<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Proteomics*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Bendz, M., Skwark, M., Nilsson, D., Granholm, V., Cristobal, S. et al. (2013)
Membrane protein shaving with thermolysin can be used to evaluate topology predictors.
Proteomics, 13(9): 1467-1480
<http://dx.doi.org/10.1002/pmic.201200517>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-90357>

Membrane protein shaving with thermolysin can be used to evaluate topology predictors.

Maria Bendz [†], Marcin Skwark, Daniel Nilsson,

Viktor Granholm, Susana Cristobal, Lukas Käll

and

Arne Elofsson [†] *

January 16, 2013

[†] To whom correspondence should be addressed

Tel: +46-8-16 4672

Email: arne@bioinfo.se

Email: maria.bendz@cbr.su.se

*Science for Life Laboratory, Center for Biomembrane Research, Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, SE-106 91 Stockholm, Sweden E-mail: arne@bioinfo.se, maria.bendz@cbr.su.se

Abstract

Topology analysis of membrane proteins can be obtained by enzymatic shaving in combination with mass-spectroscopy identification of peptides. Ideally such analysis could provide quite detailed information about the membrane spanning regions. Here, we first examined the ability of some shaving enzymes to provide large-scale analysis of membrane proteome topologies. To compare different shaving enzymes we first analyzed the detected peptides from two over-expressed proteins. Secondly, we analyzed the peptides from non over-expressed *Escherichia coli* membrane proteins with known structure to evaluate the shaving methods. Finally, the identified peptides were used to test the accuracy of a number of topology predictors. At the end we suggest that the usage of thermolysin, an enzyme working at the natural pH of the cell for membrane shaving is superior because: (i) we detect a similar number of peptides and proteins using thermolysin and trypsin. (ii) Thermolysin shaving can be run at a natural pH and (iii) the incubation time is quite short. (iv) Fewer detected peptides from thermolysin shaving origin from the transmembrane regions. Using thermolysin shaving we can also provide a clear separation between of the best and less accurate topology predictors, indicating that using data from shaving can in the future provide very valuable information when developing future topology predictors.

1 Introduction

Integral membrane proteins are critical for many cellular processes as transporters, receptors, enzymes etc. Two subgroups of membrane proteins based on the secondary structure of the protein: α -helical and β -sheet membrane proteins. The importance of membrane proteins and especially the α -helical group is reflected by the fact that in most genomes they encode for 20-30 % of the genes [1]. One unique feature of membrane proteins is that they face both hydrophilic and hydrophobic environments and therefore need to expose both hydrophobic and hydrophilic residues on their surfaces. This property is essential for their function in the cell but puts scientist in a troublesome position as it often contributes to major analysis difficulties. Usually the α -helical membrane proteins cause most problems as detergents often are needed in various experiment and they might actually affect the outcome of the analysis. The hydrophobicity of membrane proteins also results in a lower frequency of arginine and lysine within the transmembrane regions, affecting the outcome in mass spectrometry.

Membrane protein structure has been shown to be more complex than earlier thought, with long helices, deep core coils [2] and reentrant regions being common [3]. Our understanding of membrane protein structure is still to a large extent drawn from a small set of membrane proteins that can be over-expressed and crystallized. As have been shown in recent studies all properties of this small set of proteins do not transfer to all membrane proteins. Therefore, the use of large-scale methods to identify and predict accurate membrane protein topologies would be useful to increase our understanding of membrane protein biogenesis and structure.

Knowledge about protein structure is often crucial for understanding its function, and the crystallization difficulties of membrane proteins results in an underrepresentation of this group among high-resolution structures. As a compensation for the lack of membrane protein structures, topology observations and predictions, play an important role as they can give an idea of whether a protein is a membrane protein or not and give a starting point in the investigation of its function. Therefore, a number of methods to predict the topology of α -helical membrane proteins have been developed [4, 5, 6, 7, 8, 9, 10]. Topology predictions are based on

information from the amino acid sequence and trained on experimentally verified topologies. The topology information can be obtained by different biochemical methods including glycosylation mapping, reporter fusions and chemical modifications including cysteine scanning, epitope mapping and proteolysis but also from the structure of known TM-proteins. However, all these methods provide quite small datasets of low-resolution topologies of not necessarily unbiased proteins. Also, it has been noted lately that it is likely that the reported performance of the best topology prediction methods is most likely not maintained for entire genomes [11, 12].

In general membrane protein topology prediction methods are based on the observations that (i) transmembrane α -helices are significantly more hydrophobic than other regions in proteins and (ii) the charge distribution of the loops connecting the transmembrane helices follows the “positive inside” rule, which states that non translocated loops are enriched in positively charged residues compared to translocated loops [13, 14]. Available methods for topology predictions are quite good at predicting the number of transmembrane helices in a protein but quite bad at identifying exactly start and end of a TM helix [15]. Also, by adding limited experimental information about the location of C-termini prediction accuracies can be improved [16].

Traditionally used biochemical methods for topology studies can be divided into two groups, (i) post-translational modifications of side-chains and (ii) fusion based methods. The first set of approaches are based on two assumptions: a) the impermeability of hydrophilic molecules i.e. they do not enter the membrane and b) the embedded transmembrane regions of a protein are not affected by these molecules. Here, target sites in combination with unique binding partners can be used within the membrane protein of interest and upon binding provide information about the topology can be gained. The most common tags are cysteine modifications and glycosylation scanning [14, 17].

A major drawback with these biochemical methods is that a single protein or construct at a time is studied which makes them time consuming and expensive. Anyhow, providing sufficient resources global analyzes of transmembrane protein topologies is possible [18, 19, 20].

A fourth approach to experimentally determine topologies is to use a protease to cleave the

protein of interest at non-membrane loops. Proteolysis for topology studies was also originally dealing with one protein at the time [21, 22]. However, with the entrance of mass spectrometry as a tool for identifying removed peptides [23] made it possible to analyze many proteins simultaneously [24]. The advantages of enzymatic shaving versus other biochemical methods is that many proteins, maybe hundreds, may be scanned at the same time and that no genetic modification is needed. Specific enzymes like trypsin, V8 (Glu-C), trypsin and Proteinase K have been used to shave before the era of mass spectrometry [25, 26]. With the entrance of mass spectrometry as a tool for peptide analysis, focus was set on trypsin as the favored enzyme for protein identification and therefore most tools are developed for trypsin digestions. The advantage of trypsin is its specificity, the length of the produced peptides and the peptides charge distribution, often resulting in identifications. However, for membrane proteins this enzyme often fails to deliver due to the lack of available tryptic digestion sites. Therefore, other less specific enzymes such as Proteinase K, [27], pepsin [28] and elastase [29], have been suggested and given promising results, especially in combination with an N-terminal modification, the number of peptides increased drastically (Proteinase K digests) [30].

