Post-translational Modifications of Integral Membrane Proteins Resolved by Top-down Fourier Transform Mass Spectrometry with Collisionally Activated Dissociation*

Christopher M. Ryan‡, Puneet Souda‡, Sara Bassilian‡, Rachna Ujwal§, Jun Zhang§, Jeff Abramson§, Peipei Ping§¶, Armando Durazo∥, James U. Bowie∥∥, S. Saif Hasan‡‡, Danas Baniulis‡‡, William A. Cramer‡‡, Kym F. Fault‡‡§§, and Julian P. Whitelegge‡‡§§¶¶

Integral membrane proteins remain a challenge to proteomics because they contain domains with physico-chemical properties poorly suited to today’s bottom-up protocols. These transmembrane regions may potentially contain post-translational modifications of functional significance, and thus development of protocols for improved coverage in these domains is important. One way to achieve this goal is by using top-down mass spectrometry whereby the intact protein is subjected to mass spectrometry and dissociation. Here we describe top-down high resolution Fourier transform mass spectrometry with collisionally activated dissociation to study post-translationally modified integral membrane proteins with polyhelix bundle and transmembrane porin motifs and molecular masses up to 35 kDa. On-line LC-MS analysis of the bacteriorhodopsin holoprotein yielded b- and y-ions that covered the full sequence of the protein and cleaved 79 of 247 peptide bonds (32%). The experiment proved that the mature sequence consists of residues 14–261, confirming N-terminal propeptide cleavage and conversion of N-terminal Glu-14 to pyrorolidine carboxylic acid (~17.02 Da) and C-terminal removal of Asp-262. Collisionally activated dissociation fragments localized the N6-(retinyldiene) modification (266.20 Da) between residues 225–248 at Lys-229, the sole available amine in this stretch. Off-line nanospray of all eight subunits of the cytochrome b6f complex from the cyanobacterium Nostoc PCC 7120 defined various post-translational modifications, including covalently attached c-hemes (615.17 Da) on cytochromes f and b. Analysis of murine mitochondrial voltage-dependent anion channel established the amenability of the transmembrane β-barrel to top-down MS and localized a modification site of the inhibitor Ro 68-3400 at Cys-232. Where neutral loss of the modification is a factor, only product ions that carry the modification should be used to assign its position. Although bond cleavage in some transmembrane α-helical domains was efficient, other regions were refractory such that their primary structure could only be inferred from the coincidence of genomic translation with precursor and product ions that spanned them. Molecular & Cellular Proteomics 9:791–803, 2010.

Top-down proteomics uses high resolution FT-MS to define proteins with their intact masses, and subsequent dissociation of the intact protein provides primary structure information for unambiguous identification and characterization of covalent modifications (1–3). For top-down proteomics to become globally relevant, it is essential that all segments of the proteome can be addressed by this approach, including the integral membrane proteins of biological bilayer membranes that compartmentalize living cells and constitute around one-third of the proteome and a greater proportion of drug targets (4). Integral membrane proteins present many technical challenges for efficient mass spectrometry in a large part due to hydrophobic transmembrane domains that greatly limit their solubility in aqueous solvents. Technology developments that improve our ability to handle integral membrane proteins and analyze transmembrane domains are therefore of importance.

Considerable improvements in the performance of bottom-up strategies have enabled efficient identification of integral membrane proteins at a proteome-wide level. Some examples include use of alternative proteases such as Proteinase K (5), use of methanol during the digest (6), use of acid-labile surfactant (7, 8) and “gel-C MS,” the use of SDS-PAGE as a first dimension prior to in-gel digestion and recovery of peptides for further separations (9, 10). Wu and co-workers (11, 12) have described improved digestion and
chromatography conditions for better sequence coverage of transmembrane domains. However, it is apparent from this work that although some transmembrane domains yield readily to such strategies there are others that do not. One way to cover such "dark zones" is to analyze the intact protein using the top-down approach.

