Published in final edited form as:

Methods. 2011 December ; 55(4): 330–336. doi:10.1016/j.ymeth.2011.09.019.

Profiling of integral membrane proteins and their post translational modifications using high-resolution mass spectrometry

Puneet Souda¹, Christopher M. Ryan¹, William A. Cramer², and Julian Whitelegge^{1,*}

¹The Pasarow Mass Spectrometry Laboratory, NPI-Semel Institute, David Geffen School of Medicine, University of California Los Angeles, California, USA

²Department of Biological Sciences, Purdue University, Indiana, USA

Abstract

Integral membrane proteins pose challenges to traditional proteomics approaches due to unique physicochemical properties including hydrophobic transmembrane domains that limit solubility in aqueous solvents. A well resolved intact protein molecular mass profile defines a protein's native covalent state including post-translational modifications, and is thus a vital measurement toward full structure determination. Both soluble loop regions and transmembrane regions potentially contain post-translational modifications that must be characterized if the covalent primary structure of a membrane protein is to be defined. This goal has been achieved using electrospray-ionization mass spectrometry (ESI-MS) with low-resolution mass analyzers for intact protein profiling, and high-resolution instruments for top-down experiments, toward complete covalent primary structure information. In top-down, the intact protein profile is supplemented by gas-phase fragmentation of the intact protein, including its transmembrane regions, using collisionally activated and/or electroncapture dissociation (CAD/ECD) to yield sequence-dependent high-resolution MS information. Dedicated liquid chromatography systems with aqueous/organic solvent mixtures were developed allowing us to demonstrate that polytopic integral membrane proteins are amenable to ESI-MS analysis, including top-down measurements. Covalent post-translational modifications are localized regardless of their position in transmembrane domains. Top-down measurements provide a more detail oriented high-resolution description of post-transcriptional and post-translational diversity for enhanced understanding beyond genomic translation.

Keywords

Integral membrane proteins; top-down mass spectrometry; membrane protein complexes; intact protein mass spectrometry; high-resolution mass spectrometry

© 2011 Elsevier Inc. All rights reserved.

*Corresponding author: Dr. Julian Whitelegge, Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, Tel: 310-794-5156; jpw@chem.ucla.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Membrane proteins are high value targets for over half of all marketed drugs and represent 20 – 30% of all coded proteins in sequenced genomes making them important for both structure determination and mass spectrometric characterization. Both transmembrane and loop regions may contain posttranslational modifications of both functional and structural significance, and must be well understood if we are to collectively define the native covalent state of membrane proteins [1]. Mass spectrometry can be used to obtain sequence identification, deliver molecular mass profiles and define post-translational modifications (PTMs), for both soluble and membrane proteins.

Bottom-up mass spectrometry techniques involve approaches where the intact protein is enzymatically cleaved to peptides before measurements via tandem mass spectrometry. Liquid chromatography with tandem mass spectrometry (LCMSMS) is one of the most common workflows employed for separation and identification of peptides. Tandem mass spectrometry data includes both parent ion and product ion fragment masses, and are frequently good enough to assign sequence identity to short peptides (10 – 30 residues) based on comparison to translated gene sequences. Though progress has been made with technical improvements in digestion and chromatography, sequence coverage can still be marginal, and this is especially true for the transmembrane domains of integral membrane proteins [2,3]. Typically, a handful of easily recovered peptides known as ‘proteotypic’ peptides are routinely observed in these tandem mass spectrometry experiments such that bottom-up approaches are typically biased with incomplete sequence coverage and PTM information [4]. Integral membrane proteins are not ideally suited for bottom-up proteomics due to their unique physiochemical properties, yielding some peptides with poor solubility and/or ability to be ionized, especially from transmembrane domains. Another caveat with bottom-up approaches is that they are heavily dependent on underlying genomic information thus ignoring molecular heterogeneity not immediately predictable from gene sequences.

Top-down mass spectrometry addresses many of the problems of the bottom-up approach by targeting intact proteins rather than peptides for analysis. The goal is to define a protein’s primary structure by providing highly accurate structural assignment of fragments. High-resolution Fourier-transform mass spectrometry (FT-MS) is most frequently used for top-down measurements due to the need to accurately assign product ions [5–7]. The whole intact protein can be dissociated using multiple dissociation mechanisms including CAD or ECD toward full sequence and PTM coverage. Complete interrogation of the primary structure via top-down mass spectrometry usually requires larger quantities of proteins than bottom-up experiments, and is thus well suited for protein crystallography experiments where both purity and abundance are typically attained prior to MS analysis. Much progress has been made in top-down MS as proteins of increasing size and complexity are being resolved [1,8–11]. Aqueous conditions suitable for mass spectrometry of soluble proteins are often inadequate for integral membrane proteins requiring specialized sample preparation and chromatography protocols, which we will discuss presently.

In conclusion, the bottom-up approach is suitable if an overall picture of a complex proteome is required, while top-down offers more valuable information if PTMs, protein heterogeneity and complete information about the primary structure is desired.

