Published in final edited form as:

Anal Chem. 2013 February 5; 85(3): 1880–1888. doi:10.1021/ac3031527.

Top Down Proteomics of Human Membrane Proteins from Enriched Mitochondrial Fractions

Adam D. Catherman, Mingxi Li, John C. Tran, Kenneth R. Durbin, Philip D. Compton, Bryan P. Early, Paul M. Thomas, and Neil L. Kelleher*

Departments of Chemistry and Molecular Biosciences, the Chemistry of Life Processes Institute, the Proteomics Center of Excellence, and the Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Evanston, Illinois, 60208, United States

Abstract

The interrogation of intact integral membrane proteins has long been a challenge for biological mass spectrometry. Here, we demonstrate the application of Top Down mass spectrometry to whole membrane proteins below 60 kDa with up to 8 transmembrane helices. Analysis of enriched mitochondrial membrane preparations from human cells yielded identification of 83 integral membrane proteins, along with 163 membrane-associated or soluble proteins, with a median q value of 3×10^{-10} . An analysis of matching fragment ions demonstrated that significantly more fragment ions were found within transmembrane domains than would be expected based upon the observed protein sequence. Forty-six proteins from the complexes of oxidative phosphorylation were identified which exemplifies the increasing ability of Top Down Proteomics to provide extensive coverage in a biological network.

Introduction

Top Down Proteomics (TDPs) is defined by the direct introduction of intact protein molecules into a mass spectrometer. Enabling mass measurement and gas-phase fragmentation of the whole protein, TDPs routinely provides full sequence coverage of protein forms arising from combinations of coding polymorphisms, alternative splicing, post-translational modifications (PTMs), and proteolysis. These favorable characteristics have made TDPs useful for the analysis of single proteins and simple mixtures.^{1–3} While significant, proteome-wide Top Down analyses have also been reported,^{4–6} the technological challenges of analyzing intact proteins relative to peptides has caused Top Down analyses to suffer in regards to proteome coverage, dynamic range, and throughput when compared to Bottom Up methods. While great advances in mass spectrometric instrumentation^{7–8} and data analysis methods^{9–11} accelerated the progress of TDPs, front-end separations remained the primary barrier to achieving high-throughput Top Down Proteomics. Recently, efficient separation of intact proteins has been realized through the use of Gel-Eluted Liquid Fraction Entrapment Electrophoresis (GELFrEE), which separates intact proteins according to molecular weight using a tube gel in a manner analogous to traditional SDS-PAGE slab gels.¹² Several studies have highlighted the impact of GELFrEE on the field of TDPs.^{13–14} GELFrEE technology has also been coupled to solution isoelectric focusing,¹⁵ creating a robust intact protein separation platform capable of identifying more than 1000 human proteins.¹⁶

*Corresponding author. Prof. Neil Kelleher. Phone: +1-847-467-4362 Fax: +1-847-467-3276, n-kelleher@northwestern.edu.

Despite the rapid increase in the rate of technology development for Top Down, the interrogation of integral membrane proteins (IMPs) has been neglected. These proteins comprise ~70% of known drug targets and are responsible for many important cellular functions including ion transport and intercellular communication.¹⁷ Although approximately 25% of the human genome encodes IMPs,¹⁸ they are often underrepresented in proteomic studies using either Top Down or Bottom Up methods due to their hydrophobic character and low cellular abundance.¹⁹

Previous studies by Whitelegge and colleagues have shown the ability of Top Down mass spectrometry to detect and identify integral membrane proteins.^{20–22} Much of the initial work focused on bacteriorhodopsin, a seven transmembrane helix-containing proton pump, as a membrane protein standard. The first intact FTMS spectrum of the protein was published in 2003, achieving an intact mass measurement within 8 ppm of the expected mass.²³ High mass accuracy MS/MS and identification of bacteriorhodopsin was also reported.²⁴ Significant increases in the coverage of bacteriorhodopsin were achieved more recently, along with the identification of the post-translational modifications of the eight subunits of cytochrome *b₆f* complex, and demonstration of the ability of a Top Down approach to detect and fragment β -barrel proteins.²⁵ The study of the large photosystem II complex from *Galderia sulphuraria* resulted in the confident identification of eleven integral membrane subunits and five peripheral subunits.²²

Primarily focused on single protein analysis or characterization of relatively simple mixtures isolated from plant chloroplasts, their work is expanded here towards whole proteins from enriched human mitochondrial membranes. Mitochondria are responsible for important cellular functions including the Krebs cycle, oxidative phosphorylation, fatty acid oxidation, and calcium regulation.²⁶ They are also highly involved in the intrinsic apoptotic pathway.²⁷ With well-honed protocols for organelle isolation,^{28–29} mitochondria have been the target of several large-scale Bottom Up studies.^{30–31} Previous Top Down studies of bovine mitochondrial membrane proteins utilized organic extraction followed by HILIC fractionation prior to intact mass measurement³² and fragmentation³³ using both triple quadrupole and Q-TOF instruments. Manual inspection of MS/MS spectra resulted in the identification of integral membrane proteins derived from 20 unique genes.

