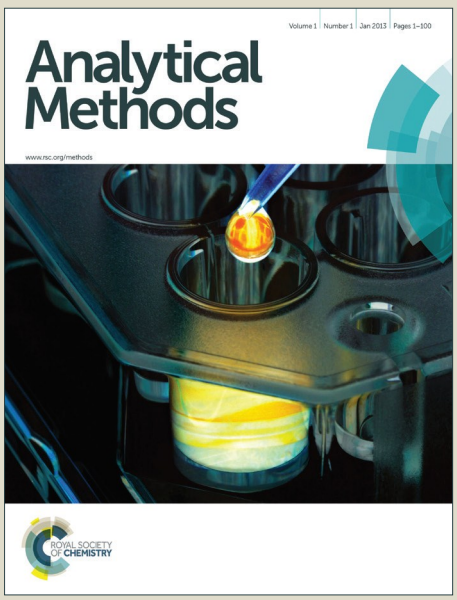
Analytical Methods

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Enhancing the Mass Spectrometric Identification of Membrane Proteins by Combining Chemical and Enzymatic Digestion Methods

*Johanna M. Smeekens, Weixuan Chen, Ronghu Wu**

School of Chemistry and Biochemistry and the Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

*Corresponding author: Phone: 404-385-1515; Fax: 404-894-7452,

E-mail: ronghu.wu@chemistry.gatech.edu

Abstract

Membrane proteins are critical for many cellular events, including cell signaling, molecular transport, and extracellular interactions. One third of the genome is estimated to encode membrane proteins, which are correlated with disease progression and can serve as promising biomarkers and drug targets. Modern mass spectrometry (MS)-based proteomics techniques facilitate the global analysis of proteins in complex biological samples; however, the hydrophobicity of membrane proteins inhibits their comprehensive analysis. Since membrane proteins are not easily accessible by proteases in aqueous solutions, a combinatorial method incorporating chemical and enzymatic digestion is presented here to improve the digestion efficiency of membrane proteins for MS analysis. Chemical digestion with 2-nitro-5-thiocyanatobenzoic acid (NTCB) was supplemented with enzymatic digestion (Glu-C, or Lys-C and trypsin) to determine the optimal combination of digestion methods. Three parallel experiments were performed with membrane protein extracts from HEK293T cells, and the results demonstrated that combining NTCB with Lys-C and trypsin resulted in the greatest number of total peptides (9,483 peptides). Comparatively, digestion with only Lys-C and trypsin allowed the identification of 7,982 total peptides, and sequential digestion with NTCB and Glu-C resulted in 3,307 peptides. By integrating chemical digestion before enzymatic digestion, NTCB could more easily access cleavage sites within membrane proteins, and the resulting peptide fragments were thus more accessible by proteases. The combination of chemical and enzymatic digestion presented here proved to be effective for membrane protein analysis.

1. Introduction

Membrane proteins play extremely important roles in biological systems and are crucial for a variety of cellular events including cell signaling, extracellular interactions and molecular transport.[1-3] They also participate in various cellular functions, including adhesion, growth and metastasis, which contribute to disease progression.[4,5] Additionally, because of their location on the cell surface, and resulting accessibility by macromolecules, membrane proteins are admirable for their potential as therapeutic and diagnostic targets.[6,7] It has been estimated that about one third of the genome encodes membrane proteins,[8] yet they represent 60-70% of FDA approved drug targets.[9,10] The comprehensive analysis of membrane proteins will facilitate a better understanding of membrane protein function and lead to the identification of membrane proteins as effective biomarkers and drug targets.[11] However, the hydrophobic nature and low abundance of membrane proteins hinders their global analysis.[12-14] Modern mass spectrometry (MS)-based proteomics techniques have proven to be very powerful for global protein analysis.[15-20] Common bottom-up proteomics techniques, where proteins are digested into peptides and subsequently analyzed with mass spectrometry (MS),[21-23] require effective digestion wherein proteins remain solubilized so that proteases or small molecules can access cleavage sites.[24] However, the accessibility of hydrophobic membrane proteins has been an existing problem because they tend to aggregate, precipitate and remain tightly folded in aqueous environments.[13]

Enzymatic methods are the most common digestion technique used for MS-based proteomics. Several enzymes are frequently used for protein digestion; the most widely used enzyme is trypsin, which cleaves the peptide bond at the C-terminus of lysine and arginine

residues (unless followed by a proline).[25] Trypsin is preferred due to the relatively high abundances of lysine and arginine, and their distribution throughout proteins,[26] which results in many peptides ideal for MS analysis. Other enzymes investigated here and frequently used for protein digestion include Lys-C and Glu-C. Lys-C cleaves at the C-terminus of lysine residues,[27] and Glu-C cleaves at the C-terminus of glutamic and aspartic acid residues.[28] Membrane protein digestion with only enzymatic methods is challenging since bulky enzymes cannot access all cleavage sites within tightly folded hydrophobic proteins. As a result, digestion exclusively with enzymes is often not sufficient to achieve comprehensive analysis of membrane proteins.

The inaccessibility of membrane proteins by proteases may be at least partially improved with chemical digestion methods, where small molecules can more easily access cleavage sites. However, chemical cleavage generally targets amino acids that are less abundant within proteins, which results in larger peptide fragments.[29] For example, cyanogen bromide (CNBr) targets methionine residues and 2-nitro-5-thiocyanobenzoic acid (NTCB) targets cysteine residues.[30] NTCB only cleaves reduced cysteine residues, so protein reduction is required before digestion can be performed. Digestion with NTCB leads to two types of products, one is a N-terminal peptide and a cyclized N-terminal cysteine, and the other is dehydroalanine, which is the product of β-elimination on the thiocyanato group of cysteine.[31]

Theoretically, the combination of chemical and enzymatic digestion methods would be ideal for membrane protein digestion because small molecules can easily access cleavage sites within membrane proteins, and the resulting fragments would be easily accessible by proteases. In this work, chemical and enzymatic methods were combined to optimize membrane protein

digestion for MS analysis; chemical digestion was first performed with NTCB, and followed by enzymatic digestion with Glu-C or Lys-C and trypsin. These combinatorial digestion methods were compared to sequential enzymatic digestion with Lys-C and trypsin. Our experimental results confirmed that the combinatorial digestion method utilizing NTCB, Lys-C and trypsin was most efficient for membrane protein digestion.

2. Materials and methods

2.1. Chemicals and reagents

NTCB was purchased from Tokyo Chemical Industry, Co., trypsin was from Promega, Lysyl endopeptidase (Lys-C) was from Wako, and Endoproteinase Glu-C from *Staphylococcus aureus* was from EMD Millipore. All other reagents were purchased from Sigma Aldrich if not stated. Zirconia/silica beads (0.5 mm diameter) were purchased from BioSpec Products. HEK293T cells were kindly provided by Dr. Gang Bao's research group at the Georgia Institute of Technology.