2. General considerations

The challenges associated with proteomics of membrane proteins arises due to their amphipathicity, a combination of polar soluble domains and apolar transmembrane domains, complicated by the presence of free thiols in the bilayer [8]. The coupling of ESI to MS has turned out to be an essential breakthrough in intact protein analysis by mass spectrometry [12].

ESI is preferred over MALDI as it produces multiply charged intact protein ions that dissociate with high efficiency for information-rich spectra that can be analyzed to deduce the protein sequence and PTMs. Liquid chromatography is easily interfaced with electrospray-ionization sources yielding a versatile, robust analytical platform for protein and peptide mass spectrometry.

2.1 LC-MS+ approach and solvent systems for integral membrane proteins

Integral membrane proteins were first analyzed by MALDI-TOF in 1992 and ESI in 1993 [13–15]. In 1998 we successfully used high formic acid concentrations with liquid chromatography and demonstrated that integral membrane proteins could be analyzed with mass accuracy similar to that achievable for soluble proteins [9]. However, high concentrations of formic acid could also lead to sporadic and unpredictable problems associated with protein formylation (+28 Da adducts). In newer and more improved approaches a high concentration of formic acid (up to 90%) is still preferred owing to its unrivaled capability to solvate proteins, but to reduce adducts, formic acid is introduced just seconds (< 120s) before mass spectrometry analysis. Tri-fluoro acetic acid (TFA) also has excellent solubilizing properties but routinely suppresses electrospray ionization, adds +114 Da adducts to proteins as well as presenting safety issues.

In order to obtain intact protein profiles such as the one shown (Fig. 1), a methodology known as LC-MS+ was developed. LC-MS+ refers to liquid chromatography with mass spectrometry and concomitant fraction collection. The technique employs a flow splitter between the HPLC and a low-resolution electrospray-ionization mass spectrometer so that half of the column eluent is diverted to collect fractions that can be used later for downstream experiments involving protein identification and PTM characterization on high-resolution Fourier transform mass spectrometers (FT-MS), if such a detailed analysis is required. Mass data from the initial LC-MS+ experiment is used to guide the subsequent top-down experiments. The LC-MS+ protocol is limited by the complexity of the protein sample and the capacity of the separations used. Both size-exclusion and reversed phase chromatography are used in the LC-MS+ protocol, depending upon sample complexity (Fig. 2). We have successfully used our size-exclusion LC-MS+ protocol to analyze a wide range of integral membrane proteins containing up to fifteen transmembrane helices [1,9,16,17] in circumstances where a single protein or a modest mixture were available after prior fractionation. LC-MS+ has also been widely applied using a reversed-phase protocol involving volatile aqueous/organic solvent mixtures compatible with membrane protein solubility and efficient ESI-MS.

Low-resolution mass spectra from LC-MS+ are deconvoluted to obtain an intact protein molecular mass profile (Fig.1) typically achieving 0.01 % (100 ppm) mass accuracy. This experiment nicely shows the extent of molecular heterogeneity of the bovine lens major intrinsic protein (MIP), an integral membrane protein of the aquaporin family. While the ionization efficiency for each molecular species could be different, this is generally not the case for intact proteins such that while the profile is semi-quantitative it can be assumed to reflect the natural heterogeneity of the preparation measured. In this example, the measured mass of the MIP protein (28,225.6 Da) was within agreement of the mass calculated for the translated gene product with no further processing involved (28,225.2 Da), a highly unusual observation for a eukaryotic protein which nearly always have PTMs that result in their measured masses being different from those calculated from the genomic translation. The observed molecular heterogeneity of this protein can be explained within experimental errors as phosphorylation (80 Da) at 28303.8 Da, cysteinylolation (119 Da) at 12343.3 Da, and additional minor C-terminal processing. The readout provided by an intact protein molecular mass profile is an important piece of information in guiding the structure determination process, in this case suggesting the use of phosphatase/reductant in order to minimize

molecular heterogeneity. A variety of integral membrane protein preparations, including the *E. coli* lactose permease and the thylakoid cytochrome *b₆f* complex, were characterized using LC-MS+ based approaches prior to successful crystallization.

Bacteriorhodopsin is a 27 kDa integral membrane protein from *Halobacterium halobium* that has a retinal chromophore for light-driven proton translocation across the membrane. Bacteriorhodopsin holoprotein was first analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) [14], and subsequently by ESI-MS with a measured mass within 0.01% of the calculated theoretical value [9]. The retinal chromophore is susceptible to hydrolysis under acidic conditions and a preliminary analysis of the apoprotein yielded just five *b*- and seven *y*-ions in what believe to be the first top-down FT-MS experiment on a polytopic integral membrane protein [8,9,18]. With improved size-exclusion chromatography and online high-resolution FT-MS data, we successfully performed top-down experiments on the holoprotein [19]. Peak parking (section 3.3.3) was used to maximize the time available for data acquisition on CAD product ions, before complete hydrolysis of the cofactor. Fig. 3(B) shows the mass spectrum obtained after reconstruction of the zero-charge molecular mass profile, clearly showing the holo- and apoforms of the protein separated by the mass of retinal (266.20 Da) [20]. By pooling data from CAD experiments on six different precursor charge states we successfully matched 67 *b*- and 55 *y*-ions, resulting in coverage of 79 of 247 peptide bonds (32%). The presence of numerous overlapping *b*- and *y*-product ions confirms full sequence coverage (Fig. 4), in agreement with the genomic translation. Many of the product ions resulted from dissociation in transmembrane domains and the covalent retinal modification was localized to a stretch of 22 amino acid residues (225 – 248) containing a single Lys residue (Lys216), which is the known cofactor site.