Previously, it was demonstrated that a variety of integral membrane proteins could be analyzed by ESI-MS on low resolution analyzers with mass accuracy similar to that achievable for soluble proteins (4, 13). Subsequently, top-down mass spectrometry was performed on small integral subunits of the cytochrome b_{6f} complex using quadrupole time-of-flight analyzers (14), and FT-MS was used for the first time on bacteriorhodopsin apoprotein, achieving mass accuracy <10 ppm (15). Preliminary top-down FT-MS of bacteriorhodopsin holoprotein (16) and a thorough top-down collisionally activated dissociation (CAD)1/electron capture dissociation (ECD) study of the c-subunit of F_{o} of the ATP synthase (17) established the feasibility of performing top-down FT-MS on integral membrane proteins. In this study, we present data that establish the general applicability of top-down FT-MS to a variety of integral membrane proteins, including bacteriorhodopsin holoprotein, the subunits of the cytochrome b_{6f} complex from *Nostoc*, and a recombinant form of the murine mitochondrial voltage-dependent anion channel (VDAC) with the transmembrane β-barrel porin motif. CAD was used to localize post-translational modifications with varying degrees of certainty. The data illustrate the wide range of susceptibility of transmembrane domains to CAD and highlight current challenges in data interpretation.

**EXPERIMENTAL PROCEDURES**

**Protein Samples**

*Bacteriorhodopsin*—The protein (1 mg; Sigma B3636 or wild-type protein from *Halobacterium halobium* L33 from the laboratory of James Bowie, UCLA) was suspended in 1 mM CHAPS (100 μl), and 10 dried aliquots were prepared using centrifugal evaporation. An aliquot (100 μg) was wetted with 10 μl of water and dissolved in 90 μl of undiluted formic acid (90%, v/v) prior to immediate injection onto a size exclusion chromatography HPLC system (Super SW2000, Tosoh Biosciences, Montgomeryville, PA) equilibrated in a buffer containing chloroform, methanol, 1% formic acid in water (4:4:1, v/v/v) at 250 μl/min and 40 °C (13, 18) to purify the bacteriorhodopsin away from small molecule contaminants including lipids and CHAPS. Eluent was directed to the standard electrospray ionization source of the LTQ-FT Ultra mass spectrometer (Ionmax) with the flow dropped manually to 10 μl/min as the UV absorbance exceeded 50 milliabsorbance units at the start of the first peak containing the protein.

*Cytochrome b_{6f} Complex*—Samples (250 μg of protein provided by the laboratory of William Cramer, Purdue University) were precipitated using acetone. The suspension was split into two microcentrifuge tubes (125 μl each), and 1 ml of 80% acetone in water (−20 °C stock) was added to each tube prior to Vortex mixing (1 min) and incubation at −20 °C for 1 h. Precipitated protein was recovered by centrifugation (10,000 × g), and the supernatant was removed. Pellets were dried briefly to allow evaporation of residual acetone (5 min, room temperature and pressure) and dissolved in 90% formic acid (total 100 μl) for immediate injection into an HPLC system prepared for reverse-phase chromatography (14, 18). A column (5 μm, 300-Å PLRP/S, 2 × 150 mm; Varian) was previously equilibrated in 95% buffer A (0.1% TFA in water), 5% buffer B (0.05% TFA, 50% acetonitrile, 50% isopropanol) at 100 μl/min at 40 °C for 30 min prior to sample injection. The column was eluted with a stepped linear gradient of increasing buffer B as described previously (14), and the eluent was directed to a liquid flow splitter delivering 50 μl/min to a low resolution mass spectrometer and 50 μl/min to a fraction collector (1 min/fraction). Fractions, selected by inspection of the low resolution MS data, were subjected to manual direct infusion nanospray analysis on the high resolution LTQ-FT Ultra.

**Voltage-dependent Anion Channel**—Aliquots of VDAC (10 mg/ml; 10 μl; from the laboratory of Jeff Abramson, UCLA) were diluted with 90% formic acid (90 μl) prior to mixing and immediate injection onto the size exclusion chromatography system described for bacteriorhodopsin. In the case of VDAC, fractions were collected to acid-washed glass vials for direct infusion nanospray analysis on the LTQ-FT Ultra.

**Direct Infusion Analysis**

HPLC fractions were individually loaded into 2-μm-inner diameter externally coated nanospray emitters (Proxenon, Cambridge, MA) and desorbed using a spray voltage of between 1.7 and 1.9 kV (versus the inlet of the mass spectrometer) using the nanospray source supplied by the manufacturer. These conditions produced a flow rate of 20–50 nl/min.