2.2. Cell culture, lysis and membrane protein enrichment

HEK293T cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) and harvested when they reached to 80% confluency. Cell pellets were washed three times with phosphate buffered saline (PBS). To remove cytosolic proteins, digitonin buffer (150 mM NaCl, 50 mM HEPES (pH=7.4), 25 μg/mL digitonin, protease inhibitor (1 tablet/10 mL)) was added to the cell pellet and incubated with end-over-end rotation at 4 °C for ten minutes. The suspended cell

pellet was centrifuged at 2,000 g for 10 minutes. Digitonin buffer was added to the cell pellet and the sample was vortexed and subsequently centrifuged at 2,000 g for 10 minutes. This digitonin wash was repeated for a total of two washes.

Lysis buffer (10 mM HEPES (pH=8), 1.5 mM $MgCl_2$, 10 mM KCl) and zirconia/silica beads were added to the cell pellet, and placed in the Mini-Beadbeater (BioSpec). Samples were subjected to three 30 second cycles with 2 minutes of resting on ice in between. A flame-heated needle was used to poke holes in the bottom of the tubes, and the contents were transferred to new tubes (leaving the beads behind) through centrifugation at 1,000 g for 3 minutes. Samples were vortexed and centrifuged at 2,500 g for 5 minutes to remove cell debris. The supernatants were transferred to new tubes and centrifuged at 25,830 g for 30 minutes. Sodium carbonate buffer (0.1M sodium carbonate, 1 mM EDTA) was added to the cell pellets, vortexed, and incubated for 30 minutes on ice. The samples were centrifuged at 25,830 g for 15 minutes, and the supernatant was removed. Urea buffer (75 mM NaCl, 50 mM HEPES (pH=7.4), 8M urea) was added to the cell pellets and incubated with shaking for 30 minutes at room temperature. The samples were centrifuged at 25,830 g for 15 minutes, the supernatants were removed, and the urea incubation was repeated once.[24] After the samples were centrifuged and supernatants were removed, 1% sodium deoxycholate (SDC) in PBS was added to samples to solubilize membrane proteins. Samples were incubated overnight with end-over-end rotation at room temperature. The following day, samples were centrifuged at 15,000 g for 15 minutes; the supernatants were transferred to new tubes and the pellets were discarded. Disulfide bonds within proteins were subjected to reduction by incubation with 5 mM dithiothreitol (DTT) for 25

minutes at 56 °C. The sample was then divided into three equal samples and transferred to 10 kDa filter columns.

2.3. Lys-C and trypsin digestion

One filter column was centrifuged at 10,000 g for 5 minutes. The alkylation reaction buffer (14 mM iodoacetamide, 0.1% sodium dodecyl sulfate (SDS), PBS (pH=8)) was added, and the sample was centrifuged at 10,000 g for 5 minutes. Reaction buffer was added again and incubated in the dark for 20 minutes at room temperature. After incubation, the sample was centrifuged at 10,000 g for 5 minutes, washed with 100 mM sodium acetate (pH=5), centrifuged at 10,000 g for 5 minutes and rinsed with water twice. Digestion buffer (50 mM NH_4OAc, 5% acetonitrile (ACN), 0.1 M urea) and Lys-C (enzyme:substrate ratio of ~1:100) were added to the sample and incubated overnight at 37 °C. The next day, trypsin was added at an enzyme:substrate ratio of ~1:100 and the sample was subsequently incubated for four hours and centrifuged at 10,000 g for 30 minutes. The flow-through was collected, and the filter column was washed with 50 mM ammonium bicarbonate (pH=8.5). The second flow-through was collected and combined with the first. Combined flow-throughs were acidified with trifluoroacetic acid (TFA) to a final pH of ~2, and centrifuged at 2,500 g for 5 minutes. The supernatant was desalted on a 50 mg C18 SepPak cartridge and dried. Eluted peptides were dissolved in 30 µL of MS solvent (5% ACN, 4% formic acid (FA)), and 4 µL were analyzed by liquid chromatography (LC)-MS/MS.

2.4. NTCB and enzymatic digestion

The other two samples were used for combined chemical and enzymatic digestion. First, samples were centrifuged at 10,000 g for 5 minutes. Reaction buffer (10 mM NTCB, 0.1% SDS, PBS (pH=8)) was added to the filter columns which were subsequently centrifuged at 10,000 g for 5 minutes. The buffer was added again to columns and incubated in the dark for 20 minutes at 40 °C, and then centrifuged at 10,000 g for 5 minutes. Sodium acetate (100 mM, pH=5) was added to filter columns, which were then centrifuged at 10,000 g for 5 minutes. Filter columns were rinsed with water and centrifuged at 10,000 g for 5 minutes twice. Next, 50 mM ammonium acetate (pH=9) was added to filter columns and incubated for one hour at 50 °C. Filter columns were centrifuged at 10,000 g for 5 minutes and rinsed with water twice. To one filter column, 200 µL PBS (pH=7.4) and Glu-C (enzyme:substrate ratio of ~1:100) were added. To the second filter column, 200 µL digestion buffer (NH_4OAc (pH=8.5), 5% ACN, 0.1M urea) and Lys-C at an enzyme:substrate ratio of ~1:100 were added. Both samples were incubated overnight at 37 °C. The next day, trypsin was added (enzyme:substrate ratio of ~1:100) to the Lys-C digestion sample, which was subsequently incubated for four hours. Both digestion samples were centrifuged at 10,000 g for 30 minutes. The flow-through was collected for each, and filter columns were washed with 50 mM ammonium bicarbonate (pH=8.5). The second flow-through for each was collected and combined with the first. Combined flow-throughs were quenched with TFA to a final pH of ~2. Acidified samples were centrifuged at 2,500 g for 5 minutes and the supernatant was desalted on a 50 mg C18 SepPak cartridge and dried. Eluted peptides were dissolved in 30 µL of MS solvent, and 4 µL were analyzed by LC-MS/MS.

2.5. LC-MS/MS analysis

Each digestion sample was analyzed with a single LC-MS/MS run. Samples were loaded onto a C18-packed microcapillary column (Magic C18AQ, 5 μm, 200 Å, 100 μm x 16 cm) using a WPS-3000TPL RS autosampler (Thermostatted Pulled Loop Rapid Separation Nano/Capillary Autosampler, Dionex). Peptides were separated by reversed-phase chromatography using an UltiMate 3000 binary pump with a 110 minute gradient of 2-100% ACN containing 0.125% FA. Samples were detected in a hybrid dual-cell quadrupole linear ion trap – Orbitrap mass spectrometer (LTQ Orbitrap Elite, ThermoFisher) using a data-dependent Top 20 method. Cycles consisted of one full MS scan (resolution: 60,000) in the Orbitrap at the Automatic Gain Control (AGC) target of 1 million, and up to 20 MS/MS of the most intense ions in the LTQ.[32-34] Selected ions were excluded from further sequencing for 90 seconds. Ions with a single or unassigned charge were not fragmented and maximum ion accumulation times were 1000 ms for each full MS scan and 50 ms for MS/MS scans.