2.2 Identification of intact proteins via top-down sequence tag analysis

Protein identification of top-down datasets is done via sequence tags [21]. Both commercial and open-source tools are available in which a raw spectrum is deconvoluted to a zero-charge profile and a monoisotopic peak list is exported. Since the tandem mass spectra of intact proteins are very complex, grouping the peaks into isotopomer envelopes is a key initial stage for their interpretation. ProSight software (see section 3.3.6) is most commonly used for top-down data analysis [22], with deconvolution algorithms based on Thrash or Xtract [23]. Development of newer algorithms and improvements on Thrash and Xtract have led to evolution of MS-Deconv, which has a unique advantage in that it scores sets of envelopes rather than individual envelopes [24,25]. It is also important to keep in mind the end point of the top-down experiment since datasets are usually not complete to the extent that every bond in the structure can be confirmed. Manual data interpretation combined with software tools is an iterative process that continues to evolve with improved software packages for top-down data analysis.

2.3 Mass spectrometry of integral membrane protein complexes

Large membrane protein complexes are also amenable to LC-MS+ and top down mass spectrometry provided the separation is not overwhelmed [11,26,27]. By combining reverse-phase LC-MS+ with high-resolution FT-MS we characterized eleven integral and five peripheral subunits of the 750 kDa photosynthetic photosystem II complex from the eukaryotic red alga, *Galdieria sulphuraria* [26]. Analyses such as this one are very much geared at defining the different subunits of a complex, with protocols that eliminate any non-covalent interactions. However, if an integral membrane complex can be analyzed in its native state, with non-covalent interactions preserved, then it is possible to use the top-down approach toward stoichiometry measurements that are less well defined in LC-MS+. An early success in this arena used laser-induced liquid bead ion desorption (LILBID) to

analyze integral membrane protein complexes and their subunits. An IR-laser is used to desorb membrane protein micelles from aqueous micro-droplets prior to time-of-flight MS, typically in negative ion mode [28]. LILBID was the first technique to allow gas-phase measurement of ATP synthase c-ring stoichiometry establishing a new paradigm for the field [29]. ESI-MS of membrane protein micelles was first reported by Robinson and coworkers in 2004 [30] and they more recently showed that n-dodecyl-beta-d-maltoside micellar solution could be successfully used as a detergent for the heteromeric adenosine 5'-triphosphate (ATP)-binding cassette transporter complex BtuC2D2 maintaining the intact complex in the gas phase of a mass spectrometer [30,31]. Such techniques will become increasingly valuable for determining subunit stoichiometry and ligand-binding properties of membrane protein complexes.

Hydrogen-deuterium exchange MS (HDX) is also becoming a valuable tool in studying membrane protein dynamics. Because of low dielectric constant and lack of competition from water, hydrogen-bonding is thought to be an important force in the membrane environment with substitution of polar residues being one of the most common disease-causing mutations in membrane proteins. A double-mutant cycle analysis for hydrogen-bonding in bacteriorhodopsin showed that hydrogen-bond interactions in membrane proteins were only modestly stabilizing, however [32]. HDX-MS was also used to map the conformational changes in microsomal glutathione transferase upon binding substrate and on chemical modification of the stress sensor [33]. Another approach targeted towards understanding dynamics of membrane proteins uses microsecond hydroxyl radical ($\cdot\text{OH}$) pulses that 'footprint' solvent-accessible residues via oxidative modifications of amino-acid side chains. The site and extent of oxidative labeling in these experiments is determined by MS [34].

2.4 Mass spectrometry of integral cytochrome *b₆f* complexes

LC-MS+ protocols have been used for analysis of preparations of cytochrome *b₆f* complex for over ten years [10]. The earliest experiments revealed time-dependent specific proteolysis of polypeptide subunits of the complex. This observation led to an effort to speed up the crystallization process, in order to obtain crystals prior to the proteolytic events. Subsequently it was found that addition of an artificial lipid, dioleoylphosphatidylcholine (DOPC), induced rapid crystal formation with sufficient improvement in resolution that the structure could be solved [35]. This original MS analysis also highlighted the likelihood of a covalently attached heme associated with the cytochrome b subunit (PetB) of the complex in addition to the two non-covalently associated b-heme groups known to be present [10,35,36]. The 3.0 Å crystal structure of cytochrome *b₆f* complex by Cramer's group, and the *Chlamydomonas* structure, soon confirmed the presence of the third c-type heme covalently bound to cys35 of PetB [35,36].

