martoglio, B. (2003) *J. Biol. Chem.* **278**, 16528–16533
16. Edbauer, D., Winkler, E., Regula, J. T., Pesold, B., Steiner, H., and Haass, C. (2003) *Nat. Cell Biol.* **5**, 486–488
17. De Strooper, B. (2003) *Neuron* **38**, 9–12
18. Takasugi, N., Tomita, T., Hayashi, I., Tsuruoka, M., Niimura, M., Takahashi, Y., Thinakaran, G., and Iwatsubo, T. (2003) *Nature* **422**, 438–441
19. Kopan, R., and Ilagan, M. X. G. (2004) *Nat. Rev. Mol. Cell Biol.* **5**, 499–504
20. Grigorenko, A. P., Moliaka, Y. K., Korovaitseva, G. I., and Rogae, E. I. (2002) *Biochemistry (Moscow)* **67**, 826–835
21. Su, A. I., Cooke, M. P., Ching, K. A., Hakak, Y., Walker, J. R., Wiltshire, T., Orth, A. P., Vega, R. G., Sapinoso, L. M., Moqrich, A., Patapoutian, A., Hampton, G. M., Schultz, P. G., and Hogenesch, J. B. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4465–4470
22. Lemberg, M. K., and Martoglio, B. (2003) *Anal. Biochem.* **319**, 327–331
23. Martoglio, B., Hauser, S., and Dobberstein, B. (1998) in *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) Vol. 2, 2nd Ed., pp. 265–274, Academic Press, San Diego
24. Nilsson, J., Persson, B., and von Heijne, G. (2000) *FEBS Lett.* **486**, 267–269
25. Martoglio, B., and Dobberstein, B. (1998) *Trends Cell Biol.* **8**, 410–415
26. Townsley, F. M., and Pelham, H. R. (1994) *Eur. J. Cell Biol.* **64**, 211–216
27. Nyborg, A. C., Kornilova, A. Y., Jansen, K., Ladd, T. B., Wolfe, M. S., and Golde, T. E. (2004) *J. Biol. Chem.* **279**, 15153–15160
28. Blobel, G., and Dobberstein, B. (1975) *J. Cell Biol.* **67**, 852–862
29. Lau, J. T., Welply, J. K., Shenbagamurthi, P., Naider, F., and Lennarz, W. J. (1983) *J. Biol. Chem.* **258**, 15255–15260
30. Doan, A., Thinakaran, G., Borchelt, D. R., Slunt, H. H., Ratovitsky, T., Podlisky, M., Selkoe, D. J., Seeger, M., Gandy, S. E., Price, D. L., and Sisodia, S. S. (1996) *Neuron* **17**, 1023–1030
31. Li, X., and Greenwald, I. (1996) *Neuron* **17**, 1015–1021
32. Li, X., and Greenwald, I. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 7109–7114
33. Nakai, T., Yamasaki, A., Sakaguchi, M., Kosaka, K., Mihara, K., Amaya, Y., and Miura, S. (1999) *J. Biol. Chem.* **274**, 23647–23658
34. Nilsson, I. M., and von Heijne, G. (1993) *J. Biol. Chem.* **268**, 5798–5801
35. Popov, M., Tam, L. Y., Li, J., and Reithmeier, R. A. (1997) *J. Biol. Chem.* **272**, 18325–18332
36. Wilson, R., Allen, A. J., Oliver, J., Brookman, J. L., High, S., and Bulleid, N. J. (1995) *Biochem. J.* **307**, 679–687
37. Oliver, J. D., van der Wal, F. J., Bulleid, N. J., and High, S. (1997) *Science* **275**, 86–88
38. Wada, I., Rindress, D., Cameron, P. H., Ou, W. J., Doherty, J. J., II, Louvard, D., Bell, A. W., Dignard, D., Thomas, D. Y., and Bergeron, J. J. (1991) *J. Biol. Chem.* **266**, 19599–19610
39. High, S., Greenfield, J. J., Meacock, S. L., and Oliver, J. D. (1999) *Biochem. Soc. Trans.* **27**, 883–888
40. Ott, C. M., and Lingappa, V. R. (2002) *J. Cell Sci.* **115**, 2003–2009
41. von Heijne, G. (1992) *J. Mol. Biol.* **225**, 487–494
42. Kall, L., and Sonnhammer, E. L. (2002) *FEBS Lett.* **532**, 415–418
43. Zelenski, N. G., Rawson, R. B., Brown, M. S., and Goldstein, J. L. (1999) *J. Biol. Chem.* **274**, 21973–21980
44. Green, D. H., and Cutting, S. M. (2000) *J. Bacteriol.* **182**, 278–285
45. Lemberg, M. K., and Martoglio, B. (2004) *FEBS Lett.* **564**, 213–218
46. Lemberg, M. K., and Martoglio, B. (2002) *Mol. Cell* **10**, 735–744
47. Wolfe, M. S., and Kopan, R. (2004) *Science* **305**, 1119–1123

Consensus Analysis of Signal Peptide Peptidase and Homologous Human Aspartic Proteases Reveals Opposite Topology of Catalytic Domains Compared with Presenilins

Elena Friedmann, Marius K. Lemberg, Andreas Weihofen, Kumlesh K. Dev, Uwe Dengler, Giorgio Rovelli and Bruno Martoglio

J. Biol. Chem. 2004, 279:50790-50798.

doi: 10.1074/jbc.M407898200 originally published online September 21, 2004

Access the most updated version of this article at doi: [10.1074/jbc.M407898200](https://doi.org/10.1074/jbc.M407898200)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 47 references, 21 of which can be accessed free at <http://www.jbc.org/content/279/49/50790.full.html#ref-list-1>