2.6. Database searches

The raw MS files were converted into mzXML format. Incorrect monoisotopic peak assignments were minimized by checking precursors for MS/MS fragmentation. All MS/MS spectra were searched using the SEQUEST algorithm (version 28)[35] against a database that included sequences of all proteins in the UniProt Human (*Homo sapiens*) Database (updated in February 2014) and common contaminants. To estimate the false discovery rate (FDR) of peptide identification,[36,37] all protein sequences were listed in the forward and reversed orientations. A 20 ppm precursor mass tolerance and 1.0 Da product ion mass tolerance were used in the database search and no enzyme was specified. Samples digested with NTCB were searched with a parameter file listing the following differential modifications: oxidation of methionine

(+15.9949), β-elimination of cysteine (-33.9877), cyclized N-terminal cysteine (+24.9952); and one fixed modification: carbamidomethylation of cysteine (+57.0214). The sample digested exclusively with Lys-C and trypsin was searched using a parameter file that listed the oxidation of methionine (+15.9949) as a differential modification and the carbamidomehylation of cysteine (+57.0214) as a fixed modification.

2.7. Data filtering

The target-decoy method was applied[36,37] to evaluate and further control FDRs of peptide identification. Linear discriminant analysis (LDA) was utilized to distinguish correct and incorrect peptide identifications based on numerous parameters including XCorr, ΔCn, and precursor mass error.[38,39] Separate linear discriminant models were trained for each raw file using forward and reversed peptide sequences to provide positive and negative training data. This approach is similar to other methods in the literature[40,41]. After scoring, peptides less than six amino acids in length were discarded and peptide spectral matches were filtered to a less than 1% FDR at the peptide level based on the number of decoy sequences in the final data set.

3. Results and discussion

3.1. Peptide and protein identification

Membrane proteins were extracted from HEK293T cells as described above, and the membrane protein-enriched sample was equally split into three aliquots for subsequent digestion with

different methods. The first sample was subjected to enzymatic digestion with Lys-C overnight and then trypsin for 4 hours (termed "LT" throughout this paper). Chemical and enzymatic methods were combined to perform the other digestions: one sample was digested with NTCB for 20 minutes and then Glu-C overnight (termed "NG"), and the other was digested with NTCB for 20 minutes, Lys-C overnight and then trypsin for 4 hours (termed "NLT"). These digested samples were purified with the stage-tip protocol and analyzed by LC-MS/MS (Figure 1). All peptides identified from each sample are presented in Supplemental Table 1.

Examples of mass spectra acquired from each digestion sample are shown in Figure 2. All three peptides are from Hspd1, a mitochondrial heat shock protein, which is involved in mitochondrial protein import and may play a subsequent role in the correct folding of imported proteins. This protein can be located in several subcellular components, including the extracellular region, cell surface and membrane.[42] The first peptide, DVANNTNEEAGDGTTTATVLAR, was identified from the sample digested with the LT method with an XCorr of 5.1 and mass accuracy of -1.02 ppm. The second peptide, NAGVEGSLIVEK, identified with an XCorr of 3.5 and mass accuracy of -0.12 ppm, was from the NLT method. The third peptide, PLVIIAEDVDGEALSTLVLNRLK, was digested with the NG method and identified with an XCorr of 3.8 and mass accuracy of -0.67 ppm. All three of these peptides have XCorr values greater than 3, which indicates strong correlation between the mass spectra acquired in these experiments and the corresponding theoretical mass spectra. Furthermore, the mass accuracy associated with each of these peptides was also very high (within ± 2 ppm).

The combination of chemical and enzymatic digestion in the NLT method allowed the identification of 9,843 total peptides, corresponding to 2,120 proteins in the membrane protein-enriched sample. Among all proteins identified with the NLT method, 1,078 or 51 % were membrane proteins. Proteins were identified as membrane proteins through cellular component clustering analysis in the Database for Annotation, Visualization and Integrated Discovery (DAVID).[43,44] Just over half of the proteins identified were membrane proteins, which is due to the fact that complete separation of membrane and non-membrane proteins remains challenging.[45] The work presented here focused mainly on the comparison of digestion methods, not complete coverage of the membrane proteome.

3.2. Comparison of three digestion methods

Compared to the LT method, which only utilized enzymatic digestion to identify 7,982 peptides, 23% more peptides were identified with the NLT method, corresponding to 20% more proteins (1,764 proteins found with LT). In an aqueous digestion environment, many cleavage sites targeted by Lys-C and/or trypsin may not be accessible by these two proteases. However, an initial chemical digestion by NTCB cleaved proteins into several fragments, allowing Lys-C and/or trypsin to access the appropriate cleavage sites. Additionally, small molecules like NTCB can more easily access cleavage sites within folded membrane proteins.

The other combinatorial digestion method employing NTCB and Glu-C (NG) provided the fewest number of total peptides (3,307) and proteins (1,037). Glutamic acid and aspartic acid are less abundant in proteins and therefore, fewer Glu-C cleavage sites exist; as a result, Glu-C is not as effective as Lys-C and/or trypsin. Additionally, the digestion efficiency of Glu-C is not as

high as Lys-C and trypsin. These results are consistent with the fact that Lys-C and trypsin are much more frequently used in bottom-up proteomics.

In addition to total peptides and proteins identified, the number of membrane proteins identified in each experiment was also compared. The greatest number of membrane proteins, 1,078 proteins, was identified from the experiment that used the NLT method. Figure 3 shows the number of membrane proteins identified with each digestion method; a similar trend is seen among membrane proteins as total peptides and proteins.

3.3. Peptide and protein overlap among three digestion methods

The overlap between peptides and proteins identified with each method was also investigated. Figure 4a shows the overlap between unique peptides identified with each digestion method; 109 peptides were identified in all samples. There was very little overlap between peptides identified in samples digested with NG and either of the other methods (170 peptides between NG and LT, and 253 between NG and NLT), which is expected due to the different cleavage site specificities of each enzyme, particularly between NG and LT which have no overlapping cleavage sites. In contrast, there is significant peptide overlap between the LT and NLT experiments; a total of 2,391 unique peptides were identified with both methods. The protein overlap between all three samples is shown in Figure 4b, and the number of proteins identified with multiple methods is markedly higher. Peptides from the same proteins are expected to be identified in multiple experiments, even if the peptides differ in sequence. More than 70% of proteins identified in the NG experiment (773 of 1,037 proteins) were also identified in the NLT experiment, and 76% proteins identified with the LT method (1,341 of 1,764 proteins) were also found with the NLT

method. The overlap between membrane proteins identified in each method was also investigated, as shown in Figure 4c; 700 membrane proteins were identified in both LT and NLT experiments, and 329 proteins were found in all three experiments. A total of 1,308 membrane proteins were identified in this work.

3.4. Missed cleavage and peptide length distributions

Datasets were further analyzed to determine the number of missed cleavages associated with each digestion method (Figure 5a). Both methods utilizing Lys-C and trypsin resulted in the highest percentages of zero missed cleavages: 94% for LT and 92% for NLT. As shown in Figure 5a, NG resulted in the greatest percentage of peptides containing missed cleavages (80% compared to 6% and 8% for LT and NLT, respectively), which further demonstrate that NG was not a robust digestion method. Although the method combining NTCB, Lys-C and trypsin gave a slightly lower percentage of zero missed cleavage sites compared to the LT method, 23% more total peptides and 20% more proteins were identified, as discussed above. Overall, the method combining NTCB, Lys-C and trypsin provided the most effective digestion of membrane proteins for MS analysis.