More recently, LC-MS+ was applied to cytochrome *b₆f* complex from *Nostoc* with a stable dimeric structure and eight subunits for a total molecular weight of 217 kDa. Covalent modifications of all 8 subunits of the complex were investigated by LC-MS+ and downstream FT-MS to define primary sequences and PTMs [20]. The subunits of cytochrome *b₆f* complex are well known and analysis of mature unit of PetD confirmed removal of initiating Met-1. *N*-acetylation of the N terminus was also confirmed in the Rieske iron-sulfur protein (PetC). Interestingly, we found that the region of PetC most accessible to CAD was its transmembrane region, which contained five of a total of 11 b-ions and 10 of 12 y-ions. Cytochrome *f* (PetA) had residues 1 – 44 removed from the N terminus and a heme attached at Cys-66/Cys-69. Intact mass measurement and the masses of smaller b-ions were consistent with N-terminal acetylation (42.010565 Da; COCH_2) [20]. In PetB the analysis confirmed removal of the initiating Met residue and covalent

attachment of a c-heme with product ions covering the complete sequence with high confidence [20].

Large integral membrane protein complexes are clearly amenable to multiple approaches of mass spectrometry that help gain insights into their primary and tertiary structure. Characterization of PTMs and identification of sequence errors makes intact protein mass spectrometry a valuable tool for membrane proteomics.

3. Experimental protocols

3.1 General remarks

The protocols described have been developed and refined over the last 25 years, and build upon pioneering work of many others over the 25 years preceding those. While the protocols we describe have proved quite general in our hands the diversity of protein structures will surely present examples that require new approaches. Predicting which proteins will fall in this category is beyond our current understanding but examples of larger polytopic integral proteins from mammalian sources are most likely to fall into this category.

3.2 Materials

Solvents and other laboratory supplies are from Fisher Scientific. Solvents achieving HPLC or Optima grade are adequate. Formic acid is ACS grade 88% purity typically assaying at 90%. Trifluoroacetic acid (TFA) is in sealed 1 mL ampules from Pierce and treated with extreme caution until diluted in water or acetonitrile.

3.3 Sample preparation

Various approaches have produced satisfactory data and optimal protocols generally take empirical development. The simplest approach is to inject the sample as supplied and rely upon reverse-phase chromatography to separate the protein from detergents and other contamination. This worked well for the experiment shown in Figure 1 where MIP was separated from detergent by chromatography prior to MS. This approach should not be attempted with the described size-exclusion system as it will precipitate the protein upon exposure to mobile phase. Often it is sufficient to acidify the sample prior to injection, however. Typically, nine volumes of formic acid are added to the sample prior to vortex mixing and injection to LC-MS+ within 120 seconds (100 μ L injected). If a sample is known to contain cysteine residues it is usual to reduce the sample prior to analysis. One half volume of 0.5 M dithiothreitol is added and the sample incubated at room temperature for 20 - 30 minutes prior to acidification as described. Nearly all soluble proteins can be analyzed alongside integral membrane proteins in the separations described, provided disulfide reduction allows general unfolding of the protein. Reduction usually improves ionization efficiency, as well as removing reversible adducts on thiols, such as glutathione, very often improving the quality of the intact protein mass profile. Some proteins can be fully solubilized with lower final acid concentrations including bacteriorhodopsin that requires just three volumes formic acid. Sooner, rather than later, it will be necessary to precipitate a protein sample. This has the immediate advantage that detergents are removed such that they never contaminate the HPLC column. Furthermore, there is a class of polytopic integral membrane proteins that requires very high concentrations of formic acid for full solubility and transfer to the mobile phase of the solvent system used. This was first noted for the twelve transmembrane helix lactose permease [1] that requires dissolution in 90% formic acid prior to solvent transfer into the chloroform/methanol/1% aqueous formic acid described below. It may appear dissolved at lower concentrations but will be trapped by the online filter in front of the HPLC column, until fully solubilized with injections of 90% formic acid. This appears to be a consistent paradigm for transporters in the 50 kDa class.