**Mass Spectrometry**

All samples were analyzed using a hybrid linear ion trap/FTICR mass spectrometer (7 tesla, LTQ-FT Ultra, Thermo Scientific, Bremen, Germany) operated with standard (up to 2000) or extended mass range (up to 4000). Ion transmission into the linear trap and further to the FTICR cell was automatically optimized for maximum ion signal. The ion count targets for the full-scan FTICR and MS/MS FTICR experiments were 2 × 10^6. The m/z resolving power of the FTICR mass analyzer was set at 100,000 (defined by m/zΔm50% at m/z 400) unless otherwise stated. Individual charge states of the multiply protonated protein molecular ions were selected for isolation and collisional activation in the linear ion trap followed by the detection of the resulting fragments in the FTICR cell (CAD). For the CAD studies, the precursor ions were activated using collision energy settings in the range of 10–15 at the default activation q-value of 0.25. FT-MS data were derived from an average of between 50 and 200 transient signals.

**Data Analysis**

FTICR spectra were processed using ProSightPC software (Pro-SightPC 1.0, Thermo Scientific) to produce monoisotopic mass lists (signal/noise = 2, minimum RI = 0.9) that were then assigned to protein sequences with various post-translational modifications (Table I). Protein identification was achieved by generating sequence tags using the sequence tag compiler and sequence tag searching tools within ProSightPC (minimum tag score, 0.01; minimum tag size, 4; tolerance, 10 ppm) and matching these tags to an appropriate database (the complete Nostoc sp. PCC 7120 proteome database as translated from the genome was downloaded from NCBI on July 18, 2008; NC_003240.faa, NC_003241.faa, NC_003267.faa, NC_003270.

---

1 The abbreviations used are: CAD, collisionally activated dissociation; ECD, electron capture dissociation; VDAC, voltage-dependent anion channel; LTQ, linear trap quadrupole.
Membrane Protein Post-translational Modifications by FT-MS

f(a, NC_003272.faa, NC_003273.faa, and NC_003276.faa). Product ion assignments for known proteins were made using ProSightPC operated in single protein mode with a 10-ppm mass accuracy threshold and with the delta mass feature deactivated. ProSightPC was used to determine mass accuracy of assigned monoisotopic mass values of precursor and product ions and the chance that some other database entry might match the same data set (P Score). Modification masses are shown in Table I. To consider the "delta 1 Da" problem, the peak list used for Fig. 1D was expanded; unmatched monoisotopic product ion masses from the six CAD experiments were used to calculate a new peak list with values ±1.00235 from those unmatched masses. This list was added to the matched peak list, and the expanded list was matched once again to the bacteriorhodopsin structure. The sequence for bacteriorhodopsin was taken from Swiss-Prot (P02945). The sequence for recombinant bacteriorhodopsin structure. The sequence for bacteriorhodopsin peak list, and the expanded list was matched once again to the bacteriorhodopsin structure. The sequence for bacteriorhodopsin was taken from Swiss-Prot (P02945). The sequence for recombinant bacteriorhodopsin was taken from Swiss-Prot (P02945). The sequence for recombinant bacteriorhodopsin was taken from Swiss-Prot (P02945). The sequence for recombinant bacteriorhodopsin was taken from Swiss-Prot (P02945). The sequence for recombinant bacteriorhodopsin was taken from Swiss-Prot (P02945).

RESULTS

Sample Preparation for Retention of Labile Post-translational Modification: Bacteriorhodopsin—Previous studies have defined conditions that lead to retention of the retinal chromophore of bacteriorhodopsin (4, 16, 20), although the Schiff linkage remains susceptible to hydrolysis under the acidic conditions used in this analysis. For top-down FT-MS on the bacteriorhodopsin holoprotein, the sample was infused directly to the source of the mass spectrometer (Lonnax on LTQ-FT Ultra) using size exclusion chromatography. The protein peak elutes at 7.5 min, and the flow rate was ramped from 250 to 10 μl/min at 7 min to maximize the time available for data acquisition during the top-down experiment (peak parking). This enabled top-down CAD analysis of the holoprotein, although the cofactor was completely hydrolyzed by the time the protein was fully eluted after 15–20 min. With resolution set at 750,000 (at 400 m/z), this allowed for averaging of ~100 transients prior to data storage. Fig. 1A shows the mass spectrum as the bacteriorhodopsin starts to elute, and Fig. 1B shows the mass spectrum 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 Da). Although the duration of this experiment was adequate for a reasonable CAD analysis, this was apparently not the case for efficient ECD analysis (data not shown). A small amount of the 26,583-Da species was frequently observed and may be due to in-source dissociation.