Here we report proof-of-concept for identification and characterization of proteins from enriched human mitochondrial membranes by GELFrEE coupled to LC-MS/MS. By identifying 246 proteins, including 83 that are integral to a biological membrane, this study suggests that TDPs may be well suited to high-throughput analyses of IMPs. Interestingly, when subjected to collisionally-induced dissociation, the transmembrane helices of IMPs fragment with high propensity relative to the soluble domains, often leading to hyper-confident identifications. Of the 246 total identifications, forty-six were subunits of oxidative phosphorylation complexes, yielding substantial coverage of this biological network. Further, many of these proteins were identified with post-translational modifications, including the addition of a myristoyl group, a trimethyl group, and the formation of pyroglutamate, highlighting the ability of TDPs in discovery mode to deliver detailed characterization of membrane proteins. In summary, the improving aptitude of Top Down Proteomics to detect, identify, and characterize membrane proteins on a large scale is illustrated herein.

Experimental Section

Preparation of Mitochondrial Membrane Proteins

HeLa S3 cells were grown in spinner flasks at a density of $1\text{--}3 \times 10^5$ cells/mL in Joklik's modified MEM supplemented with antibiotics and 10% newborn calf serum. Cells were

harvested by centrifugation at $200 \times g$ for 5 min., and washed twice with ice-cold phosphate buffer saline (PBS), centrifuging after each wash. The cell pellets, consisting of $\sim 2 \times 10^9$ cells, were suspended in 40 mL of sucrose-Tris buffer [250 mM sucrose, 10 mM Tris-HCl, pH 7.4, 0.1 mM EGTA, 10 mM sodium butyrate, 1 mM dithiothreitol, 1% protease and phosphatase inhibitors (Pierce)]. Cells were lysed using ~ 100 strokes of a Teflon homogenizer. The lysate was centrifuged at $100 \times g$ for 5 min. to pellet unbroken cells. The cell lysis procedure was repeated until only a small pellet remained. The cell lysate was centrifuged at $1000 \times g$ several times for 5 min. each to remove nuclei. The supernatant was centrifuged at $20000 \times g$ for 10 min., and the pellet was washed with sucrose-Tris buffer and centrifuged again. The pellet was resuspended in several milliliters of sucrose-Tris buffer, layered over a buffered Percoll solution (50% Percoll, 10 mM Tris-MOPS, pH 7.4, 0.1 mM EGTA, 10 mM sodium butyrate, 1 mM dithiothreitol, 1% protease and phosphatase inhibitors), and centrifuged at $45000 \times g$ for 2 h. The most abundant band was carefully collected with a pipette, diluted with sucrose-Tris buffer and centrifuged at $20000 \times g$ for 10 min. After washing and centrifugation to remove residual Percoll material, the pelleted mitochondria were resuspended in 2 mL lysis buffer (10 mM HEPES, pH 7.9, 10 mM sodium butyrate, 1 mM dithiothreitol, 1% protease and phosphatase inhibitors) and stirred at 4°C for 30 min. Six sonication pulses (8 s each) were used to lyse the mitochondria. The lysate was centrifuged at $20000 \times g$ for 10 min. to pellet the membrane. The pellet was resuspended in 1 mL of extraction buffer (15 mM Tris, 4% SDS, 10 mM sodium butyrate, 1 mM dithiothreitol, 1% protease and phosphatase inhibitors) aliquoted, flash frozen in liquid nitrogen, and stored at -80°C until use.

Western Blotting

Protein concentrations from whole cell lysate or subcellular fractions were quantified using the Bradford assay and 20 μg of protein from each sample was separated by SDS-PAGE. The membranes were probed with primary antibodies (1:1000; Cell Signaling) against α -tubulin, pyruvate dehydrogenase kinase 1 (PDK1), and cytochrome *c* oxidase subunit IV (COXIV) in 0.05% Tween 20/PBS and then with an HRP-labeled secondary antibody (1:5000; Santa Cruz). Proteins were visualized by chemiluminescence with ECL Western blotting detection reagents (Pierce).

GELFrEE Separation

Protein samples (400 μg) were precipitated with three volumes of acetone followed by suspension in 150 μL of loading buffer. Protein fractionation was performed using the GELFrEE 8100 Fractionation System (Expedeon) using either 10% or 12% cartridges. This fractionation was visualized on SDS-PAGE slab gels, loading 10 μL of each GELFrEE fraction followed by silver staining.

Liquid Chromatography-Mass Spectrometry

GELFrEE fractions were precipitated using MeOH/ CH_3Cl / H_2O as previously described for the removal of SDS.³⁴ The fractions were then resuspended in 20–30 μL of buffer A (95% H_2O , 5% acetonitrile, 0.2% formic acid) or 60% formic acid. Each sample (6.4 μL) was injected onto a 2 cm, 150- μm i.d. PLRP-S ($d_p = 5 \mu\text{m}$, pore size = 1000 $\text{\AA}$) trap column and washed at 5 $\mu\text{L}/\text{min}$ with buffer A using an Ultimate 3000 RSLCnano system (Thermo Scientific Dionex). A 10 cm, 75- μm i.d. PLRP-S column was used for separation. The gradient was delivered at 300 nL/min starting at 5% B (95% acetonitrile or 50/50 acetonitrile/isopropanol, 5% H_2O , and 0.2% formic acid) and rose to 20% B at 7 min, 65% B at 50 min, and 85% B at 58 min. The column was return to 5% B over 5 min. and re-equilibrated for 12 min.