3.3.1. Precipitation—Membrane protein samples are best left in appropriate stabilizing detergents until ready for mass spectrometry. Mild detergent treatments are preferred as they help keep the protein complexes in a native covalent state. Some membrane protein preparations can be directly loaded into reverse-phase chromatography systems in detergent but typically protein is precipitated prior to re-dissolution in formic acid. To remove detergents and salts, samples are treated with organic solvents to precipitate the proteins. A modified procedure involving chloroform, methanol and water is typically used [38] and works effectively even in the presence of large amounts of SDS and/or Triton X100. Up to 250 μ L of aqueous sample containing microgram quantities of protein (in a 1.5 mL microcentrifuge tube) is diluted with 600 μ L methanol followed by 200 μ L chloroform generating a single phase solution. To this is added 400 μ L water inducing a phase separation. After vigorous mixing the phases are fully separated by centrifugation for 2 minutes at $14,000 \times g$. The protein is precipitated at the interface though the sample should be observed carefully as sometimes the disc of protein slides onto the side of the tube. The upper phase is removed and discarded. The lower phase is sometimes recovered using a Hamilton-type syringe as it may contain a small subset of integral membrane proteins that remain soluble and partition into the chloroform-enriched phase as proteolipids. The protein disc is washed with 600 μ L methanol, recovered again by centrifugation and dried at room temperature and pressure for 5 minutes prior to dissolution in formic acid and immediate analysis. It is sometimes necessary to fully dry the pellet, if it is to be shipped for example, but re-dissolution will be more challenging. Acetone precipitation (80% acetone at -20°C) is also effective for preparations lacking detergents. Each membrane protein preparation can potentially behave differently and the optimal sample preparation protocol is determined empirically.

3.3.2 Size-exclusion chromatography—The size-exclusion chromatography protocol was first developed for the lactose permease of *E. coli* [1] and has become a general workhorse method in the laboratory for membrane and soluble proteins provided they are reduced first. It is also suitable for peptides with a tendency to aggregate such as amyloid. The size-exclusion column (4.6 mm $\times$ 30 cm; SW2000XL, Tosoh Biosciences) is washed with 0.1 % formic acid in water and then 80% methanol prior to equilibration in chloroform/methanol/1% formic acid in water (4/4/1; v/v). The column is protected with a 2 μ m inline porous frit filter which must be washed with formic acid regularly to avoid ‘ghost’ peaks, as new injections of proteins in formic acid remove older proteins previously trapped by the filter. The flow rate is 250 μ L/minute though this can be ramped down to 5 μ L/minute or lower to ‘peak park’ when an extended online top-down MS experiment is required. With this scale of chromatography 2 – 100 μ g protein is required to achieve reasonable signal/noise.

3.3.3 Reverse-phase chromatography—Polymeric reverse-phase stationary phases at elevated temperature have proved robust over many years, allowing the first ESI-MS analysis of a G-protein coupled receptor by ESI-MS [9,10]. The column (2 $\times$ 150 mm, PLRP/S, 5 μ m, 300 Å; Agilent Technologies) is equilibrated in 95 % A, 5 % B (A: 0.1 % TFA in water; B: 0.1 % TFA in acetonitrile/ isopropanol, 1/1, v/v, prepared freshly) at 100 μ L/minute at 40°C prior to gradient elution with increasing buffer B. A typical program includes 30 minutes or longer equilibration time prior to sample injection, followed by 5 minutes at 5 % B, linear gradient ramp to 40 % B at 30 minutes followed by linear gradient ramp to 100 % B at 150 minutes. Columns are not kept at 100 % B for long and are stored in 80 – 90 % methanol. An inline filter is used to protect the column as in 3.3.2. Back pressure will be observed to elevate after a few runs and the column can be regenerated by equilibrating in the ‘4/4/1’ solvent system described in 3.3.2, followed by two blank injections with formic acid. Some membrane proteins elute with partial efficiency in these

experiments and are readily observed to yield ‘ghost’ peaks in subsequent gradients, unless the column is regenerated. Separations involving large intact membrane proteins are demanding and columns should not be expected to last for long. If ten good analytical runs are achieved before retiring a column, each separation has cost in the order of fifty dollars.

3.3.4 Online low-resolution mass spectrometry with fraction collection (LC-MS +)

—Column eluent is directed to a low dead-volume flow splitter and capillary lengths adjusted to yield an approximate 50/50 split with half to a fraction collector (1 minute fractions) and half to the ESI-MS source. Any robust quadrupole, ion-trap or TOF mass analyzer is suitable provided the source, ion transmission and mass range for intact proteins is suitable. Care should be taken with calibration to achieve 100 ppm mass accuracy. Regular microcentrifuge tubes are adequate for fractions collected in reverse-phase experiments and are generally stable at - 80 °C for some months. Fractions from experiments in ‘4/4/1’ are more difficult as the chloroform leaches plasticizers from microcentrifuge tubes. If collected into glass vials care should be taken to acid wash the vials to minimize leaching of Na⁺ which causes adducts in MS. The alternative is to do online LC-MS experiments with peak parking as described in section 3.3.2.

3.3.5 Offline top-down high-resolution mass spectrometry—Fractions are conveniently analyzed by static nanospray ESI-MS on high-resolution FT-MS systems. These experiments are classified as ‘data-directed’ because information from the low-resolution LC-MS+ experiments is used to drive choice of fraction and ion selection for top-down analysis. Long, steady ion currents in nanospray experiments allow for extended averaging of FT-MS transients, essential for maximizing information capture in top-down experiments. We now routinely average 1000 transients averaged during CAD and ECD experiments on 50 kDa class integral membrane proteins. Fractions from LC-MS+ experiments are easily treated with cyanogen bromide (CNBr; 1 gm/mL in acetonitrile stock, 1/10 th volume of fraction is added and incubated in the dark for 5 hours) for middle-down analysis of protein fragments resulting from Metspecific backbone cleavage.