The length of peptides generated from each digestion method was also investigated. Figure 5b shows the peptide length distribution for each digestion method. There is no significant difference between methods, except that the overall distribution of peptide length in the NG experiment includes a greater number of larger peptides. Because chemical digestion was always used in conjunction with enzymatic methods, the differences in peptide length typically seen between enzymatic and chemical cleavage could be compensated for with the

sequential enzymatic methods. Additionally, MS is biased towards a specific peptide length range (10-30 amino acids), and based on the parameters used in these experiments, peptides with too few or too many amino acid residues may not be effectively detected.

3.5. Membrane protein clustering

Membrane proteins identified in samples digested with the NLT method were further studied through clustering analysis. Proteins were clustered according to biological process and molecular function using DAVID[43,44] (Figure 6). Biological process clustering revealed that establishment of localization was most highly enriched with a *P*-value of 4.5E-127. Proteins related to membrane organization and oxidation reduction were also highly enriched, and it is well-known that many oxidation and reduction reactions occur among membrane proteins in the mitochondria. A number of proteins with functions corresponding to cell adhesion and cell motion were also enriched. Molecular functions such as substrate-specific transporter activity, oxidoreductase activity, and protein binding were highly enriched among membrane proteins identified, which is consistent with known functions of membrane proteins.

3.6. SNARE complex proteins

One specific group of membrane proteins were further investigated here. The soluble *N*-ethylmaleimide-sensitive fusion protein attachment protein receptor (SNARE) complex assists in the fusion of transport vesicles with their targeted membranes.[46] They are often found on various membranes throughout the cell, and are essential for intracellular membrane trafficking. Nine proteins in this complex were identified in this experiment: SNAP23, STX2, STX4, STX6, STX7, STX8, STX10, STX12, and VAMP2. For each protein, multiple unique peptides were

identified, and several examples are listed in Table 1; a complete list is presented in Supplemental Table 2. The greatest number of total peptides was identified with the NLT method (77 peptides), and notable overlap exists between the peptides identified with the LT and NLT methods. Fewer total peptides were identified with the NG method (22 peptides), but they complement those identified with either the LT or NLT method. For example, a total of seven unique peptides were identified in the protein STX2, among which two peptides were identified with both the NLT and LT methods, four were identified with only the NLT method and one was identified with only the NG method. The combination of chemical and enzymatic methods demonstrated to be effective to analyze SNARE complex proteins, which play critical roles in membrane trafficking.

4. Conclusion

Membrane proteins are extremely important in biological systems due to their involvement in a variety of cellular processes, including signal transduction, molecular transport, cell-cell communication and cell-environment interactions. Membrane proteins are notoriously difficult to analyze, even with powerful modern MS-based proteomics techniques, because of their hydrophobicity and overall low abundance. This work presents combinatorial methods incorporating chemical and enzymatic digestion to cleave proteins for MS analysis. Parallel experiments clearly demonstrated that the combination of NTCB with Lys-C and trypsin can provide 23% more total peptides and 20% more proteins identified than the common Lys-C and trypsin digestion method. Chemical methods utilizing small molecules can more easily access

cleavage sites within membrane proteins, compared to enzymatic methods which face steric hindrance. The combination of NTCB and Glu-C was not as effective, shown through the number of missed cleavages identified. Between the three digestion methods compared here, over 1,300 membrane proteins were identified. The combination of chemical and enzymatic methods demonstrated to be effective for membrane protein digestion, and further implementation of this method will allow the comprehensive and quantitative analysis of membrane proteins in complex biological samples.

Acknowledgements

This work is supported by Georgia Tech start-up funds. R.W is also supported by the Blanchard Assistant Professorship.

Figure Captions:

Fig. 1 Experimental procedure comparing three digestion methods for the comprehensive analysis of membrane proteins.

Fig. 2 Tandem mass spectra corresponding to peptides identified from the protein Hspd1 using digestion methods combining (a) Lys-C and trypsin, (b) NTCB, Lys-C and trypsin, and (c) NTCB and Glu-C.

Fig. 3 Number of total peptides, unique peptides, proteins and membrane proteins identified using each digestion method.

Fig. 4 Overlap between (a) peptides, (b) proteins, and (c) membrane proteins identified using each of the three digestion methods.

Fig. 5 (a) Number of missed cleavages among peptides identified with each digestion method; (b) Distribution of peptide length for each digestion method. (Red: Lys-C and trypsin; Blue: NTCB, Lys-C and trypsin; Yellow: NTCB and Glu-C).

Fig. 6 Clustering of membrane proteins identified in the NTCB, Lys-C and trypsin digestion sample according to (a) biological process and (b) molecular function.

Table 1 Examples of SNARE proteins and corresponding peptides identified in this work.

Gene Symbol	Peptide	LT	NLT	NG
SNAP23	A.LNIGNEIDAQNPQIK.R	✓	✓	
	E.IQQRAHQITDE.S			✓
	K.QPGPVTNGQLQQPTTGAASGGYIK.R	✓	✓	
	Q.RAHQITDESLESTR.R		✓	
	R.ILGLAIESQDAGIK.T	✓		
	R.ITNDAREDEMEENLTQVGSILGNLK.D		✓	
STX2	E.RSKGRIQRQLE.I			✓
	K.AIEQSFDQDESGNR.T		✓	
	K.NDDGDTVVVVEK.D	✓	✓	
	K.NHSIILSAPNPEGK.I		✓	
	K.PSIFTSDIISDSQITR.Q	✓	✓	
	M.FVETQGEMINNIER.N		✓	
STX4	E.KNILSSADYVE.R			✓
	E.VFVSNILKD.T			✓
	K.EEADENYNSVNTR.M	✓	✓	
	K.NILSSADYVER.G	✓		
	K.TQHGVLSQQFVELINK.C		✓	
STX6	D.MKDQMSTSSVQALAE.R			✓
	E.LLQDPSTATREE.I			✓
	K.AVNTAQGLFQR.W		✓	
	R.QALLGDSGSQNWSTGTTDK.Y	✓	✓	
STX7	A.NVENAEVHVQQANQQLSR.A	✓	✓	
	E.FTTSLTNFQKVQRQAAE.R			✓
	N.QLGTPQDSPELR.Q	✓	✓	
	R.LVAEFTTSLTNFQK.V	✓	✓	
	R.NLVSWESQTQPQVQVQDEEITEDDLR.L		✓	
	R.TLNQLGTPQDSPELR.Q	✓	✓	
STX8	D.ALSSIISRQKQMGQE.I			✓
	K.IIQEQDAGLDALSSIISR.Q		✓	
	R.GLGFDEIR.Q	✓	✓	
	R.QNLLDDLVTR.E		✓	
STX10	E.ANPGKFKLPAGDLQE.R			✓
	E.ILAGKPAAQKSPSDLLDASAVSATSRYIEE.Q			✓
	K.SPSDLLDASAVSATSR.Y	✓	✓	
	M.VSGSIQVLK.H	✓		
STX12	M.QSQEDEVAITEQDLELIK.E		✓	
	R.ISQATAQIK.N	✓	✓	
	R.LMNDFSAALNNFQAVQR.R		✓	
	R.NPGPSGPQLR.D	✓		
	R.QLEADILDVNQIFK.D		✓	
VAMP2	D.IMRVNVDKVLE.R			✓
	K.LSELDDRADALQAGASQFETSAAK.L		✓	
	R.ADALQAGASQFETSAAK.L	✓	✓	
	R.LQQTQAQVDEVVDIMR.V	✓	✓	
	T.AATAPPAAPAGEGGPPAPPPNLTSNR.R		✓	