Extended Mass Range for Improving Sequence Coverage—Bacteriorhodopsin is typical of many integral membrane proteins in that it has less ionizable side-chain functionalities such that it is relatively poorly charged in ESI-MS compared with soluble proteins. Strong ion currents are observed for several ions over the 2000-Da limit usually used on the linear ion trap so a number of experiments were performed using the high mass range capability of the instrument (up to 4000 m/z). The most obvious effect on the CAD spectrum was a quite dramatic widening of the range over which useful product ions were recovered (shown here for CAD on the 2460 precursor ion (1250–3400 Da; Fig. 1C). This translated to improved sequence coverage, an important consideration for characterization of post-translational modification sites. The disadvantage was that the instrument had to run at higher resolution (~100,000 at 2000 m/z) to achieve unit resolution of the higher m/z ions, thereby increasing duty cycle. Fig. 1C also shows the ion isolation used for the CAD experiment on the 2460-Da precursor. Note that the transmission window is widened to allow full transmission of the 2460-Da ion to maximize the yield of useful product ions. The disadvantage of this is that the ion isolation includes minor adduct species arising from methionine oxidation (+16 Da) and formylation (+28 Da), which add to the complexity of data interpretation. Recalibration—Careful inspection of some of the high m/z ions revealed modest deviation from calibration compared with ions below 2000 m/z (0.022 at 2460 m/z). Internal calibration improved mass accuracy on parent ion assignments in the high mass range (required to achieve 5 ppm; see Table II) while having little effect on the number of product ion assignments (data not shown). We concluded that there was little benefit to recalibration provided the FTICR analyzer was adequately externally calibrated, a process performed daily.

Delta 1 Da Problem and Misassignment of Monoisotopic Peaks—Even after recalibration nearly all precursor ion monoisotopic mass estimates from Xtract were 1 Da too high with that calculated for the known primary structure. To test the accuracy of monoisotopomer extrapolation, the isotopomer distribution for the known atomic composition was calculated (isotopident tool). Interestingly, the known
FIG. 1. **Top-down mass spectrometry of bacteriorhodopsin holoprotein.** A, ESI mass spectrum of bacteriorhodopsin after purification by size exclusion chromatography in chloroform/methanol/aqueous formic acid. The spectrum shown was recorded by the FTICR operated below unit resolution to illustrate the typical charge state distribution observed for this integral membrane protein. The retinal chromophore is hydrolyzed from the polypeptide in acidic conditions such that paired signals arise from both apo- and holoforms at each charge state. Approximate average masses and charge states are labeled. Ion statistics for individual ions were poor in this full scan experiment. B, deconvolution of apo- and holoforms. The zero-charge molecular mass profile obtained after deconvolution of a 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) of the polypeptide. With the instrument operated at 750,000 resolution at m/z 400, a resolution of around 60,000 was achieved for the 11-charge ions at m/z 2460. C, CAD of the holoprotein. The ion isolation on the 11-charge
precursor (2460; inset top left) and its CAD tandem mass spectrum are shown. The ion isolation is widened to ensure maximal signal strength. Note that this experiment uses the extended mass range of the ion trap with useful product ions up to nearly 3200 m/z. Unit resolution was achieved on all product ions by operating the instrument at 750,000 resolution at 400 m/z. D, ion assignments for the bacteriorhodopsin holoprotein. CAD experiments were performed on 11-, 12-, 13-, 14-, 17-, 18-charge precursors, and the product ions were matched to the known primary structure of bacteriorhodopsin with its retinal cofactor at Lys-216 using ProSightPC software (version 1.0) operated at 10-ppm tolerance with the delta mass feature deactivated. Matched peak lists were pooled, and the composite list was again 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 3.9e−150. Transmembrane domains are boxed. AU, arbitrary units; R, resolution.