3.3.6 Intact protein bioinformatics and structural assignments—Low-resolution ESI-MS spectra are transformed to zero-charge molecular mass profiles (average mass based upon natural ¹³C abundance) using standard software packages. High-resolution top-down FT-MS data is analyzed using software that transforms mass spectra to monoisotopic zero charge mass. This is a less than perfect process where monoisotopic assignments are frequently ‘off by 1 or more Da’ and newer programs such as MS-Deconv [27] are designed to accommodate this problem. Prosight PTM (<https://prosightptm.northwestern.edu/>) and Prosight PC (Thermo Scientific) software is used to match top-down data to protein primary structure, and to localize PTMs, in what remains a largely manual, iterative process. A paradigm-shifting leap forward would result if real-time data interpretation could be used to drive selection of MS² and MS³ experiments ‘on the fly’.

4. Conclusion

Molecular mass profiling is a valuable and robust tool for crystallographers and biochemists interested in the true primary structure of an isolated protein and its diversity of post-translational modifications. Solvent systems suitable for intact membrane proteins and the components of membrane-embedded complexes have been developed. The LC-MS+ protocol that includes a low-resolution ‘preview’ of primary structure with concomitant fraction collection allows for data-directed top-down experiments for complete sequence and post-translational modification characterization via high-resolution mass spectrometry.

Membrane polypeptide chains up to 50 kDa can now be analyzed routinely by top-down mass spectrometry, though overall sequence coverage can be somewhat limited. Continued improvements in sensitivity, detection and resolution of mass spectrometers will yield better performance as more widespread MS³ becomes possible. Improved bio-informatics capabilities such as real-time data interpretation that directs the use of CAD and ECD experiments could revolutionize the top-down experiment toward complete characterization of a protein's primary structure within the timeframe of a chromatographic peak.

While we have analyzed individual components of a membrane protein complexes as large as 750 kDa by high-resolution MS, we look forward to the time when current technologies to spray intact non-covalent integral membrane protein complexes can be successfully migrated to high-resolution instrumentation. Such development would solidify the remarkable progress that has been made in studies involving subunit stoichiometry and lipid- and ligand-binding properties of membrane protein complexes.

Acknowledgments

We applaud the energetic support of Dr. Fred McLafferty and Dr. Neil Kelleher for the field of intact protein mass spectrometry and the development of top-down mass spectrometry. The MIP sample in Figure 1 was provided by Dr. Guido Zampighi (NIH grant : 2R01-EY004110). Dr. James Bowie is thanked for preparations of bacteriorhodopsin used in this work.

References

- Whitelegge JP, le Coutre J, Lee JC, Engel CK, Privé GG, Faull KF, et al. Toward the bilayer proteome, electrospray ionization-mass spectrometry of large, intact transmembrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* 1999; 96:10695–10698. [PubMed: 10485888]
- Speers AE, Blackler AR, Wu CC. Shotgun analysis of integral membrane proteins facilitated by elevated temperature. *Anal. Chem.* 2007; 79:4613–4620. [PubMed: 17500534]
- Blackler AR, Speers AE, Wu CC. Chromatographic benefits of elevated temperature for the proteomic analysis of membrane proteins. *Proteomics.* 2008; 8:3956–3964. [PubMed: 18780350]
- Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, et al. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotech.* 2007; 25:125–131.
- Kelleher NL, Lin HY, Valaskovic GA, Aaserud DJ, Fridriksson EK, McLafferty FW. Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *Journal of the American Chemical Society.* 1999; 121:806–812.
- Kelleher NL, Zubarev RA, Bush K, Furie B, Furie BC, McLafferty FW, et al. Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxylglutamic acid. *Anal. Chem.* 1999; 71:4250–4253. [PubMed: 10517147]
- Jebanathirajah JA, Pittman JL, Thomson BA, Budnik BA, Kaur P, Rape M, et al. Characterization of a new qQq-FTICR mass spectrometer for post-translational modification analysis and top-down tandem mass spectrometry of whole proteins. *J. Am. Soc. Mass Spectrom.* 2005; 16:1985–1999. [PubMed: 16271296]
- Whitelegge J, Halgand F, Souda P, Zabrouskov V. Top-down mass spectrometry of integral membrane proteins. *Expert Rev Proteomics.* 2006; 3:585–596. [PubMed: 17181473]
- Whitelegge JP, Gunderson CB, Faull KF. Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. *Protein Sci.* 1998; 7:1423–1430. [PubMed: 9655347]
- Whitelegge JP, Zhang H, Aguilera R, Taylor RM, Cramer WA. Full subunit coverage liquid chromatography electrospray ionization mass spectrometry (LCMS+) of an oligomeric membrane protein: cytochrome b(6)f complex from spinach and the cyanobacterium *Mastigocladus laminosus*. *Mol. Cell Proteomics.* 2002; 1:816–827. [PubMed: 12438564]
- Baniulis D, Yamashita E, Whitelegge JP, Zatsman AI, Hendrich MP, Hasan SS, et al. Structure-Function, Stability, and Chemical Modification of the Cyanobacterial Cytochrome b6f Complex from *Nostoc* sp. PCC 7120. *Journal of Biological Chemistry.* 2009; 284:9861–9869. [PubMed: 19189962]