Data were collected with a 12 Tesla (T) LTQ-Velos-FT-Ultra mass spectrometer (Thermo Fisher) fitted with a digitally controlled nanospray ionization source (PicoView DPV-550, New Objective.). MS¹ data were collected using either the FT-ICR (6 microscans, 180,000 resolving power at m/z 400) or the linear ion trap (25 microscans) with scan range of m/z 500–2000. Fragmentation was achieved using data-dependent CID or source-induced dissociation (SID).¹⁶ Data-dependent fragmentation was performed using a top 2 MS² acquisition strategy with 25 m/z isolation window and 6 microscans at 90,000 resolving power at m/z 400. Dynamic exclusion was enabled with a repeat count of 2, exclusion duration of 5000 s, and a repeat duration of 240 s. SID utilized a potential of 95 or 100 V and data was collected at 90000 resolving power, using 6 microscans and a scan range of 500–2000 m/z .

Data Analysis

Intact precursor and fragment masses from LC-MS/MS files were determined using in-house software (cRAWler), which utilizes THRASH for monoisotopic neutral masses from high-resolution precursor and fragment ion spectra and kDecon¹¹ to generate average masses from ion trap precursor data. The lists of precursor and fragment masses were analyzed with a 168-core computer cluster running a distributed version of ProSightPC 2.0 SP1 (Thermo Fisher Scientific). Data were processed according to the type of data acquisition: FT/FT CID, FT/FT SID, IT/FT SID, and SID (in which a precursor could not be determined). CID data were analyzed using a tiered, iterative absolute mass search logic that used a 2.3 Da window followed by 2000, 20000, and 100000 Da windows of MS¹ mass values. FT/FT SID data used 200, 2000, and 100000 windows. IT/FT SID used 2000, 20000, and 100000 Da. When a precursor mass could not be determined, the entire database was searched. The database used for windows of 2000 Da or less (homo_sapiens_2001_04_top_down_complex) was created by shotgun-annotation of known alternative splice forms, peptide cleavage events, and modifications to create 8,450,000 theoretical protein species from the UniProt release 2011_04. Searches with an intact mass window over 2000 Da used a simplified database consisting of 54,190 base sequences, modified only with N-terminal acetylation and initial methionine cleavage (where applicable) creating a total of approximately 160,000 forms (homo_sapiens_2011_04_top_down_simple). Databases are built regularly and can be downloaded from <ftp://prosightftp.gsX1gON@prosightpc.northwestern.edu>.

To estimate the false discovery rate (FDR), a q value evaluation approach was applied as previously described.¹⁶ A sequence-scrambled decoy database was built from the forward intact database.³⁵ All data were searched against both the forward and decoy databases separately using identical search parameters. All search hits were scored using a Poisson-based model⁹ and a posterior probability-based q value was calculated for each hit to estimate the FDR for each identification event.^{36–37}

TMHMM v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM>) was used for prediction of helical transmembrane domains of the identified proteins³⁸ based upon the sequence output of ProSight.

Results and Discussion

A block diagram of the methodology used for the analysis of intact membrane proteins is shown in Figure 1. In contrast to previous studies utilizing GELFrEE for Top Down analysis, this work emphasizes the sample and data flows tailored for the focused analysis of membrane proteins. In this case, intact mitochondria were enriched and their membrane proteins extracted prior to GELFrEE and LC-MS/MS. The use of differential centrifugation coupled to Percoll density gradient centrifugation enabled the concentration of

mitochondrial membrane proteins from HeLa S3 cells. The level of mitochondrial purity was assessed by Western blotting that indicated the enrichment of pyruvate dehydrogenase kinase 1 (PDK-1), a protein found in the mitochondrial matrix, and cytochrome *c* oxidase subunit 4 (COXIV), a marker of mitochondrial inner membranes (Fig. 1, top right). A 400 μ g portion of the mitochondrial membrane sample was separated according to molecular mass by GELFrEE. Visualization of the separation on a SDS-PAGE gel shows 12 well-resolved fractions below 75 kDa, demonstrating the extensibility of GELFrEE to membrane proteins (Fig. 1, lower right). While the use of SDS provides strongly solubilizing conditions for high recovery of IMPs, it must be removed prior to LC-MS/MS analysis. Here, detergent removal is accomplished by methanol/chloroform/water precipitation.³⁴ Following resuspension, the fractions were further separated and analyzed using online nanoLCMS/MS and the resulting precursor and fragmentation data searched using ProSightPC 2.0 SP1 to generate protein identifications. The number and location of helical transmembrane domains (TMDs) was determined using the TMHMM prediction software (Fig. 1, lower left), enabling a careful evaluation of the number of TMDs for the intact species that actually eluted from the nanoLC columns used in this study.

Figure 2 shows example identifications of a membrane associated protein and two IMPs (one with 1 and one with 4 TMDs) from GELFrEE fraction 3 following resuspension in buffer A. These three mitochondrial proteins (including transmembrane protein 14C which was detected with a relatively low S/N) were fragmented extensively, providing hyper-confident identifications (q value $<10^{-52}$), 100% sequence coverage and intact precursor masses with <5 ppm error. This example also demonstrates the ability of TDPs to routinely identify the correct protein form as each of these three proteins exhibited modifications beyond the DNA-predicted sequence. Transmembrane protein 14C was found to be N-terminally acetylated, cytochrome *c* oxidase polypeptide 7A2 was observed with a 23 residue signal sequence removed from the N-terminus, and ATP synthase subunit ϵ was identified in the Met-off form. The analysis of GELFrEE fraction 3 with a single injection resulted in 33 total unique protein identifications in which 14 were predicted integral membrane proteins. Analysis of 9 fractions from a single GELFrEE separation, resuspended in either Buffer A or 60% formic acid, resulted in a total of 90 unique identifications and 36 predicted IMPs.