Enzymatic shaving of the membrane have been described in the literature as a method for analyzing surface exposed peptides, i.e. to shave of the loops from membrane proteins. It has been claimed that the proteases cannot enter the membrane. Proteinase K at high pH, using the so-called hppk method, and trypsin has also been suggested to be useful in topology analysis of membrane protein [31, 32]. Peptides from the shaved loops are then identified by mass spectrometry. In addition to increasing the number of transmembrane proteins with, at least partial, topology information the large amount of data received in a reliable shaving experiment could also help to improve prediction methods especially when it comes to exactly span of α -helices.

Here, we have examined shaving using two different enzymes thermolysin and trypsin, for their ability to obtain topological information about transmembrane proteins in *E. coli*. Two other enzymes, the hppk method (Proteinase K at high pH) and pepsin were also tried but in our hands using these enzymes only a handful of peptides were reliably identified and therefore

the results from these methods are not shown here. All shavings were done in duplicates.

Primarily we examined peptides identified from two over-expressed protein, thereafter we used membrane proteins with known structure in *E coli* to systematically examine how well the different shaving protocols performed and finally we also examined the identified peptides from non-overexpressed proteins. Here we wanted to examine both the properties of the two shaving protocols and properties of different topology predictors. Here, the assumption is that a method that is better at distinguishing between membrane and non-membrane regions of a protein is superior to a method showing less separation. To our knowledge, this is the first study to control the quality of shaving protocols and systematically test if it corrects to assume that these enzymes do not detect membrane-spanning peptides and thermolysin has not been widely used in this type of studies before, but it is known that thermolysin is a “gentle” protease [33]. We also show that a set of topology predictors, previously identified to be among the best ones, clearly appear to be superior to other predictors when using the identified peptides as the standard of truth, indicating that shaving protocols can provide valuable data for evaluating topology predictors. However, the usefulness of shaving data for benchmarking would increase significantly if shaving were could be performed so that only peptides from one side of the membrane were cleaved off. Given the short incubation time of thermolysin this might be possible.

2 Materials and Methods

2.1 Chemicals

LB-Broth Miller was purchased from Formedium (Hunstanton, England). Lysozyme and Tri-ethanolamine were purchased from Fluka (Sigma-Aldrich, Stockholm, Sweden). Pepsin, thermolysin, iodoacetamide and solvents for high-performance liquid chromatography (HPLC) were purchased from Sigma-Aldrich (Stockholm, Sweden). Sequenced-grade-modified trypsin and Proteinase K was purchased from Promega (SDS, Falkenberg, Sweden). Ethylenedi-

aminetetraacetic acid, EDTA, was purchased from Scherlau (Thermo Fischer, Scientific, Västra Frölunda, Sweden). Dithiothreitol (DTT) was purchased from Pierce (Nordic Biolabs AB, Täby, Sweden), Prefabloc was purchased from Roche Diagnostics (Roche Diagnostics Scandinavia AB, Bromma, Sweden). DC Protein Assay was purchased from Bio-Rad Laboratories (Sundbyberg, Sweden). Kanamycin and Chloroamphenicol were purchased from USB Corporation (Cleveland, USA). Isopropyl β -D-1-thiogalactopyranoside (IPTG) was purchased from Apollo Scientific (Bredbury, Stockport, England) MES 1-(Nicotinoyloxy) Succinimide Ester, Nic-NHS, was a kind gift from Prof Peter James, Lund University and was synthesized as described by Jansson et al [30].

2.2 Cell growth and protein expression

Plasmids containing LepB-TEV-His8 and YiiP-His8 from the in-house library were transformed into *E. coli* BL21(DE3)LysS competent cells and left to grow on LB-Broth agar plates. A colony was inoculated into 20 mL of LB-Broth medium containing 50 μ g/mL kanamycin and 30 μ g/mL chloroamphenicol incubated at 37°C and 175 rpm over night. The culture was diluted 1:50 in LB-Broth medium with 50 μ g/mL kanamycin and 30 μ g/mL chloroamphenicol added and the culture was grown 37°C and 175 rpm until the OD₆₀₀ reached 1.8. 500 μ L of 1M IPTG was added to induce protein expression and it was incubated for another 3 hours at 37°C and 175 rpm. The cells were harvested by centrifugation for 20 minutes at 3000 x g. The cell pellet was washed with 40 mL of PBS and collected by centrifugation at +4°C for 20 minutes at 3000 x g. The supernatant were discharged and the cell pellet was stored at -80°C. For an overview see Figure 1.

2.3 Inner membrane purification

The cell pellet was thawed on ice and washed with 40 mL of 50 mM Triethanolamine, 250 mM Sucrose, 1 mM EDTA, 1 mM DTT, pH 7.5 and pelleted again by centrifugation for 20 minutes at 3000 x g. The pellet was resuspended in 15 mL of 50 mM Triethanolamine, 250

mM Sucrose, 1 mM EDTA, 1 mM DTT, 1mg/mL Prefabloc, 10 μ L/mL Lysozyme, pH 7.5. The suspension was incubated at 30°C and 175 rpm for 20 minutes and cells were lysed using the EmulsiFlex-C3 (Avestin Inc, USA). The cell debris was pelleted by centrifugation at 8000g for 20 minutes at +4°C. The supernatant was collected and distributed on top of a two step sucrose gradient (8.8:55% (w/v) sucrose) and centrifuged at 210 000 x g for 2.5 hours at +4°C with no break. The membrane fraction was collected with a syringe and inner- and outer- membranes were separated on a six-step sucrose gradient as described previously [34, 35, 36]. The inner membrane fraction was collected with a syringe and diluted 1:2 with 100 mM Triethanolamine, 2 mM EDTA, 2 mM DTT, pH 7.5 and pelleted at 148 000 x g at +4°C for 1 hour. The inner membrane pellet was dissolved in 1 mL of PBS and the protein concentration was measured using the DC Protein Assay kit (Bio-rad Laboratories).