Figures

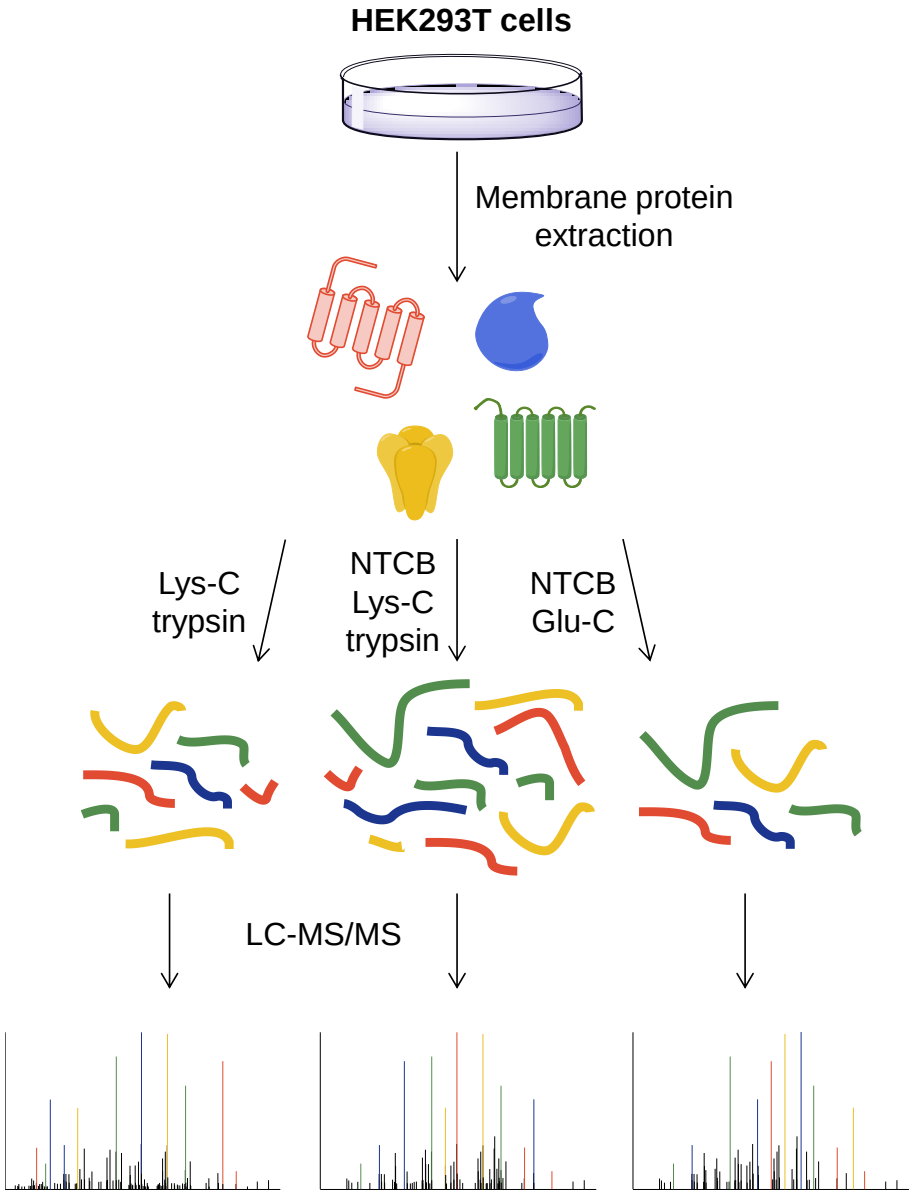


Fig. 1 Experimental procedure comparing three digestion methods for the comprehensive analysis of membrane proteins.