Large Scale Identification of Membrane Proteins

Analysis of 30 GELFrEE fractions from samples enriched in mitochondria employed 44 LC-MS/MS injections to yield 246 unique identifications at a protein level FDR of 5%. At a 1% FDR, 219 proteins were identified corresponding to only an 11% decrease in unique identifications. The median q value for the identifications was 3×10^{-10} at 5% FDR and 1×10^{-11} at 1% FDR. These identifications were accomplished from three GELFrEE separations, utilizing less than 1.5 mg of total protein. Of the 246 identifications, 78 were predicted IMPs from TMHMM. Five additional proteins were annotated by UniProt to contain potential transmembrane helices. These proteins also had significant TMD probability based upon TMHMM, however below the cutoff for a clear helical prediction. With the addition of these five, a total of 83 integral membrane proteins were identified in this study at a 5% FDR and 76 at a 1% FDR. This represents ~34% of the total identified proteins. All 83 identified IMPs are listed in Table S-1.

The distribution of TMDs (Fig. S-1A) shows that proteins with up to 8 transmembrane helices were identified as intact species by nanoLC-MS/MS. Thirty-three of the integral membrane proteins were found to contain more than one TMD. Grand average of hydropathicity (GRAVY) scores were also determined for each of the identified protein sequences (Fig. S-1B). Forty-two total proteins exhibited GRAVY values greater than zero. The median GRAVY value was -0.48, while the IMP set alone had a median value of -0.07,

indicating that the membrane proteins identified contained significant hydrophilic regions. By analyzing the intact proteins, consisting of hydrophobic and hydrophilic regions, the overall protein solubility in acidic aqueous solutions is not always limiting, allowing for their analysis by the Top Down approach.

Determination of the subcellular localization of the identified proteins was also performed to assess the proteome coverage of the mitochondrion and the effectiveness of the subcellular fractionation protocol (Fig. S-2A). As expected, the 107 annotated mitochondrial proteins comprised the largest set of proteins by subcellular localization. Nuclear and cytosolic proteins were the largest contaminants, likely due to the large number of highly abundant proteins from these compartments. Of the 69 nuclear proteins, 27 were histones including proteoforms from each of the core histones as well as the linker histone H1. Fifty-one ribosomal proteins were also identified, eight from the mitochondrial ribosome, which along with histones, are often easily identified with TDPs due to their high cellular abundance and low molecular weight.¹³ While the purification did not yield pure mitochondrial membranes, it allowed for a large number of mitochondrial and integral membrane protein identifications as well as depletion of abundant proteins within other cellular components, as indicated by Western blotting (Figure 1). The use of subcellular fractionation as a method for increasing Top Down proteome coverage is also evident, as 80 proteins identified in this work were not identified in the 1043 proteins reported in a recent large-scale TDPs study.¹⁶ Additional work could be aimed at further increasing mitochondrial purity, perhaps relying on tissue samples where protocols are more refined and sample amounts are not usually limiting.

Concentrating on proteins with known mitochondrial localizations, the large majority, 66, were annotated as present in the inner membrane (Fig. S-2B). Ten proteins, of which seven were not annotated as inner membrane proteins, were predicted to be proteins of the mitochondrial outer membrane. These ten included four subunits of the translocase of the outer membrane (TOM complex). Additionally, cytochrome *c* oxidase assembly protein COX16 homolog and transmembrane protein 14C were additionally annotated as “mitochondrial membrane” but not specifically in the inner or outer membrane. Overall, 75 proteins were annotated as a membrane component within mitochondria, representing 16% of the proteins within the Swiss-Prot database containing the mitochondrial membrane annotation. Only nine proteins were annotated as matrix proteins, four of which were ATP synthase proteins (which also have annotations as mitochondrial inner membrane proteins).

A large number of the inner mitochondrial membrane proteins identified were found to be subunits of the large membrane protein complexes involved in oxidative phosphorylation. Eleven of the 15 subunits of ATP synthase were identified (Table S-2). Twenty-one of the ~45 subunits that comprised the 1 MDa NADH dehydrogenase complex were found. Nine of the 20 subunits from cytochrome *c* oxidase and five of the 11 subunits of cytochrome *bcl* complex were also identified. Other important proteins involved in oxidative phosphorylation were also identified: cytochrome *c*, cytochrome *c* oxidase assembly protein COX16 homolog, cytochrome *c* oxidase subunit 7A-related protein, and ATPase inhibitor. None of the four subunits of succinate dehydrogenase were identified.