The inner membrane was pelleted again by centrifugation at 48 000 x g at + 4 °C for 40 minutes. The pellet was washed three times. First with 10 mM sodium carbonate buffer pH 11, the second time with ddH₂O and then a third time with PBS. Between each wash the membranes were pelleted by centrifugation at 48 000 x g at + 4 °C for 40 minutes. After the final centrifugation PBS was added to each pellet in order give a calculated total protein concentration of 6 mg/mL and it was then stored at -80°C.

2.4 Reducing and modifying Cysteine

The membrane sample was reduced by adding 500 mM DTT to give a final concentration of 5 mM. Thereafter, the sample was incubated for 20 minutes at 60 °C. It was briefly chilled on ice before being alkylated by adding 500 mM iodoacetamide, IAA, to give a final concentration of 15 mM IAA in the sample and incubation in the dark at room temperature for 20 minutes.

2.5 Thermolysin shaving

The sample, 1200 mg of alkylated protein, was pelleted by centrifugation at 48 000 x g and + 4 °C for 40 minutes. The pellet was dissolved in 100 mM Tris, 10mM CaCl₂ pH 7.4. The

sample was then split into 6 parts (25 μ L in each). Four of the samples were incubated at 65°C for 10, 20, 30 or 60 minutes. Two of the samples were incubated at 85°C for 10 or 60 minutes. The membrane lipids including the membrane embedded peptides were then pelleted by centrifugation at 135 000 x g at 4 °C for 45 minutes. The supernatant was collected.

A digestion temperature of 65 °C gave more peptides originating from the overexpressed protein than a digestion at 85 °C. Increasing the digestion time to more than 10 minutes did not improve the results. We therefore chose to analyze the thermolysin shaving samples with the incubating temperature set to 65 °C and with an incubation time of 10 minutes, using the LTQ Orbitrap XL (Thermo Scientific, USA).

2.6 Trypsin shaving

In the trypsin experiments 12, μ g of trypsin were added to 400 μ g of alkylated total protein in PBS, and it was incubated over night at +37°C. The membrane lipids and the membrane embedded parts of the protein were pelleted by centrifugation at 135 000 x g at 4 °C for 45 minutes. The supernatant containing the shaved off peptides was collected.

2.7 Proteinase K shaving

Proteinase K shaving of membrane. A total amount of 400 μ g alkylated protein, was pelleted by centrifugation at 48 000 x g and +4 °C for 40 minutes. The pellet was dissolved in 50 μ L of 200 mM sodium carbonate buffer pH 11 containing 8 μ g (2 mg/mL) of Proteinase K. After being incubated for 3 hours at 37 °C, another 8 μ g of Proteinase K in 200 mM sodium carbonate buffer pH 11 were added. The sample was incubated for another 1.5 hours at 37 °C. The membrane lipids plus the membrane embedded peptides were then pelleted by centrifugation at 135 000 x g at 4 °C for 45 minutes. The supernatant containing the shaved off peptides was collected. The shaved off sample was speed vacuumed until dryness and resuspended in 200 μ L of 200 mM triethyl-ammonium bicarbonate buffer pH 7.

2.8 Pepsin shaving of membrane

A total of 400 μg alkylated total protein, was pelleted by centrifugation at 48 000 x g and + 4 °C for 40 minutes. The pellet was dissolved in 50 μL 200 mM MES buffer pH 6. 8 μg of pepsin in 200 mM MES buffer pH 6 (2 mg/mL pepsin) were added to the membrane sample. After being incubated for 3 hours at 37 °C, another 8 μg of pepsin in 200 mM MES buffer pH 6 were added. The sample was incubated for another 1.5 hours at 37 °C. The membrane lipids plus the membrane embedded peptides were then pelleted by centrifugation at 135 000 x g at 4 °C for 45 minutes.

The supernatant was collected, dried by speed vacuum and the peptides were resuspended in 200 mM triethyl-ammonium bicarbonate buffer pH 7.

2.9 Modification of peptides

The pH of the supernatant was adjusted to 7 in the samples containing non-tryptic peptides and the peptides were modified by adding 200 mM Nic-NHS dissolved in DMF to a final concentration of 14 mM Nic-NHS. The reaction was carried out on ice for 1 hour. Any side reactions were removed by adding 600 μM hydroxylamine in 25 mM Tris buffer pH 8.5 to the sample to reach a final concentration of 20 μM hydroxylamine and incubated at room temperature for 20 minutes. The pH of the sample was increased to 11-12 by adding 6M NaOH, and the sample was incubated for 20 minutes at room temperature.

2.10 Preliminary mass spectrometry analysis of thermolysin digest

All thermolysin shaved samples were desalted using StageTips (Proxeon, Odense, Denmark), speed vaced to dryness to remove acetonitrile, and resuspended in 2.5% formic acid before they were analyzed on a Q-tof Ultima API (Waters, Manchester, UK) coupled to a Waters CapLC system. 6 μL of sample were injected and trapped on the pre-column (C18, 300 μm x 5 mm, 5

μm , 100 Å, LC-Packings) and separated on a reversed phase analytical column (Atlantis, C18, 75 μm x 150 mm, 3 μm , 100 Å, Waters). The flow rate was set to 200 nL/min and solvent A consisted of 2% acetonitrile and 98% water with 0.1% formic acid while solvent B consisted of 90% acetonitrile and 10% water and 0.1% formic acid. The instrument was set to analyze +1 and +2 ions.

All files generated from the Q-tof were processed using ProteinLynx V2.2. The data were analyzed using Mascot version 2.3.01 with the enzymes set to none, peptide mass tolerance +/- 0.1, fragment mass tolerance +/- 0.1, against the SwissProt database 2011_5. The searches were done with NicNHS N-terminal modification and carbamidomethyl modification of cysteine set to be fixed, while the NicNHS lysine modification was set to be variable. Two of the samples were chosen for further studies using the LTQ Orbitrap XL (Thermo Scientific, USA) see below.