12. Covey TR, Bonner RF, Shushan BI, Henion J. The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry. *Rapid Commun. Mass Spectrom.* 1988; 2:249–256. [PubMed: 2577836]
13. le Maire M, Deschamps S, Møller JV, Le Caer JP, Rossier J. Electrospray ionization mass spectrometry on hydrophobic peptides electroeluted from sodium dodecyl sulfate-polyacrylamide gel electrophoresis application to the topology of the sarcoplasmic reticulum Ca²⁺ ATPase. *Anal. Biochem.* 1993; 214:50–57. [PubMed: 8250254]
14. Schey KL, Papac DI, Knapp DR, Crouch RK. Matrix-assisted laser desorption mass spectrometry of rhodopsin and bacteriorhodopsin. *Biophys. J.* 1992; 63:1240–1243. [PubMed: 1477275]
15. Schindler PA, Van Dorsselaer A, Falick AM. Analysis of hydrophobic proteins and peptides by electrospray ionization mass spectrometry. *Anal. Biochem.* 1993; 213:256–263. [PubMed: 8238899]
16. Turk E, Kim O, le Coutre J, Whitelegge JP, Eskandari S, Lam JT, et al. Molecular characterization of *Vibrio parahaemolyticus* vSGLT: a model for sodium-coupled sugar cotransporters. *J. Biol. Chem.* 2000; 275:25711–25716. [PubMed: 10835424]
17. le Coutre J, Whitelegge JP, Gross A, Turk E, Wright EM, Kaback HR, et al. Proteomics on full-length membrane proteins using mass spectrometry. *Biochemistry.* 2000; 39:4237–4242. [PubMed: 10757971]
18. Schey KL, Papac DI, Knapp DR, Crouch RK. Matrix-assisted laser desorption mass spectrometry of rhodopsin and bacteriorhodopsin. *Biophys. J.* 1992; 63:1240–1243. [PubMed: 1477275]
19. Ryan CM, Souda P, Halgand F, Wong DT, Loo JA, Faull KF, et al. Confident assignment of intact mass tags to human salivary cystatins using top-down Fourier-transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2010; 21:908–917. [PubMed: 20189825]
20. Ryan CM, Souda P, Bassilian S, Ujwal R, Zhang J, Abramson J, et al. Post-translational modifications of integral membrane proteins resolved by top-down Fourier transform mass spectrometry with collisionally activated dissociation. *Mol. Cell Proteomics.* 2010; 9:791–803. [PubMed: 20093275]
21. Mørtz E, O'Connor PB, Roepstorff P, Kelleher NL, Wood TD, McLafferty FW, et al. Sequence tag identification of intact proteins by matching tandem mass spectral data against sequence data bases. *Proc. Natl. Acad. Sci. U.S.A.* 1996; 93:8264–8267. [PubMed: 8710858]
22. Taylor GK, Kim Y-B, Forbes AJ, Meng F, McCarthy R, Kelleher NL. Web and database software for identification of intact proteins using “top down” mass spectrometry. *Anal. Chem.* 2003; 75:4081–4086. [PubMed: 14632120]
23. Horn DM, Zubarev RA, McLafferty FW. Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* 2000; 97:10313–10317. [PubMed: 10984529]
24. Liu X. MS-Deconv. n.d.
25. Liu X, Inbar Y, Dorrestein PC, Wynne C, Edwards N, Souda P, et al. Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Mol. Cell Proteomics.* 2010; 9:2772–2782. [PubMed: 20855543]
26. Thangaraj B, Ryan CM, Souda P, Krause K, Faull KF, Weber APM, et al. Data-directed top-down Fourier-transform mass spectrometry of a large integral membrane protein complex: photosystem II from *Galdieria sulphuraria*. *Proteomics.* 2010; 10:3644–3656. [PubMed: 20845333]
27. Whitelegge JP, Zhang H, Aguilera R, Taylor RM, Cramer WA. Full Subunit Coverage Liquid Chromatography Electrospray Ionization Mass Spectrometry (LCMS+) of an Oligomeric Membrane Protein. *Molecular & Cellular Proteomics.* 2002; 1:816–827. [PubMed: 12438564]
28. Morgner N, Kleinschroth T, Barth H-D, Ludwig B, Brutschy B. A novel approach to analyze membrane proteins by laser mass spectrometry: From protein subunits to the integral complex. *J Am Soc Mass Spectrom.* 2007; 18:1429–1438. [PubMed: 17544294]
29. Meier T, Morgner N, Matthies D, Pogoryelov D, Keis S, Cook GM, et al. A tridecameric c ring of the adenosine triphosphate (ATP) synthase from the thermoalkaliphilic *Bacillus* sp. strain TA2.A1 facilitates ATP synthesis at low electrochemical proton potential. *Molecular Microbiology.* 2007; 65:1181–1192. [PubMed: 17645441]