Fig. 1—continued
atomic composition predicted the most abundant isotopomer as $a + 16$, whereas use of average by Xtract predicted $a + 15$. Thus, it was possible to account for the delta 1 Da problem and achieve mass accuracy of less than 5 ppm on the precursor ions (see Table II). The actual relative atomic abundances of bacteriorhodopsin were compared with the average model, showing that the protein was relatively enriched in total carbon (33.0\% versus 31.7\%), although this did not appear to have a significant effect on isotopomer abundance. It is likely that other delta 1 Da assignments arise because of suboptimal data quality (lower signal to noise ions) and imperfections in the software used for these assignments.

Sequence Coverage in CAD—Preliminary top-down analysis of bacteriorhodopsin using the normal mass range of the instrument gave just five $b$-ions and seven $y$-ions matched for a single CAD experiment on the holoprotein (16). By improving protocols and working with the high mass range capability of the instrument, we have now considerably extended coverage of the protein (Fig. 1D). We also observed that CAD analysis of different precursor ions generated mainly overlapping but also some novel product ions and therefore performed the top-down experiment of six different precursors and then pooled the peak lists from all six experiments. In this way, it was possible to match 67 $b$- and 55 $y$-ions, giving coverage of 79 of 247 peptide bonds (32\%). The presence of numerous overlapping $b$- and $y$-product ions confirms full sequence coverage. Comparison of bond coverage for three lower $m/z$ precursors versus the three higher $m/z$ precursors confirms that more bonds were covered with the higher $m/z$ precursor ions, although several unique dissociation sites were observed with the lower $m/z$ set. It is clear that dissociation of different precursor ion charge states yields complementary information with best coverage achieved by pooling the results from several different precursor ions. The covalent retinal modification is localized to a stretch of 22 amino acid residues (y14 and y36) surrounding the known site at Lys-216. Considering the known chemistry of the modification, requiring a primary amine as the attachment site, Lys-216 is the only possibility within the segment localized. Sequence coverage in transmembrane domains was substantial in the first four helices yet poor in the last three (Fig. 1D, boxed). To assess the contribution of delta 1 Da errors to the interpretation process, 1.00235 Da was subtracted from the unmatched masses from each of the six different parent CAD experiments, and the results were assembled into a peak list with the good matches described above. This peak list was found to cover 86 bonds (up from 79). When ±1.00235 Da was considered, the coverage increased to 95 bonds (80 $b$- and 69 $y$-ions; 38\%), although the P Score dropped to 1.98e−52 as many “peaks” were now unmatched, limiting the practical advantage of performing such an operation. These extra bond assignments may include some false positives because the expanded peak