2.11 Mass spectrometry analysis of shaving experiments.

Samples originating from trypsin, Proteinase K and pepsin shaving were desalted in the same way as the thermolysin shaving samples. Samples from shaving using trypsin, Proteinase K, pepsin and the two thermolysin experiments were analyzed on a LTQ Orbitrap XL (Thermo Scientific, USA) coupled to an Eksigent nano HPLC. The auto sampler injected x μL of sample with a speed of 10 $\mu\text{L}/\text{min}$ and the peptides were trapped on a pre-column (Zorbax 300SB-C18 5 x 0.3 mm, 5 μm , Agilent Technologies), and separated on a reversed phase analytical column (Zorbax 300SB-C18 150 x 0.75 mm, 3.5 μm , Agilent Technologies) with a flow rate of 350 nL/min. Solvent A consisted of 0.1% formic acid in water and solvent B consisted of 0.1% of formic acid in acetonitrile. Prior to the sample injection the trap column was flushed for 10 minutes with 95% of solvent A. After the samples injection the column was washed 100% solvent A for 15 minutes and then solvent B was raised up to 40% over 85 minutes, from 40% to 80% over 5 minutes, held at 80% for 15 minutes and then quickly reducing to 5% over 1 minute. The total runtime was 105 minutes and the instrument was set to analyze +1, +2 and

+3 ions during all runs.

2.12 Peptide and protein identification

The raw files of Orbitrap HPLC-MS/MS data were converted to the ms2-file format using ProteoWizard [37] and analyzed using Crux [38] version 1.36 and Percolator [39] version 2.02 with the embedded protein inference algorithm Fido [40]. Crux was run in the sequest-search mode using non-specific enzyme cleavage, a 15 ppm precursor window, fixed carbamidomethylation (+57.02 Da) on C and variable oxidation (+15.99 Da) on M. Additionally, for all enzymes except trypsin, fixed and variable NicNHS-modifications (+105.02 Da) on the N-terminal and K, respectively, were applied. The peptide-spectrum matches from Crux were further post-processed by Percolator with cleavage features specific for each enzyme. For all searches, an *E coli* BL21-GoldDE3pLys_{AG} protein sequence database obtained from Pathosystems Resource Integration Center (VA, USA) on Dec 20, 2011, was used. For estimating error rates, the protein sequences were reversed to produce a decoy database that was searched separately [41].

2.13 Topology prediction and analysis

One of the goals with this analysis was to examine how the different shaving enzymes could be used for detailed topology studies, but also how well different topology predictors can accurately identify the peptides.

First, we developed a scheme to examine if the identified peptides overlapped with the membrane regions or where found in loops domains. For each identified peptide it was compared with either the topology according to the OPM classification [42] for membrane proteins of known structure as illustrated in Figure 2.

Homology search against the OPM database was performed using a single round of context-sensitive homology detection method - CS-BLAST [43] against a dedicated database of evolutionary profiles of protein sequences contained in OPM. Only matches with E-value lower than

10^{-10} was kept. Use of CS-BLAST allowed for greater sensitivity than regular BLAST run, while retaining similar computational costs. Each alignment received from CS-BLAST was then contrasted with the topology stored in OPM and resulted in a homology-derived topology of a query protein. Topologies obtained from OPM should be quite precise in their description of the membrane locations although the exact location of the membrane borders can sometimes be difficult to determine.

To evaluate the performance of different topology predictors we used the fact that only a fraction of the identified proteins have close homologs in OPM. For all identified proteins the topology was predicted using TOPCONS [10] and TOPCONS-single [44]. These predictions are constructed from a consensus of different topology prediction algorithms, including: SCAMPI-seq [9] (single sequence mode), SCAMPI [9] (multiple sequence mode), PRODIV-TMHMM [6], PRO-TMHMM [6], OCTOPUS [8], Memsat [45], sTMHMM [6], toppred2 [4].

Although that the accuracy of the best topology predictors is estimated to be above 70%, it is well established that topology predictors cannot exactly locate the transmembrane regions [46]. Frequently the predicted TM-helices represent the most hydrophobic region and not necessarily the exact part that is found within the membrane [47]. Therefore the ability to accurately predict the location of the peptides in the shaving procedure should provide at least some information about the ability of a certain topology predictor to accurately identify the membrane regions of a protein.

3 Results

Although trypsin is the most common enzyme in mass spectrometry based proteomics, other enzymes including Proteinase K, pepsin and elastase have also been used for digesting proteins. These proteins have also been used for enzymatically shaving the membrane in order to get topology information. To make enzymatic shaving successful, however, it is important to (i) keep the pH neutral during shaving preventing structural changes close to the membrane and (ii) has as short incubation time as possible during shaving, especially if working with

whole cells. Thermolysin, an enzyme working at pH 7 and having short digesting time fulfills these demands.

3.1 Detection of peptides in over-expressed proteins.

Enzymatic membrane shaving has been presented several times in the literature but to the best of our knowledge no detailed study of the obtained topology information from these different methods has been reported. Here, we over-expressed independently two different *E coli* membrane proteins of known structure, YiiP (FieF) and LepB, from the inner membrane and purified the membrane of interest by a two-step sucrose gradient centrifugation. The inner membrane was shaved using four different enzymes, pepsin, Proteinase K, thermolysin or trypsin, and the resulting peptides were analyzed using mass spectrometry. A schematic figure of the setup is visualized in Figure 1. To improve the number of identifications and the individual score of each peptide among the non-tryptic shavings, N-terminal modification have been added to those peptides as described earlier [30]. However, even with these modifications the number identified peptides with significant scores from the pepsin and Proteinase K experiments were very small (less than 10). Therefore these two enzymes were ignored in the further analysis. So at the end we use the results of four experiments, over-expressed YiiP or LepB and shaving using either thermolysin or trypsin.

In all four experiments a number of peptides were identified in the over-expressed proteins, see Figure 3. Most of the identified peptides come from the long C-terminal parts of the over-expressed proteins. However, using trypsin two peptides outside these regions are found. In YiiP one peptide from a transmembrane region and in LepB one peptide from the long intracellular loop in LepB was found. This indicates that at least with trypsin shaving it is possible to pick up peptides from the transmembrane regions, but also that both shaving protocols primarily detect peptides in the long extracellular domains. However, the main advantage of shaving protocols for topology identification is the possibility to scan a large number of proteins simultaneously. Therefore, we proceeded to analyze the peptides identified from non-over-expressed

proteins.