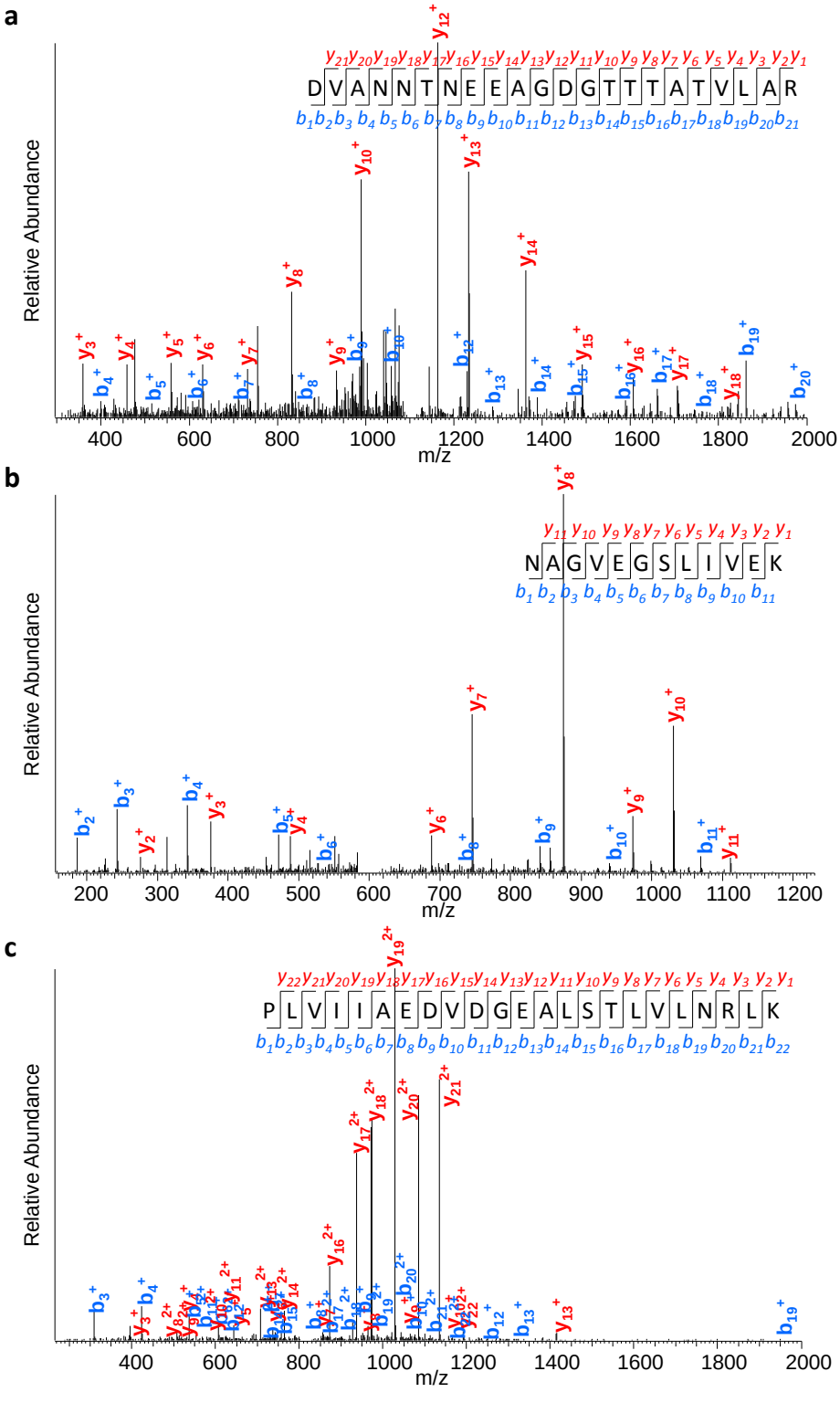


Fig. 2 Tandem mass spectra corresponding to peptides identified from the protein Hspd1 using digestion methods combining (a) Lys-C and trypsin, (b) NTCB, Lys-C and trypsin, and (c) NTCB and Glu-C.

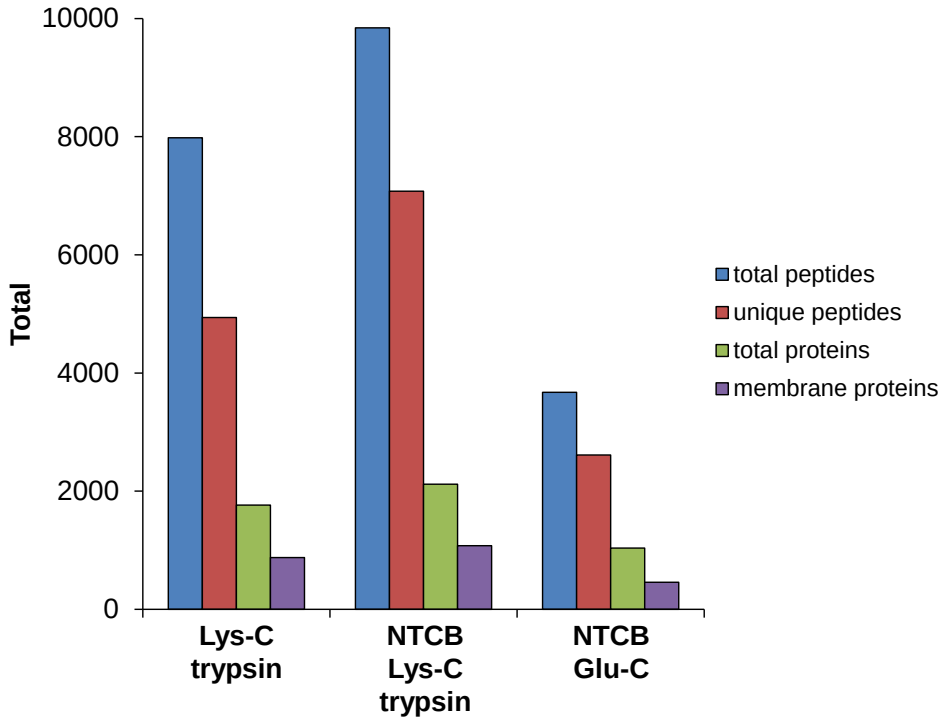


Fig. 3 Number of total peptides, unique peptides, proteins and membrane proteins identified using each digestion method.

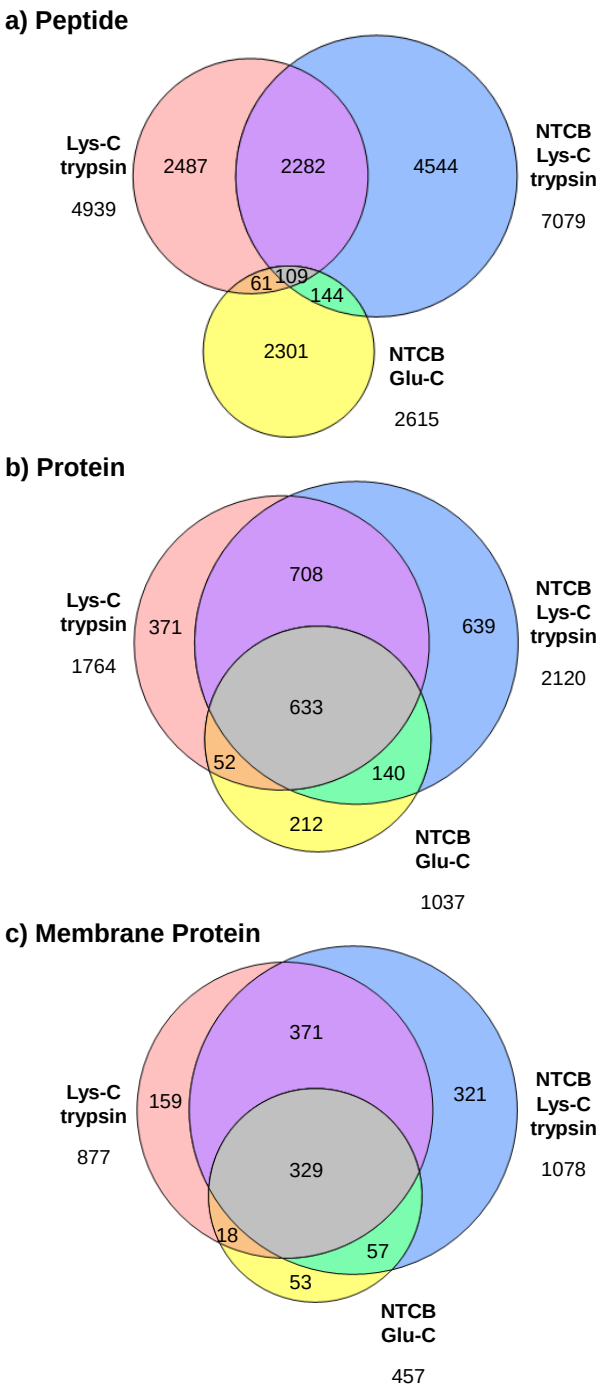


Fig. 4 Overlap between (a) unique peptides, (b) proteins, and (c) membrane proteins identified using each of the three digestion methods.