<table>
<thead>
<tr>
<th>Protein</th>
<th>SwissProt accession</th>
<th>Atomic formula</th>
<th>Calculated</th>
<th>Measured</th>
<th>Delta (ppm)</th>
<th>P Score</th>
<th>Modifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterio-rhodopsin</td>
<td>P02945</td>
<td>C$<em>{1268}$H$</em>{2369}$N$<em>{334}$O$</em>{334}$S$_{9}$</td>
<td>27032.3256</td>
<td>27032.3443</td>
<td>4.5$^a$</td>
<td>6.84e-43</td>
<td>Residues 14–261$^{b}$, N-terminal pyroglu, retinal at 229</td>
</tr>
<tr>
<td>PetD</td>
<td>Q93SX1</td>
<td>C$<em>{858}$H$</em>{1328}$N$<em>{232}$O$</em>{232}$S$_{9}$</td>
<td>17393.4150</td>
<td>17393.4581</td>
<td>2.5</td>
<td>1.37e-35</td>
<td>Residues 2–160</td>
</tr>
<tr>
<td>PetD DP-cleaved</td>
<td>Q93SX1</td>
<td>C$<em>{537}$H$</em>{827}$N$<em>{123}$O$</em>{123}$S$_{2}$</td>
<td>11185.2005</td>
<td>11185.2100</td>
<td>0.87</td>
<td>1.05e-35</td>
<td>Residues 59–160</td>
</tr>
<tr>
<td>PetC</td>
<td>Q93SX0</td>
<td>C$<em>{858}$H$</em>{1328}$N$<em>{232}$O$</em>{232}$S$_{9}$</td>
<td>19092.4045</td>
<td>19092.5230</td>
<td>6.2</td>
<td>1.56e-08</td>
<td>Residues 2–179, N-terminal acetylation, 2 disulfides</td>
</tr>
<tr>
<td>PetA</td>
<td>Q93SW9</td>
<td>C$<em>{1432}$H$</em>{2522}$N$<em>{332}$O$</em>{332}$S$_{9}$Fe</td>
<td>31746.1339</td>
<td>31746.0932</td>
<td>1.34</td>
<td>9.42e-49</td>
<td>Residues 45–233, heme c (Fe$^{III}$), one E to Q amidation</td>
</tr>
<tr>
<td>PetB</td>
<td>P0A384</td>
<td>C$<em>{1117}$H$</em>{1755}$N$<em>{270}$O$</em>{270}$S$_{10}$Fe</td>
<td>24739.7572</td>
<td>24739.7929</td>
<td>1.40</td>
<td>3.4e-28</td>
<td>Residues 2–215, heme c (Fe$^{III}$), one E to Q amidation</td>
</tr>
<tr>
<td>PetG</td>
<td>P58246</td>
<td>C$<em>{189}$H$</em>{302}$N$<em>{44}$O$</em>{50}$S$_{1}$</td>
<td>4020.2162</td>
<td>4020.2218</td>
<td>1.28</td>
<td>3.25e-55</td>
<td>N-formyl on Met1</td>
</tr>
<tr>
<td>PetM</td>
<td>P0A3Y1</td>
<td>C$<em>{1472}$H$</em>{2218}$N$<em>{394}$O$</em>{443}$S$_{9}$</td>
<td>3571.9364</td>
<td>3571.9384</td>
<td>0.43</td>
<td>2.35e-82</td>
<td>N-formyl on Met1</td>
</tr>
<tr>
<td>PetL</td>
<td>Q8YVQ2</td>
<td>C$<em>{158}$H$</em>{255}$N$<em>{35}$O$</em>{36}$S$_{1}$</td>
<td>3250.8919</td>
<td>3250.9025</td>
<td>3.12</td>
<td>2.32e-41</td>
<td>N-formyl on Met1</td>
</tr>
<tr>
<td>PetN</td>
<td>P61048</td>
<td>C$<em>{156}$H$</em>{242}$N$<em>{36}$O$</em>{36}$S$_{2}$</td>
<td>3259.7653</td>
<td>3259.7636</td>
<td>0.6</td>
<td>1.5e-32</td>
<td>N-formyl on Met1</td>
</tr>
<tr>
<td>CyTC</td>
<td>P62894</td>
<td>C$<em>{3533}$H$</em>{585}$N$<em>{147}$O$</em>{150}$S$_{1}$Fe</td>
<td>12222.007</td>
<td>12222.1769</td>
<td>2.07</td>
<td>1.12e-35</td>
<td>N-formyl on Met1</td>
</tr>
<tr>
<td>VDAC1 Un-modified$^{c}$</td>
<td>Q60932</td>
<td>C$<em>{1428}$H$</em>{2218}$N$<em>{394}$O$</em>{443}$S$_{5}$</td>
<td>32134.1741</td>
<td>32134.2473</td>
<td>2.21</td>
<td>6.57e-13</td>
<td>Residues 1–295</td>
</tr>
<tr>
<td>VDAC1 modified</td>
<td>Q60932</td>
<td>C$<em>{1442}$H$</em>{2248}$N$<em>{394}$O$</em>{444}$S$_{6}$</td>
<td>32408.4043</td>
<td>32408.4692</td>
<td>2.79</td>
<td>1.08e-9</td>
<td>Residues 1–295, Modification not yet placed</td>
</tr>
<tr>
<td>VDAC1 modified</td>
<td>Q60932</td>
<td>C$<em>{1442}$H$</em>{2248}$N$<em>{394}$O$</em>{444}$S$_{6}$</td>
<td>32408.4043</td>
<td>32408.4692</td>
<td>2.79</td>
<td>6.86e-12</td>
<td>Residues 1–295, Modification not yet placed</td>
</tr>
</tbody>
</table>

$^a$ Average of three recalibrated, parent ions from extended mass range >2000.
$^b$ C-terminal D262 removed from bacteriorhodopsin.
$^c$ recombinant murine VDAC1.
lists cover a large amount of number space even at the 10-ppm tolerance used. The delta 1 Da problem leads us to believe that future improvements in performance of the algorithms that assign monoisotopic mass will likely yield improved sequence coverage, although some uncertainties will remain where ion statistics are poor.