3.2 Trypsin and thermolysin shaving identify a similar number of proteins.

A general goal of many proteomics studies is to reliably detect as many proteins as possible. However before the introduction of proper statistical models, such as decoy searches [41, 48] it has been hard to accurately tell correct from incorrect identifications. Particularly it has been challenging to find score thresholds that is common for different proteases [39]. To address this problem we have used a pipeline providing a good estimate of false discovery rate (q value) both for identified proteins and peptides.

In Figure 4 and Table II the number of identified proteins and peptides in the four experiments is shown. We have compared the number of identifies amino acids, peptides and proteins at different false positive rates. In all four samples in the order of 200 proteins, 300 peptides and a few thousand unique amino acids are identified at comparable q values. It can be noted that at stricter q value trypsin shaving detects more peptides and proteins than thermolysin, while at q values higher than approximately 0.1 the opposite is found. Further, trypsin consistently detects more unique amino acids than thermolysin as the detected peptides are on average slightly longer, 17 vs. 10 amino acids. In summary, shaving with either trypsin or thermolysin identifies a comparable number of peptides. However, as noted in the study of the over-expressed proteins the identified peptides might not all origin from extracellular loops in membrane proteins and were therefore analyzed further using all peptides identified at a false discovery rate of 5%.

At a false discovery rate of 5% trypsin shaving identified slightly more proteins. In all four experiments about one third non-membrane proteins are found, see Table II. This indicates that our purification protocol substantially enhances the fraction of membrane proteins but that some non-membrane proteins that might be located in the vicinity of the membrane proteins are still present in the sample.

In Table: III the number of proteins jointly identified by two experiments is shown. On average 57% of the proteins detected in one experiments is detected in another experiment. However, the two different trypsin experiments show a substantially larger overlap (69%) is found, but it should be remembered that more proteins are detected with trypsin shaving. Anyhow, the quite substantial number of proteins identified by the different enzymes shows that both methods can be reliable used to identify quite some membrane proteins in *E coli*.

3.3 Using the identified peptides to provide topology.

In table IV the number, and fraction of all, of amino acids detected in the membrane regions, extracellular and periplasmic regions of the identified membrane proteins are shown. First it is clear that both shaving protocols identify some TM regions. However, it is also clear that a significantly ($P < 10^{-4}$ by Fishers's exact test) lower fraction of the transmembrane residues than the non-transmembrane residues are We also examined how close to the membrane regions the identified peptides were located, see figure 5. Clearly both for trypsin and thermolysin there is a strong increase of the number of amino acids detected starting roughly ten residues away from the membrane. However, it can also be noted that it appears to be slightly more residues detected within or close to the TM region using trypsin than thermolysin, as was also observed for the two over-expressed proteins, see Figure 3.

4 Discussion

Trypsin is the most commonly used enzyme within mass spectrometry based proteomics, however when it comes to membrane proteins important information might be lost due to the low number accessible trypsin digestion sites. Most available tryptic digestion sites in membrane proteins are located in large loops or at soluble parts at the N-or C-terminal. One drawback with trypsin is the time needed for digestion, usually 16-18 hours, while one major advantages of trypsin shaving is that no N-terminal modification is required as the charge of a tryptic

peptide is located at the N- and C-termini.

It is likely that the hydrophobicity of transmembrane helices in membrane proteins is affected by the pH. By changing the pH of the sample buffer a sample the hydrophobicity of individual transmembrane helices or just parts of them is likely to alter. A high pH will especially affect the solubility of amino acids with a hydroxyl group on its side-chain. Threonine, tyrosine and serine become more soluble at high pH and this phenomenon have been used in separations of peptides using HPLC [49]. This could likely affect the results for the hppk method. However, in this study we did not obtain sufficient number of identified peptides with hppk, not with pepsin. Both enzymes appeared to be too efficient in digesting and the resulting peptides were short. One explanation for the short peptides in the Proteinase K experiments could be that the pH of the shaving buffer was lowered during shaving by the carbon dioxide from the air, resulting in a more efficient enzyme giving short peptides. Pepsin can unfortunately not be controlled by pH in the same way as Proteinase K. The short peptides could also be an issue if working with whole cells as they might break during the long incubation.

By the introduction of thermolysin in this study we appear to be able to avoid some of the problems with trypsin. In particular the incubation time is much shorter. One other advantage of thermolysin shaving over hppk method is that the experiments can be run at neutral pH. The high pH of the hppk method opens the membrane again making it impossible to shave only one side of the membrane. A drawback, as with most non-tryptic peptides, is the need of adding a N-terminal modification step, to the procedure in order to increase the number of identified peptides. The Nic-NHS modification forces the charge of the peptide to the N-terminal, decreasing the number of internal ions in the MS/MS spectrum during the mass spectrometry analysis. Tryptic peptides have minor problems with these internal ions, as the charge of the peptide is naturally located at Arginine or Lysine (at the C-terminal) and at the N-terminal. As with the hppk method it is possible to identify peptides from thermolysin shaving experiments without adding the N-terminal modification, however, the number of peptides reliably identified with the modification is greater than without. Further, as shown by Jansson et al [30] the modifications increase the number of peptides originating from hydrophobic membrane

proteins having more than one transmembrane helices increases compared to without the modification.

In summary: (i) trypsin is quite good at separating transmembrane and non-transmembrane peptides, but that most of the detected peptides found are from large domains/long loops. (ii) Pepsin clearly detects fewer peptides than the other methods and also primarily within long loops. (iii) Both thermolysin and in particular Proteinase K has a tendency to include membrane regions among the detected peptides, but these methods can also detect peptides from short loops. This problem can, at least for thermolysin, be minimized if a minimum length of the peptide was assumed. (iv) Thermolysin detects a significantly larger number of peptides than the other methods.