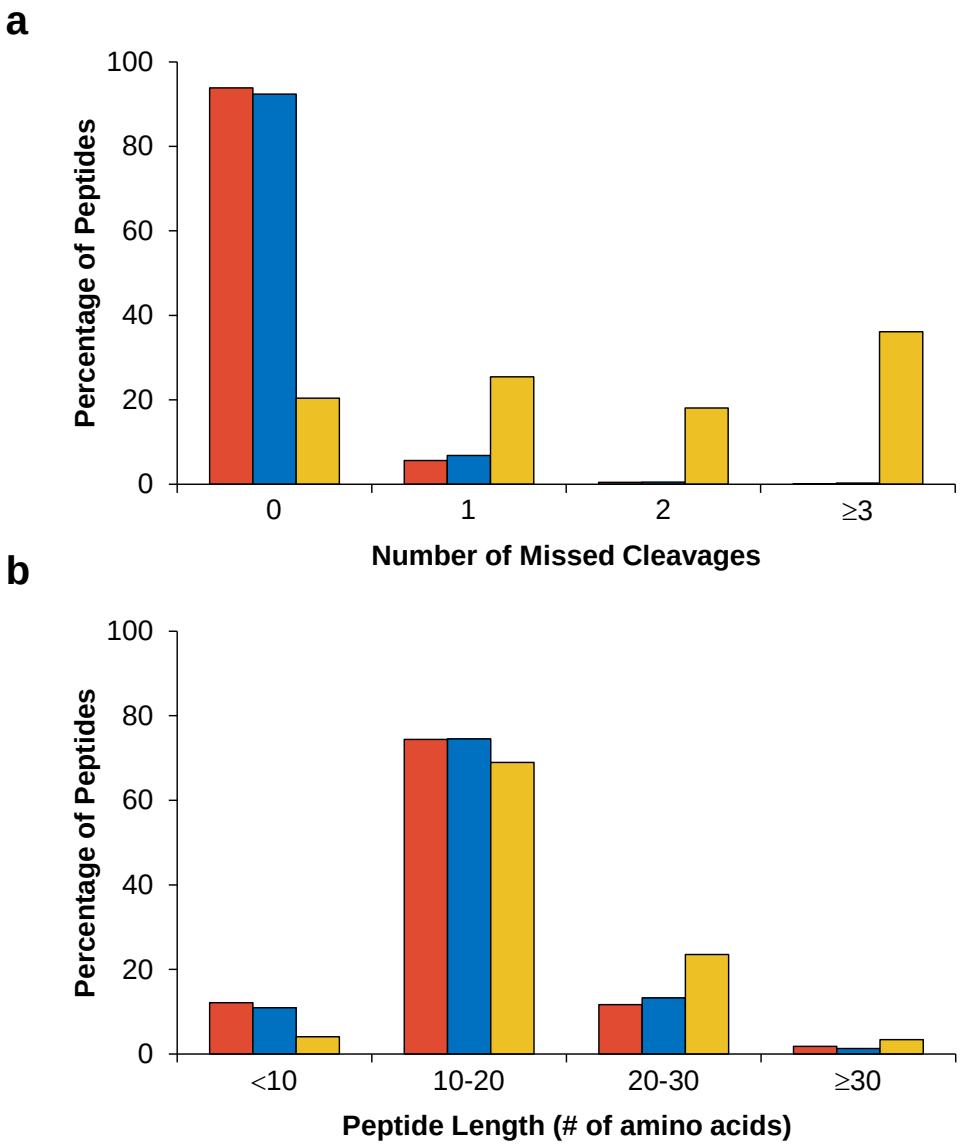


Fig. 5 (a) Number of missed cleavages among peptides identified with each digestion method; (b) Distribution of peptide length for each digestion method. (Red: Lys-C and trypsin; Blue: NTCB, Lys-C and trypsin; Yellow: NTCB and Glu-C).

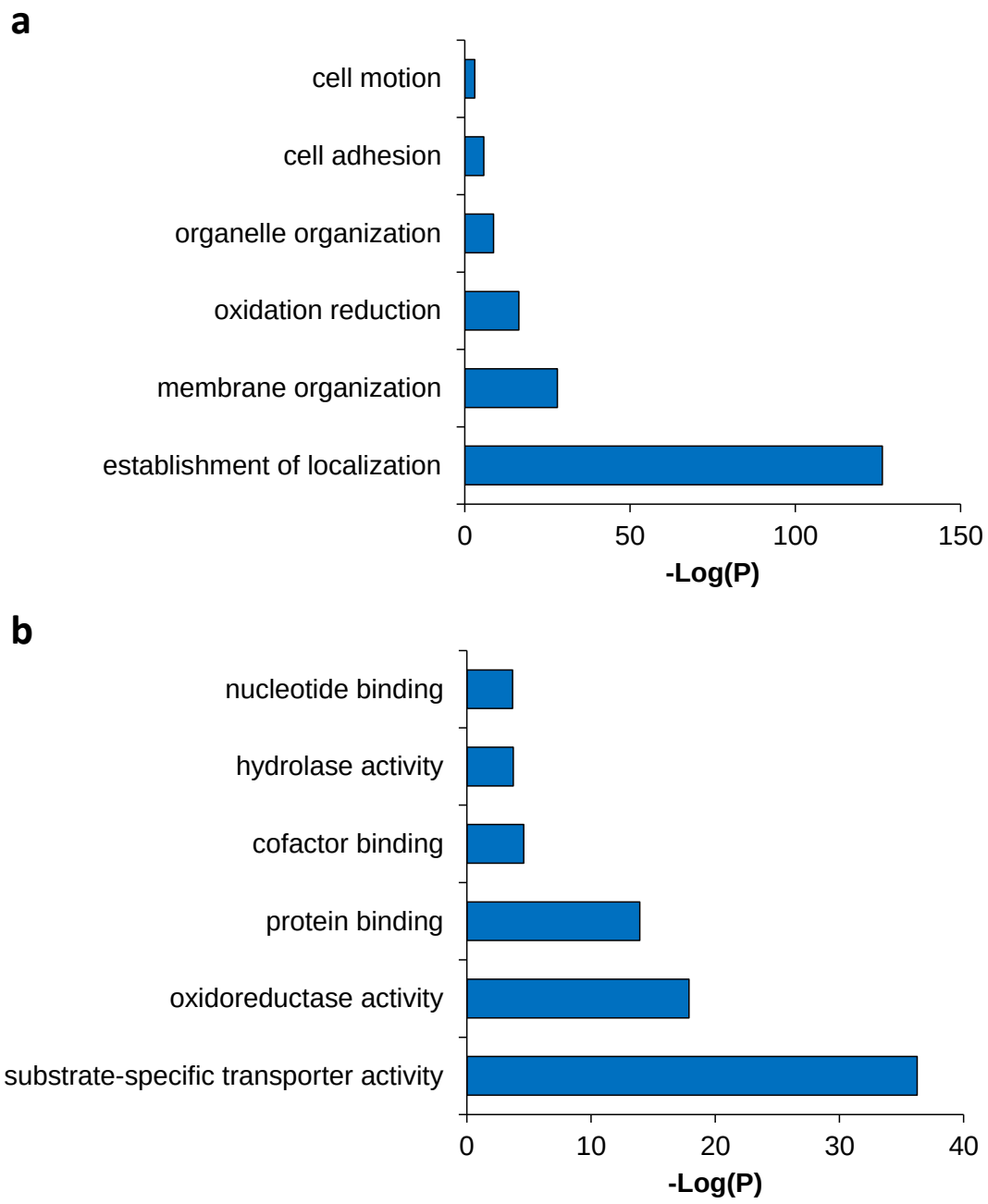


Fig. 6 Clustering of membrane proteins identified in the NTCB, Lys-C and trypsin digestion sample according to (a) biological process and (b) molecular function.