Identification of Post-translational Modifications

Many of the proteins identified in this study exhibited post-translational modifications or sequence modifications that were not predicted by the DNA sequence. N-terminal processing, which is common in eukaryotic cells, was widely evident, as 58 of the 83 integral membrane proteins were identified with an N-terminal acetylation and/or cleavage of the initial Met of signal sequence (Table S1). Furthermore, myristoyl lipid modifications were detected on the N-terminal glycine residues of ADP-ribosylation factor 1 (P84077), coiled-coil-helix-coiled-coil-helix domain-containing protein 3 (Q9NX63), and NADH

dehydrogenase 1 beta subcomplex subunit 7 (P17568) (Fig. 3A). Lysine 43 of ATP synthase subunit c (P05496) was found to be trimethylated (Fig. 3B). Previous reports have detected this trimethylation at K43 in humans³⁹ as well as dogs⁴⁰ and cows⁴¹, yet the function of the modification is not yet fully understood. ATP synthase subunit c isolated from chloroplasts of *Arabidopsis thaliana* was found to not be trimethylated, although the protein shows little sequence homology with the mammalian mitochondrial form.⁴² The trimethylation found in this work had to be manually determined as *it was not present in the SwissProt database* and therefore was not shotgun annotated into the databases used for protein identification. Calculation of the mass differences between a dozen pairs of fragment ions resulted in an average mass difference 42.04 ± 0.01 Da (1σ), consistent with the previously reported trimethylation (+42.0470 Da) rather than acetylation (+42.0106 Da). Another significant finding revealed a -14 Da mass shift on the intact mass and 10 fragment ions of 6.8 kDa mitochondrial proteolipid (P56378) were consistent with the known I9V SNP. The formation of pyroglutamic acid following signal cleavage on the N-terminus of transmembrane protein 109, a four TMD-containing protein, was also determined as all 27 *b*-ions were found to match only upon the removal of 17.0265 Da from the N-terminus (Fig. 3C). This modification has not been reported previously, although the formation of pyroglutamic acid from N-terminal glutamine is known to occur on many proteins, including bacteriorhodopsin. Given the high mass accuracy used for the PTM determination and the relatively common occurrence of pyroglutamate, this assignment is considered confident although the N-terminal region was not extensively sequenced.

The putative uncharacterized protein encoded by LINC00116 (Q8NCU8), which is listed in UniProt as probable non-coding RNA (and to our knowledge has not been previously observed as a protein product) was confidently identified (Fig. 3D). The predicted protein product listed in the database was incorrect and Met-83 was found to be the actual initiator, giving rise to a 55 amino acid protein following initial methionine cleavage. This protein was detected with a monoisotopic neutral mass of 6391.60 Da, matching the expected mass of the product to 1 ppm. The resulting protein product has one expected TMD, which gave rise to the majority of the matching fragments.

Two IMPs were identified to have mis-annotated signal sequences within the ProSight database which relied on the predicted cleavage sites listed within SwissProt. Apolipoprotein O (Q9BUR5), containing a single TMD, was identified with an additional alanine-proline that was expected to be cleaved. Growth hormone-inducible transmembrane protein (Q9H3K2), which is a 6 TMD containing species, was confidently identified based on 15 matching *b*-ions and found to have an additional tyrosine cleaved from the signal sequence. The secreted protein UPF0723 protein C11orf83 (Q6UW78) was previously determined to undergo a 23 amino acid signal sequence cleavage.⁴³ We identified the pro-protein with 14 matching fragment ions derived from the proposed signal peptide sequence. A portion of the intact sequence was included in a TMD as predicted from TMHMM. It is possible that this protein is imbedded in a biological membrane and then cleaved upon a downstream secretion event.

TDPs is also routinely capable of enabling gene and isoform specific identifications which cannot often be made using Bottom Up approaches. For example, ER lumen protein retaining receptor 1 (P24390) and ER lumen protein retaining receptor 2 (P33947) isoform 1, which are coded by two different genes, were both unambiguously identified in this study (Fig. 4). These ~24 kDa proteins contain 4 TMDs according to TMHMM and 7 according to UniProt. Sequence alignment reveals that the proteins have an 83% identity and a 94% similarity. Three of the matching *y*-ions are also isobaric caused by a difference of LV to VI. The proteins were identified by source induced dissociation scans approximately 30 s apart

so the composite set of matching fragment ions are shown considering all fragmentation scans.

Fragmentation Propensities of Intact Integral Membrane Proteins

When mapping the predicted transmembrane domain regions onto graphical fragment maps, it appeared that many of the fragmentation events were occurring in these hydrophobic regions (Fig. 5A). In some cases, many of these proteins would not be identified above the FDR without matching fragment ions in these domains. Interestingly, a comparison of the q value distribution between the soluble/membrane associated and integral membrane proteins (Fig. 5B) demonstrates that integral membrane proteins were, on average, identified with a greater confidence. A median q value of 6×10^{-15} for the IMPs was obtained while the combined set of soluble and membrane associated proteins exhibited a less confident median q value of 1×10^{-8} . Considering only the proteins with an expected intact mass <25 kDa, the IMPs display a q value of 1×10^{-15} , once again lower than the 4×10^{-9} for the soluble/membrane associated subset. However, given the diversity in the identified proteins within each group as well as the subcellular fractionation enriching for membrane proteins, a number of factors could potentially explain this observation, including protein abundance, molecular weight distribution, and ionization efficiency.