4.1 Can shaving be used to benchmark topology predictors ?

One of the goals of this study was to examine if large-scale proteomics data could be used to evaluate, and in the future improve, methods developed to predict the topology of membrane proteins. At the moment, the data we have obtained does not contain any direct information about the locations of the non-membrane regions, which of course would have been more informative. However, already with the experiments performed here it seems to be possible to quite accurately locate the transmembrane regions in a much larger set of proteins than can be done by traditional biochemical experiments. Therefore, we set out to try to examine if the experimental data used here could be used to differentiate the performance of different topology predictors.

It was first noted that for a much higher fraction of residues that are predicted to be in the membrane are found in the trypsin experiments (0.7-1.9% excluding PRODIV) than in the thermolysin experiments (0.2 - 0.5%). Even using PRODIV [6] that is known to over-predict TM-regions [12] more predicted membrane regions are identified using trypsin than thermolysin. This might indicate that trypsin shaving actually is of lesser value for comparing the different predictors and therefore we only used the thermolysin experiments to evaluate the

predictors below.

In Table V it can be seen that some methods, including SCAMPI-seq, TOPCONS and OCTOPUS predict fewer of the identified peptides to be in the membrane than other methods. The ratio of non-membrane/membrane residues in these peptides is also improved for these methods. In general methods that use multiple sequence alignments together with SCAMPI-seq perform best here. Further, it is also clear that PRODIV [6] predicts many more of the peptides to be located in the membrane than the other methods. All these results are in agreement with recent observations of the performance of these predictors on large datasets [12], indicating that shaving data can be used to optimize the performance of topology predictor or even be used in combination with them.

5 Conclusions

Traditional biochemical methods like cysteine scanning and glycosylation assays for determination of the topology of membrane proteins often deals with one protein at a time and are therefore time consuming for global studies. To speed up the analysis and to produce large dataset to improve topology prediction methods, large-scale methods using mass spectrometry have been suggested. We have investigated the performance of some of those large-scale methods. Our results show thermolysin is a promising candidate for shaving studies. In particular relatively fewer membrane-spanning peptides are identified by thermolysin than by other methods. Finally, we show that already today shaving studies can be used to differentiate between the qualities of topology predictors. This indicates that in the future it should be possible to significantly enlarge the amount of reference points when training such predictors.

However, still one major drawback with the shaving data is that there is no separation between peptides origin from the cytoplasmic and periplasmic sides of the membrane. One possibility to obtain sidedness information could be to perform shaving on spheroplasts, but there is a significant risk that even the most gently shaving protocols breaks the membrane so that cytoplasmic peptides also are cleaved.

One problem both with biochemical methods and the proteomics methods used in this study is that they all somehow modify the protein and that the time-scale for measuring the topology is long enough to allow for structural rearrangements. Therefore methods that make minor modifications in vivo would have clear advantages.

It has recently been shown that oxidisation in vivo of hydroxyl radicals followed by digestion and identifications can be used to identify external loops in outer membrane proteins [50, 51]. Alternatively the FASP method developed by Wisniewski et al [52] can be used to determine the topology of a large set of membrane proteins [53].

However, as we show here the significantly simpler approaches used here provide sufficient information for benchmarking of membrane protein topology predictors.

6 Acknowledgement

We thank Dan Daley and Isolde Palembo at the Department of Biochemistry and biophysics at Stockholm University for kindly providing plasmids of LepB and YiiP (FieF) for this study. We also thank Prof Peter James at Dept Immunotechnology, Lund University for kindly letting MB use his mass spectrometers for our analysis and for his kind gift of Nic-NHS chemical used in this article. We thank Karin Hansson at Dept Immunotechnology, Lund University for helping with running the samples.

Support by BILS (Bioinformatics Infrastructure for Life Sciences) through Fredrik Levan-der is gratefully acknowledged. Finally we thank Per-Olof Edlund, Biovitrum and Dept of Analytical Chemistry, Stockholm University, for suggesting the usage of thermolysin.

This work was supported by grants from the Swedish Research Council (VR-NT 2009-5072, VR-M 2007-3065 to AE and VR-NT 2008-3454 to LK), SSF (the Foundation for Strategic Research). MB was supported by a Swedish Research Council postdoctoral grant (2009-698). MS was supported by the EU 7th framework through the ITN project Transys (FP7-PEOPLE-2007-1-1-ITN). The EU 6th Framework Program is gratefully acknowledged for support to the EDICT project, contract No: FP7-HEALTH-F4-2007-201924 .

References

- [1] Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001, 305:567–580.
- [2] Kauko, A., Illergard, K. & Elofsson, A. Coils in the membrane core are conserved and functionally important. *J Mol Biol* 2008, 380:170–180.
- [3] Viklund, H., Granseth, E. & Elofsson, A. Reentrant regions in α -helical transmembrane proteins are divided in 3 structural classes and abundant in small residues. *J. Mol. Biol.* 2006, 361:591–603.
- [4] Claros, M. & von Heijne, G. TopPred II: an improved software for membrane protein structure prediction. *Comput. Appl. Biosci.* 1994, 10:685–686.
- [5] Kall, L. & Sonnhammer, E. Reliability of transmembrane predictions in whole-genome data. *FEBS Lett.* 2002, 532:415–418.
- [6] Viklund, H. & Elofsson, A. Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. *Protein Sci.* 2004, 13:1908–1917.
- [7] Jones, D. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* 2007, 23:538–544.
- [8] Viklund, H. & Elofsson, A. OCTOPUS:Improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* 2008, 24:1662–1668.
- [9] Bernsel, A., Viklund, H., Falk, J., Lindahl, E. *et al.* Prediction of membrane-protein topology from first principles. *Proc Natl Acad Sci U S A* 2008, 105:7177–7181.