References

1 B. Alberts, Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. *Molecular Biology of the Cell*; Garland Science: New York, 2002.
2 J. M. Gilmore and M. P. Washburn, *J Proteomics*, 2010, **73**, 2078-2091.
3 J. K. Lee and R. M. Stroud, *Curr. Opin. Struct. Biol.*, 2010, **20**, 464-470.
4 K. R. Kampen, *J Membrane Biol*, 2011, **242**, 69-74.
5 B. A. Macher and T. Y. Yen, *Mol. Biosyst.*, 2007, **3**, 705-713.
6 D. Grimm, J. Bauer, J. Pietsch, M. Infanger, J. Eucker, C. Eilles and J. Schoenberger, *Curr Med Chem*, 2011, **18**, 176-190.
7 M. Rucevic, D. Hixson and D. Josic, *Electrophoresis*, 2011, **32**, 1549-1564.
8 E. Wallin and G. von Heijne, *Protein Sci*, 1998, **7**, 1029-1038.
9 A. L. Hopkins and C. R. Groom, *Nat Rev Drug Discov*, 2002, **1**, 727-730.
10 M. A. Yildirim, K. I. Goh, M. E. Cusick, A. L. Barabasi and M. Vidal, *Nat Biotechnol*, 2007, **25**, 1119-1126.
11 A. B. Weinglass, J. P. Whitelegge and H. R. Kaback, *Curr Opin Drug Disc*, 2004, **7**, 589-599.
12 L. A. Eichacker, B. Granvogl, O. Mirus, B. C. Muller, C. Miess and E. Schleiff, *J Biol Chem*, 2004, **279**, 50915-50922.
13 A. O. Helbig, A. J. R. Heck and M. Slijper, *J Proteomics*, 2010, **73**, 868-878.
14 X. Wei, C. Dulberger and L. J. Li, *Anal Chem*, 2010, **82**, 6329-6333.
15 C. C. Wu and J. R. Yates, *Nat Biotechnol*, 2003, **21**, 262-267.
16 J. N. Savas, B. D. Stein, C. C. Wu and J. R. Yates, *Trends Biochem.Sci.*, 2011, **36**, 388-396.
17 C. C. Wu, M. J. MacCoss, K. E. Howell and J. R. Yates, *Nat Biotechnol*, 2003, **21**, 532-538.
18 Y. X. Zhao, W. Zhang, Y. J. Kho and Y. M. Zhao, *Anal Chem*, 2004, **76**, 1817-1823.
19 T. Nilsson, M. Mann, R. Aebersold, J. R. Yates, A. Bairoch and J. J. M. Bergeron, *Nat Methods*, 2010, **7**, 681-685.
20 R. H. Wu, W. Haas, N. Dephoure, E. L. Huttlin, B. Zhai, M. E. Sowa and S. P. Gygi, *Nat Methods*, 2011, **8**, 677-683.
21 A. L. McCormack, D. M. Schieltz, B. Goode, S. Yang, G. Barnes, D. Drubin and J. R. Yates, *Anal Chem*, 1997, **69**, 767-776.
22 R. Aebersold and D. R. Goodlett, *Chem. Rev.*, 2001, **101**, 269-295.
23 L. L. Sun, A. S. Hebert, X. J. Yan, Y. M. Zhao, M. S. Westphall, M. J. P. Rush, G. J. Zhu, M. M. Champion, J. J. Coon and N. J. Dovichi, *Angew. Chem.-Int. Edit.*, 2014, **53**, 13931-13933.
24 N. Nagaraj, A. P. Lu, M. Mann and J. R. Wisniewski, *J Proteome Res*, 2008, **7**, 5028-5032.
25 J. V. Olsen, S. E. Ong and M. Mann, *Mol Cell Proteomics*, 2004, **3**, 608-614.
26 P. Brownridge and R. J. Beynon, *Methods*, 2011, **54**, 351-360.
27 R. Raijmakers, P. Neerincx, S. Mohammed and A. J. R. Heck, *Chem Commun*, 2010, **46**, 8827-8829.
28 G. R. Drapeau, J. Houmard and Y. Boily, *J Biol Chem*, 1972, **247**, 6720-&.
29 D. L. Crimmins, Mische, S. M. and Denslow, N. D. , *Current Protocols in Protein Science*, 2005.
30 H. Y. Tang and D. W. Speicher, *Anal Biochem*, 2004, **334**, 48-61.
31 G. R. Jacobson, M. H. Schaffer, G. R. Stark and T. C. Vanaman, *J Biol Chem*, 1973, **248**, 6583-6591.
32 W. X. Chen, J. M. Smeekens and R. H. Wu, *J Proteome Res*, 2014, **13**, 1466-1473.
33 W. X. Chen, J. M. Smeekens and R. H. Wu, *Mol Cell Proteomics*, 2014, **13**, 1563-1572.
34 J. M. Smeekens, W. X. Chen and R. H. Wu, *J Am Soc Mass Spectr*, 2015 (in press).
35 J. K. Eng, A. L. Mccormack and J. R. Yates, *J Am Soc Mass Spectr*, 1994, **5**, 976-989.
36 J. M. Peng, J. E. Elias, C. C. Thoreen, L. J. Licklider and S. P. Gygi, *J Proteome Res*, 2003, **2**, 43-50.
37 J. E. Elias and S. P. Gygi, *Nat Methods*, 2007, **4**, 207-214.
38 E. L. Huttlin, M. P. Jedrychowski, J. E. Elias, T. Goswami, R. Rad, S. A. Beausoleil, J. Villen, W. Haas, M. E. Sowa and S. P. Gygi, *Cell*, 2010, **143**, 1174-1189.

39 R. H. Wu, N. Dephoure, W. Haas, E. L. Huttlin, B. Zhai, M. E. Sowa and S. P. Gygi, *Mol Cell Proteomics*, 2011, **10**, 10.1074/mcp.M1111.009654.
40 X. Du, S. J. Callister, N. P. Manes, J. N. Adkins, R. A. Alexandridis, X. Zeng, J. H. Roh, W. E. Smith, T. J. Donohue, S. Kaplan, R. D. Smith and M. S. Lipton, *J Proteome Res*, 2008, **7**, 2595-2604.
41 L. Kall, J. D. Canterbury, J. Weston, W. S. Noble and M. J. MacCoss, *Nat Methods*, 2007, **4**, 923-925.
42 T. U. Consortium, *Nucleic Acids Res*, 2014, **42**, D191-D198.
43 D. W. Huang, B. T. Sherman and R. A. Lempicki, *Nat Protoc*, 2009, **4**, 44-57.
44 D. W. Huang, B. T. Sherman and R. A. Lempicki, *Nucleic Acids Res*, 2009, **37**, 1-13.
45 Y. Lin, K. B. Wang, Y. J. Yan, H. Y. Lin, B. Peng and Z. H. Liu, *J. Sep. Sci.*, 2013, **36**, 3026-3034.
46 A. T. Brunger, *Q. Rev. Biophys.*, 2005, **38**, 1-47.