30. Ilag LL, Ubarretxena-Belandia I, Tate CG, Robinson CV. Drug binding revealed by tandem mass spectrometry of a protein-micelle complex. *J. Am. Chem. Soc.* 2004; 126:14362–14363. [PubMed: 15521749]
31. Barrera NP, Di Bartolo N, Booth PJ, Robinson CV. Micelles protect membrane complexes from solution to vacuum. *Science.* 2008; 321:243–246. [PubMed: 18556516]
32. Joh NH, Min A, Faham S, Whitelegge JP, Yang D, Woods VL, et al. Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature.* 2008; 453:1266–1270. [PubMed: 18500332]
33. Busenlehner LS, Codreanu SG, Holm PJ, Bhakat P, Hebert H, Morgenstern R, et al. Stress sensor triggers conformational response of the integral membrane protein microsomal glutathione transferase 1. *Biochemistry.* 2004; 43:11145–11152. [PubMed: 15366924]
34. Pan Y, Brown L, Konermann L. Kinetic folding mechanism of an integral membrane protein examined by pulsed oxidative labeling and mass spectrometry. *J. Mol. Biol.* 2011; 410:146–158. [PubMed: 21570983]
35. Kurisu G, Zhang H, Smith JL, Cramer WA. Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. *Science.* 2003; 302:1009–1014. [PubMed: 14526088]
36. Stroebel D, Choquet Y, Popot J-L, Picot D. An atypical haem in the cytochrome b(6)f complex. *Nature.* 2003; 426:413–418. [PubMed: 14647374]
37. Wessel D, Flüggé UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 1984; 138:141–143. [PubMed: 6731838]

Fig. 1. Zero-charge intact protein molecular mass profile (MMP) of bovine major intrinsic protein (MIP). The data was collected using a low-resolution triple quad mass spectrometer and transformed to obtain the zero-charge molecular mass profile shown. The molecular heterogeneity of different protein species is clearly visible with phosphorylation (80) at 28303.8, cysteinylolation (119) at 28,343.3 and also C-terminal processing by removal of Ala and Leu/Ile residues. Mass calculated from the MIP gene sequence was 28,223.1 which reflects a delta of 2.1 or 0.0075%, coincident within experimental error. Bovine MIP is one of very few eukaryotic proteins whose mature primary sequence is exactly as predicted from the genomic translation without post-translation processing.

LC-MS+

Fig. 2. Schematic workflow for integral protein mass spectrometry. The technique employs a flow splitter between HPLC and mass spectrometer to facilitate collection of fractions for later use in downstream experiments for protein identification and PTM characterization on high-resolution Fourier transform mass spectrometers.

Fig. 3.

Top-down mass spectrometry of bacteriorhodopsin holoprotein. (A) A typical charge state distribution of bacteriorhodopsin after purification by size exclusion chromatography (SEC) in chloroform/methanol/aqueous formic acid. Paired signals are generated as a result of partial hydrolysis of the retinal chromophore in acidic conditions. (B) Zero change molecular mass profile obtained after deconvolution of selected ion monitoring experiment (m/z 100 width; 40 transients averaged) on the 11-charge ion shows both forms differing by the mass of retinal (266 Da) as well as mild oxidation (+16 Da) and formylation (+28 Da). (C) CAD of the holoprotein is shown with ion isolation of the 11-charge precursor (2460, inset top-left) and its CAD tandem mass spectrum. Unit resolution was achieved on all product ions by operating the instrument at 750,000 resolution at 400 m/z . Note that the ion isolation experimental window mass kept wide enough that there was some contamination of the desired molecular ion with adducts of higher mass, reflective of the ever-present pressure to maximize signal strength for better sequence coverage in the dissociation experiment.

Fig. 4.

Ion assignments for the bacteriorhodopsin holoprotein. Matched peak lists from CAD experiments on 6 different precursor ions were pooled, and the composite list was matched to the structure to give the ion assignments shown. 67 b- and 55 y-ions were matched, giving coverage of 79 of 247 peptide bonds (32%) and a P Score of $3.9\text{e-}150$. The experiment confirms proteolytic processing that removes N-terminal residues 1–13 and the C-terminal Asp262 residue, as well as cyclization of the N-terminus to pyroglutamate and the retinal chromophore between residues 225 – 248, at Lys 229. Transmembrane domains are boxed. Note that while the experiment completely defines the native covalent state of mature bacteriorhodopsin, the regions where bonds were not cleaved rely upon genomic translation for sequence assignment.