- [10] Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res* 2009, 37:W465–8.
- [11] Fagerberg, L., Jonasson, K., von Heijne, G., Uhlen, M. & Berglund, L. Prediction of the human membrane proteome. *Proteomics* 2010, 10:1141–1149.
- [12] Tsirigos, K., Hennerdal, A., Kall, L. & Elofsson, A. A guideline to proteome-wide alpha-helical membrane protein topology predictions. *Proteomics* 2012, 12:2282–2294.
- [13] von Heijne, G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol* 1992, 225:487–494.
- [14] van Geest, M. & Lolkema, J. S. Membrane topology and insertion of membrane proteins: search for topogenic signals. *Microbiol Mol Biol Rev*, 2000, 64:13–33.
- [15] Cuthbertson, J. M., Doyle, D. A. & Sansom, M. S. Transmembrane helix prediction: a comparative evaluation and analysis. *Protein Eng Des Sel*, 2005, 18:295–308.
- [16] Rapp, M., Drew, D., Daley, D., Nilsson, J. *et al.* Experimentally based topology models for E. coli inner membrane proteins. *Protein Sci* 2004, 13:937–945.
- [17] Cheung, J. C. & Reithmeier, R. A. Scanning N-glycosylation mutagenesis of membrane proteins. *Methods* 2007, 41:451–459.
- [18] Daley, D., Rapp, M., Granseth, E., Melen, K. *et al.* Global topology analysis of the Escherichia coli inner membrane proteome. *Science* 2005, 308:1321–1323.
- [19] Griffin, N. & Schnitzer, J. Overcoming key technological challenges in using mass spectrometry for mapping cell surfaces in tissues. *Mol Cell Proteomics* 2011, 10:R110.000935.
- [20] Savas, J., Stein, B., Wu, C. & Yates, J., 3rd Mass spectrometry accelerates membrane protein analysis. *Trends Biochem Sci* 2011, 36:388–396.

- [21] Andersson, H., Bakker, E. & von Heijne, G. Different positively charged amino acids have similar effects on the topology of a polytopic transmembrane protein in *Escherichia coli*. *J Biol Chem* 1992, 267:1491–1495.
- [22] Dumont, M., Trewhella, J., Engelman, D. & Richards, F. Stability of transmembrane regions in bacteriorhodopsin studied by progressive proteolysis. *J Membr Biol* 1985, 88:233–247.
- [23] Moore, C., Yates, J., 3rd, Griffin, P., Shabanowitz, J. *et al.* Proteolytic fragments of the nicotinic acetylcholine receptor identified by mass spectrometry: implications for receptor topography. *Biochemistry* 1989, 28:9184–9191.
- [24] Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* 2003, 422:198–207.
- [25] Lindstrom, J., Gullick, W., Conti-Tronconi, B. & Ellisman, M. Proteolytic nicking of the acetylcholine receptor. *Biochemistry* 1980, 19:4791–4795.
- [26] Conti-Tronconi, B. M., Dunn, S. M. & Raftery, M. A. Functional stability of Torpedo acetylcholine receptor. Effects of protease treatment. *Biochemistry* 1982, 21:893–899.
- [27] Wu, C., MacCoss, M., Howell, K. & Yates, J., 3rd A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* 2003, 21:532–538.
- [28] Rietschel, B., Bornemann, S., Arrey, T., Baeumlisberger, D. *et al.* Membrane protein analysis using an improved peptic in-solution digestion protocol. *Proteomics* 2009, 9:5553–5557.
- [29] Baeumlisberger, D., Arrey, T., Rietschel, B., Rohmer, M. *et al.* Labeling elastase digests with TMT: informational gain by identification of poorly detectable peptides with MALDI-TOF/TOF mass spectrometry. *Proteomics* 2010, 10:3905–3909.

- [30] Jansson, M., Warell, K., Levander, F. & James, P. Membrane protein identification: N-terminal labeling of nontryptic membrane protein peptides facilitates database searching. *J Proteome Res* 2008, 7:659–665.
- [31] Tjalsma, H., Lambooy, L., Hermans, P. & Swinkels, D. Shedding & shaving: disclosure of proteomic expressions on a bacterial face. *Proteomics* 2008, 8:1415–1428.
- [32] Blackler, A., Speers, A., Ladinsky, M. & Wu, C. A shotgun proteomic method for the identification of membrane-embedded proteins and peptides. *J Proteome Res* 2008, 7:3028–3034.
- [33] Cline, K., Werner-Washburne, M., Andrews, J. & Keegstra, K. Thermolysin is a suitable protease for probing the surface of intact pea chloroplasts. *Plant Physiol* 1984, 75:675–678.
- [34] Stenberg, F., Chovanec, P., Maslen, S., Robinson, C. *et al.* Protein complexes of the Escherichia coli cell envelope. *J Biol Chem* 2005, 280:34409–34419.
- [35] De Vrije, T., Tommassen, J. & De Kruijff, B. Optimal posttranslational translocation of the precursor of PhoE protein across Escherichia coli membrane vesicles requires both ATP and the protonmotive force. *Biochim Biophys Acta*, 1987, 900:63–72.
- [36] Osborn, M., Gander, J., Parisi, E. & Carson, J. Mechanism of assembly of the outer membrane of Salmonella typhimurium. Isolation and characterization of cytoplasmic and outer membrane. *J Biol Chem* 1972, 247:3962–3972.
- [37] Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* 2008, 24:2534–2536.
- [38] Park, C., Käll, L., Klammer, A., MacCoss, M. & Noble, W. Rapid and accurate peptide identification from tandem mass spectra. *Journal of proteome research* 2008, 7:3022.

- [39] Käll, L., Canterbury, J., Weston, J., Noble, W. & MacCoss, M. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods* 2007, 4:923–925.
- [40] Serang, O., MacCoss, M. & Noble, W. Efficient marginalization to compute protein posterior probabilities from shotgun mass spectrometry data. *Journal of proteome research* 2010.
- [41] Moore, R., Young, M. & Lee, T. Qscore: an algorithm for evaluating SEQUEST database search results. *Journal of the American Society for Mass Spectrometry* 2002, 13:378–386.
- [42] Lomize, M., Lomize, A., Pogozheva, I. & Mosberg, H. OPM: orientations of proteins in membranes database. *Bioinformatics* 2006, 22:623–625.
- [43] Biegert, A. & Soding, J. Sequence context-specific profiles for homology searching. *Proc Natl Acad Sci U S A* 2009, 106:3770–3775.
- [44] Hennerdal, A. & Elofsson, A. Rapid membrane protein topology prediction. *Bioinformatics* 2011, 27:1322–3.
- [45] Jones, D., Taylor, W. & Thornton, J. A model recognition Approach to the Prediction of All-Helical Membrane Protein Structure and Topology. *Biochemistry* 1994, 33:3038–3049.
- [46] Papaloukas, C., Granseth, E., Viklund, H. & Elofsson, A. Estimating the length of trans-membrane helices using Z-coordinate predictions. *Protein Sci* 2008, 17:271–278.
- [47] Kauko, A., Hedin, L., Thebaud, E., Cristobal, S. *et al.* Repositioning of transmembrane alpha-helices during membrane protein folding. *J Mol Biol* 2010, 397:190–201.
- [48] Peng, J., Elias, J., Thoreen, C., Licklider, L. & Gygi, S. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* 2003, 2:43–50.