To further explore these reasons behind the differences in fragmentation, the matching ion fragmentation data for the most confident identification of each identified integral membrane protein were localized to their predicted protein architecture to determine if gas-phase fragmentation was more or less likely to occur in these regions. TMHMM helix probability scores were determined individually at each individual backbone position cleaved during tandem mass spectrometry and as a composite across the protein. The average probability score for the matching fragment ions was then compared to the composite score (Fig. 5C). These results show that gas-phase fragmentation is significantly more likely to occur within transmembrane helices than can be predicted by random chance (two sample paired t-test: $p < 2 \times 10^{-16}$). Previous work has shown several examples in which this extensive fragmentation of transmembrane regions allows for manual protein sequencing and identification relying on these long sequence tags.³³ The hydrophobic transmembrane regions, largely devoid of charged residues as well as proline, do not have highly sequestered protons or weak amide bonds such as Asp-Pro, thus allowing for collisionally-induced fragmentation to occur more randomly along the backbone, likely giving rise to rich stretches of fragmentation and subsequently very confident identifications.. Previous study of bacteriorhodopsin displayed high sequence coverage from four of the seven transmembrane helices.²⁵ The same paper showed a larger number of matching fragment ions derived from the transmembrane regions of PetB and PetC compared to their soluble regions. However, when considering the entire study the authors concluded that matching fragment ions were not correlated with transmembrane domains, adding that some helices were appeared amenable to dissociation while others were resistant. The study of ATP synthase subunit c from *A. thaliana* revealed that all of the matching fragment ions were derived from the transmembrane helices, although IR activated ECD significantly increased the number of matching fragment ions.⁴² It will be of great interest to study the fragmentation patterns achieved using electron-based fragmentation for integral membrane proteins in the same high throughput manner as described above as well as investigating the role of tertiary structure in the gas phase as it relates to membrane protein fragmentation. Considering the number of IMP identifications and full fragmentation analysis achieved by absolute mass searches of shotgun annotated databases, the results in this work provide the most thorough examination of fragmentation of TMD-containing proteins to date, largely consistent with the results of previous Top Down studies. These results indicate that the transmembrane regions of integral membrane proteins should

significantly ease their identification if the proteins can be significantly solubilized before introduction into the mass spectrometer.

Conclusions

To our knowledge, this work represents the largest Top Down study targeting membrane proteins. The use of GELFrEE-LC-MS/MS has allowed for identification of 83 integral membrane proteins from a density gradient purified mitochondrial membrane sample. The data suggests that many hydrophobic proteins can be analyzed using the Top Down approach. Proteins with up to 8 transmembrane domains or those containing lipid modifications are amenable to this platform. Based upon the q values of the identified IMPs, it became clear that these proteins were often identified in high confidence relative to their soluble counterparts. Additionally, a greater number of the matching fragments ions were derived from the transmembrane helical regions than would be expected by random fragmentation. The work also demonstrates the importance of subcellular fractionation, which will provide Top Down Proteomics with an effective means to improve proteome coverage globally. With 48 proteins from the large complexes of oxidative phosphorylation and 107 total mitochondrial proteins identified, this work represents a large step forward in the use of TDPs for the study of isolated organelles. Oxidative phosphorylation and other mitochondrial pathways play important roles in antiproliferative states such as apoptosis and senescence. Additionally, the recent hypothesis that protein acetylation of mitochondrial proteins^{44–45} is a major regulatory mechanism of primary metabolism positions this work well to study the post-translational events underlying these important biological phenomena.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM067193. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additional support was provided by the UIUC Center for Neuroproteomics on Cell to Cell Signaling (P30 DA018310), the Chicago Biomedical Consortium with support from The Searle Funds at The Chicago Community Trust, and the Robert H. Lurie Comprehensive Cancer Center. AC would also like to acknowledge the ACS Division of Analytical Chemistry and the Society for Analytical Chemists of Pittsburgh (SACP) for their support.

References

1. Siuti N, Roth MJ, Mizzen CA, Kelleher NL, Pesavento JJ. *J. Proteome Res.* 2006; 5:233–239. [PubMed: 16457587]
2. Pesavento JJ, Bullock CR, Leduc RD, Mizzen CA, Kelleher NL. *J. Biol. Chem.* 2008; 283:14927–14937. [PubMed: 18381279]
3. Resemann A, Wunderlich D, Rothbauer U, Warscheid B, Leonhardt H, Fuchser J, Kuhlmann K, Suckau D. *Anal. Chem.* 2010; 82:3283–3292. [PubMed: 20329751]
4. Meng FY, Cargile BJ, Patrie SM, Johnson JR, McLoughlin SM, Kelleher NL. *Anal. Chem.* 2002; 74:2923–2929. [PubMed: 12141648]
5. Roth MJ, Forbes AJ, Boyne MT, Kim YB, Robinson DE, Kelleher NL. *Mol. Cell. Proteomics.* 2005; 4:1002–1008. [PubMed: 15863400]
6. Collier TS, Hawkrige AM, Georgianna DR, Payne GA, Muddiman DC. *Anal. Chem.* 2008; 80:4994–5001. [PubMed: 18512951]
7. Syka JEP, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, Ueberheide B, Garcia B, Busby S, Muratore T, Shabanowitz J, Hunt DF. *J. Proteome Res.* 2004; 3:621–626. [PubMed: 15253445]

8. Schaub TM, Hendrickson CL, Horning S, Quinn JP, Senko MW, Marshall AG. *Anal. Chem.* 2008; 80:3985–3990. [PubMed: 18465882]
9. Meng FY, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL. *Nat. Biotechnol.* 2001; 19:952–957. [PubMed: 11581661]
10. Zamdborg L, LeDuc RD, Glowacz KJ, Kim YB, Viswanathan V, Spaulding IT, Early BP, Bluhm EJ, Babai S, Kelleher NL. *Nucleic Acids Res.* 2007; 35:W701–W706. [PubMed: 17586823]
11. Durbin KR, Tran JC, Zamdborg L, Sweet SMM, Catherman AD, Lee JE, Li MX, Kellie JF, Kelleher NL. *Proteomics.* 2009; 10:3589–3597. [PubMed: 20848673]
12. Tran JC, Doucette AA. *Anal. Chem.* 2008; 80:1568–1573. [PubMed: 18229945]
13. Lee JE, Kellie JF, Tran JC, Tipton JD, Catherman AD, Thomas HM, Ahlf DR, Durbin KR, Vellaichamy A, Ntai I, Marshall AG, Kelleher NL. *J. Am. Soc. Mass Spectrom.* 2009; 20:2183–2191. [PubMed: 19747844]
14. Kellie JF, Catherman AD, Durbin KR, Tran JC, Tipton JD, Norris JL, Witkowski CE, Thomas PM, Kelleher NL. *Anal. Chem.* 2012; 84:209–215. [PubMed: 22103811]
15. Tran JC, Doucette AA. *Anal. Chem.* 2009; 81:6201–6209. [PubMed: 19572727]
16. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SMM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL. *Nature.* 2011; 480 254–U141.
17. Hopkins AL, Groom CR. *Nat. Rev. Drug Discov.* 2002; 1:727–730. [PubMed: 12209152]
18. Wallin E, von Heijne G. *Protein Sci.* 1998; 7:1029–1038. [PubMed: 9568909]
19. Wu CC, Yates JR. *Nat. Biotechnol.* 2003; 21:262–267. [PubMed: 12610573]
20. Whitelegge JP, Gundersen CB, Faull KF. *Protein Sci.* 1998; 7:1423–1430. [PubMed: 9655347]
21. Gomez SM, Nishio JN, Faull KF, Whitelegge JP. *Mol. Cell. Proteomics.* 2002; 1:46–59. [PubMed: 12096140]
22. Thangaraj B, Ryan CM, Souda P, Krause K, Faull KF, Weber APM, Fromme P, Whitelegge JP. *Proteomics.* 2010; 10:3644–3656. [PubMed: 20845333]
23. Whitelegge JP. *Photosynth. Res.* 2003; 78:265–277. [PubMed: 16245055]
24. Whitelegge J, Halgand F, Souda P, Zabrouskov V. *Expert Rev. Proteomics.* 2006; 3:585–596. [PubMed: 17181473]
25. Ryan CM, Souda P, Bassilian S, Ujwal R, Zhang J, Abramson J, Ping PP, Durazo A, Bowie JU, Hasan SS, Baniulis D, Cramer WA, Faull KF, Whitelegge JP. *Mol. Cell. Proteomics.* 2010; 9:791–803. [PubMed: 20093275]
26. Scheffler IE. *Mitochondrion.* 2001; 1:3–31. [PubMed: 16120266]
27. Green DR, Kroemer G. *Science.* 2004; 305:626–629. [PubMed: 15286356]
28. Frezza C, Cipolat S, Scorrano L. *Nat. Protoc.* 2007; 2:287–295. [PubMed: 17406588]
29. Cox B, Emili A. *Nat. Protoc.* 2006; 1:1872–1878. [PubMed: 17487171]
30. Forner F, Foster LJ, Campanaro S, Valle G, Mann M. *Mol. Cell. Proteomics.* 2006; 5:608–619. [PubMed: 16415296]
31. Sickmann A, Reinders J, Wagner Y, Joppich C, Zahedi R, Meyer HE, Schonfisch B, Perschil I, Chacinska A, Guiard B, Rehling P, Pfanner N, Meisinger C. *Proc. Natl. Acad. Sci. U. S. A.* 2003; 100:13207–13212. [PubMed: 14576278]
32. Carroll J, Fearnley IM, Walker JE. *Proc. Natl. Acad. Sci. U. S. A.* 2006; 103:16170–16175. [PubMed: 17060615]
33. Carroll J, Altman MC, Fearnley IM, Walker JE. *Proc. Natl. Acad. Sci. U. S. A.* 2007; 104:14330–14335. [PubMed: 17720804]
34. Wessel D, Flugge UI. *Anal. Biochem.* 1984; 138:141–143. [PubMed: 6731838]
35. Elias JE, Gygi SP. *Nat. Methods.* 2007; 4:207–214. [PubMed: 17327847]
36. Benjamini Y, Hochberg Y. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 1995; 57:289–300.
37. Storey JD, Tibshirani R. *Proc. Natl. Acad. Sci. U. S. A.* 2003; 100:9440–9445. [PubMed: 12883005]
38. Sonnhammer EL, von Heijne G, Krogh A. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1998; 6:175–182. [PubMed: 9783223]