- [49] Kong, R., Siu, S., Lee, S., Lo, C. & Chu, I. Development of online high-/low-pH reversed-phase-reversed-phase two-dimensional liquid chromatography for shotgun proteomics: a reversed-phase-strong cation exchange-reversed-phase approach. *J Chromatogr A* 2011, 1218:3681–3688.
- [50] Zhu, Y., Guo, T., Park, J., Li, X. *et al.* Elucidating in vivo structural dynamics in integral membrane protein by hydroxyl radical footprinting. *Mol Cell Proteomics* 2009, 8:1999–2010.
- [51] Zhu, Y., Guo, T. & Sze, S. Elucidating structural dynamics of integral membrane proteins on native cell surface by hydroxyl radical footprinting and nano LC-MS/MS. *Methods Mol Biol* 2011, 790:287–303.
- [52] Wisniewski, J., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* 2009, 6:359–362.
- [53] Wisniewski, J., Nagaraj, N., Zougman, A., Gnäd, F. & Mann, M. Brain phosphoproteome obtained by a FASP-based method reveals plasma membrane protein topology. *J Proteome Res* 2010, 9:3280–3289.

7 Tables

Method	Release year	Algorithm	URL	MSA
MEMSAT	1994	ANN + Grammar	http://single.topcons.net/	-
OCTOPUS	2008	ANN + HMM	http://octopus.cbr.su.se/	+
PRO	2004	HMM	http://topcons.net/	+
PRODIV	2004	HMM	http://topcons.net/	+
SCAMPI	2008	Hydrophobicity + Model	http://topcons.net/	+
SCAMPI-seq	2008	Hydrophobicity + Model	http://single.topcons.net/	-
sTMHMM	2004	HMM	http://single.topcons.net/	-
TOPCONS	2009	HMM - Consensus	http://single-topcons.net/	+
TopPred2	1994	Hydrophobicity profiles	http://single.topcons.net/	-

Table I: The membrane topology predictors used in the current study. The key features of each are listed. MSA: the predictor uses multiple sequence alignments as input.

Experiment	No. Prot.	No Pept	No Residues	Non-TM
LepB Thermolysin	195	536		37%
LepB Trypsin	264	744		28%
YiiP Thermolysin	201		417	38%
YiiP Trypsin	235	488		32%

Table II: Overview of the number of identified protein and peptides in the four experiments at FDR of 5%. The first column shows the number of identified proteins, the second the number of identified peptides, the third the number of detected residues and the last the fraction of the proteins that were predicted to be non-transmembrane

	LepB		YiiP	
Experiment	Thermolysin	Trypsin	Thermolysin	Trypsin
LepB Thermolysin		63%	61%	57%
LepB Trypsin	47%		48%	65%
YiiP Thermolysin	59%	63%		59%
YiiP Trypsin	47%	73%	51%	

Table III: Overview of the overlap between of identified proteins in the four experiments at FDR of 5%. The four columns show the overlap between protein identification in the different experiments.

Enzyme	YiiP			LepB		
	M	P	E	M	P	E
Thermolysin	0 (0%)	176 (4.0%)	81 (2.8%)	12 (0.48%)	83 (1.9%)	176 (5.86%)
Trypsin	47 (1.9%)	427 (8.2%)	133 (5.7%)	48 (1.5%)	453 (6.6%)	391 (12.3%)

Table IV: Number of amino acids in identified peptides, that are located in the Membrane, Periplasmic, and Extracellular regions.

Method	YiiP		LepB	
	Enhancement	No. Residues	Enhancement	No. Residues
SCAMPI-seq	22.9	22	18.3	28
TOPCONS	24.1	22	14.4	37
OCTOPUS	22.6	23	13.1	40
SCAMPI	22.6	23	13.1	40
PRO	12.6	38	13.1	38
hmmtop	16.2	30	10.5	48
stmhmm	12.5	37	10.6	45
memsat	12.6	38	10.6	46
toppred	10.6	47	12.4	41
PRODIV	5.6	112	4.4	150

Table V: Enhancement, column 2 and 4, and number, column 1 and 3 of residues found in predicted TM-regions using thermolysin shaving. Enhancement is calculated as the increase in the ratio membrane/non-membrane residues among the detected peptides compared with a random set of peptides.

8 Figures

Figure 1: Overview of shaving methodology used in this study. *E. coli* was grown then the cells were harvested by centrifugation. Then the inner membrane was purified using a high pH wash and alkylation. Thereafter the samples were shaved using proteolytic enzymes and for the non-specific enzymes modified by adding NicNHS. Finally the peptides were identified using HPLC couples with MS/MS analysis.

Figure 2: Schematic illustration of the comprehensive bioinformatic analysis of membrane protein topology.

Figure 3: Peptides identifies from the two over-expressed protein, YiiP (a,b) and LepB (c,d) mapped on the topology of the proteins. Membrane regions are shown as grey boxes, intracellular regions as the lower lines and extracellular regions as the upper lines. The location of all identified peptides with q value less than 0.05 are shown as boxes on top of the topology. To the left identified peptides using trypsin are shown and to the right the peptides identified after thermolysin shaving.

Figure 4: Mass spectrometry results for the four experiment, with YiiP (a,b,c) over-expressed and LepB (d,e,f) . The figures show the number of identifications of proteins, peptides and amino acids using the different enzymes for the samples with the two different proteases are plotted against the q value.

Figure 5: Location of identified amino acids using the different shaving enzymes relative to the membrane borders. Position 0 is defined as the last/first position of a membrane spanning helix. In the left the results for proteins detected in the experiment with YiiP over-expressed are shown and to the right the LepB results are shown. It can be noted that trypsin appears to find slightly more TM-residues than thermolysin.