39. Katz ML, Gao CL, Tompkins JA, Bronson RT, Chin DT. *Biochem. J.* 1995; 310:887–892. [PubMed: 7575423]
40. Katz ML, Christianson JS, Norbury NE, Gao CL, Siakotos AN, Koppang N. *J. Biol. Chem.* 1994; 269:9906–9911. [PubMed: 8144584]
41. Chen RM, Fearnley IM, Palmer DN, Walker JE. *J. Biol. Chem.* 2004; 279:21883–21887. [PubMed: 15010464]
42. Zabrouskov V, Whitelegge JP. *Journal of Proteome Research.* 2007; 6:2205–2210. [PubMed: 17441748]
43. Zhang ZM, Henzel WJ. *Protein Sci.* 2004; 13:2819–2824. [PubMed: 15340161]
44. Kim SC, Sprung R, Chen Y, Xu YD, Ball H, Pei JM, Cheng TL, Kho Y, Xiao H, Xiao L, Grishin NV, White M, Yang XJ, Zhao YM. *Mol. Cell.* 2006; 23:607–618. [PubMed: 16916647]
45. Schwer B, Eckersdorff M, Li Y, Silva JC, Fermin D, Kurtev MV, Giallourakis C, Comb MJ, Alt FW, Lombard DB. *Aging Cell.* 2009; 8:604–606. [PubMed: 19594485]

Figure 1. Platform for high-throughput Top Down Proteomics of mitochondrial membrane proteins

Cultured HeLa cells are subjected to subcellular fractionation to yield enriched mitochondria which are further fractionated to yield mitochondrial membrane proteins. These proteins are then separated according to molecular weight by GELFrEE, yielding 12 fractions. Each fraction is injected onto a nanoLC column and intact mass measurement and fragmentation spectra are collected with a 12 Tesla LTQ Velos FT Ultra. ProSightPC is used for identification prior to analysis of transmembrane domain content.

Figure 2. Representative data for the LC-MS/MS analysis of a single GELFrEE fraction containing <15 kDa proteins

The total ion chromatogram (TIC) shows the separation of proteins according to hydrophobicity, demonstrating the orthogonal nature of GELFrEE and RPLC. Intact spectra for three proteins are shown which allow for monoisotopic intact masses to be determined within 5 ppm of the expected molecular weight. Collisionally-induced fragmentation of each of the three proteins resulted in highly confident identifications, each demonstrating 100% sequence coverage using 10 ppm mass accuracy on the fragment ions. ATP synthase subunit epsilon was identified as the Met off form, cytochrome *c* oxidase polypeptide 7A2 with a 23

residue transit peptide cleavage, and transmembrane protein 14C with N-terminal acetylation. The TMHMM predicted transmembrane domain helices are displayed in red.

Figure 3. Identification of post-translational modifications and a mis-annotated start site on a newly observed human protein

A, Myristoylation identified on the N-terminus of the methionine cleaved form of NADH dehydrogenase 1 beta subcomplex subunit 7. *B*, Lysine 43 trimethylation was identified on ATP synthase subunit c following signal sequence cleavage. *C*, Signal sequence cleavage also occurs on transmembrane protein 109, leaving an N-terminal glutamine which cyclizes to a pyroglutamate. *D*, First evidence of a protein from a gene expected to be non-coding was identified as a 55 amino acid, single TMD containing product.

ER Lumen Protein Retaining Receptor 1

-M-N-L-F-R-F-L-G-D-L-S-H-L-L-A-I-I-L-L-L-L-K-I-W-K-S-R-S-C-A-G-I-S-G-I-K-S-Q-V-L-F-A-V-V-F-T-
-A-R-Y-L-D-L-F-//D-L-I-A-I-V-A-G-L-V-Q-T-T-V-L-Y-C-D-F-F-Y-L-Y-I-T-K-V-L-K-G-K-K-L-S-L-P-A-

Identical

Isobaric

Unique

ER Lumen Protein Retaining Receptor 2

-M-N-I-F-R-L-T-G-D-L-S-H-L-A-A-I-V-I-L-L-L-K-I-W-K-T-R-S-C-A-G-I-S-G-K-S-Q-L-L-F-A-L-V-F-T-
-T-R-Y-L-D-L-F-//D-L-I-A-V-V-A-G-V-Q-T-T-I-L-Y-C-D-F-F-Y-L-Y-I-T-K-V-L-K-G-K-K-L-S-L-P-A-

Figure 4. A selected example of gene-specific identifications illustrate that this is routine in Top Down Proteomics

ER lumen protein retaining receptor 1 and 2 were unambiguously differentiated from one another. Although many of the matching fragments were identical or isobaric between the two species, several unique ions reveal each individual identification.

Figure 5. Examination of the propensity for transmembrane helices to fragment in the gas phase
 A) Graphical fragment maps for two identified proteins demonstrating the propensity of transmembrane domain helices (shown in red) to fragment relative to the soluble regions. Also displayed is the average TMHMM value for all of the corresponding backbone positions as well as the average value for the entire protein. B) Distribution of lowest obtained q values for each of the identified proteins separated as predicted integral membrane and soluble/membrane associated proteins. The whiskers represent the 10th/90th percentiles. The distribution demonstrates that integral membrane proteins were, on average, identified with a greater confidence than the soluble proteins. C) Scatter plot displaying the TMHMM scores for fragment ions and entire proteins (as in A) where each points represents

a unique protein. Integral membrane proteins are shown as blue circles and soluble/membrane associated proteins in red circles. The $y=x$ line is also plotted to demonstrate where the data points would lie if the fragmentation was truly random. Points above this line indicate that the fragments were more likely to be derived from TMDs.