**Fig. 2.** Top-down mass spectrometry for identification of post-translationally modified unknown. A, sequence tags from an unknown species of mass 11,185 Da. A top-down experiment was performed on the m/z 1245 ion (9+), corresponding to a previously unseen species within the preparation. The sequence tag compiler function of ProSightPC 1.0 was used to generate a list of potential sequence tags using default parameters. These tags were then matched to the complete Nostoc proteome database using the same software. B, result of sequence tag search. The sequence tag search identified a known subunit of the cytochrome b_{6}_{f} complex as the best match. Subunit 4 (PetD) was matched with two tags of 4 and 5 amino acid residues in length as highlighted. C, ion assignments without refinement of primary structure. The PetD sequence was compared with the top-down product ion peak list with 10 y-ions matched, strongly suggesting modification of the N terminus. D, ion assignments after refinement of N terminus. The first 58 amino acid residues were removed from the N terminus, resulting in close agreement of measured and calculated masses for this species (\( \pm 1\) ppm). An additional set of 19 b-ions was subsequently matched, confirming the N terminus with a P Score of 7.3e–35. It was concluded that the 11,185-Da molecule was an artifact resulting from DP cleavage due to brief exposure to concentrated formic acid.
Fig. 3. Post-translational modifications of large subunits of the Nostoc cytochrome $b_{f}$ complex. A, N-terminal acetylation of the Rieske Fe-S subunit (PetC). CAD was performed on a 19-charge precursor at m/z 1006. Product ions were matched to PetC, confirming N-terminal acetylation after removal of the initiating Met residue. The iron-sulfur center was apparently displaced with subsequent oxidation of Cys.
De Novo Identification: Novel Subunit of the Cytochrome b_{6f} Complex?—Although the subunits of the cytochrome b_{6f} complex are well known, there is always the possibility of an extra subunit appearing when a new species is investigated. During our analysis of the *Nostoc* preparation, a polypeptide of mass 11,185 Da was observed to be quite abundant. A precursor ion of m/z ~1245 (9+) was chosen for CAD, and the product ion spectrum was recorded. The “compile sequence tag” feature of ProSightPC was used to generate a number of short sequence tags from the data (Fig. 2A), and these tags were used to perform a sequence tag search through the entire *Nostoc* database. The strongest hit, with a pair of sequence tags of 4 and 5 amino acid residues (Fig. 2B), turned out to be subunit 4 (PetD) of the cytochrome b_{6f} complex, suggesting that the 11,185-Da species was in fact a truncated form of one of the known subunits (usual PetD mass was 17,393 Da; Table II). When the peak list was matched to the subunit 4 sequence, 10 y-ions were matched within 10 ppm, implicating an N-terminal cleavage (Fig. 2C). By removing the first 58 amino acid residues, the measured mass was within 0.87 ppm of that calculated for the truncated sequence, and 19 additional b-ions were matched (Fig. 2D) to give an overall P Score of 7.3e−35. Because DP cleavage is a known chemistry of concentrated formic acid, it was concluded that the 11,185-Da species resulted from hydrolysis of the peptide bond between Asp-58 and Pro-59 and that this event was an unnatural artifact due to the use of 90% formic acid for sample solubilization despite a protocol that limits exposure to less than 2 min.

Post-translational Modification of Subunits of Cytochrome b_{6f} Complex—Top-down mass spectrometry was used to investigate covalent modifications of the large subunits of the cytochrome b_{6f} complex whose masses had previously been measured on a low resolution instrument (21). Top-down analysis of the mature subunit of PetD confirmed removal of initiating Met residue with a P Score of 1.4e−35 (Table II). Analysis of the Rieske iron-sulfur protein (PetC) confirmed the presence of N-acetylation of the N terminus after removal of Met-1 (Fig. 3A) as reported previously. Both the intact mass measurement and the mass of smaller b-ions were more consistent with N-terminal acetylation (42.010565 Da; COCH_{3}) as opposed to a mutation that changed Ala-2 to Leu/Ile (delta mass, 42.04695 Da; C_{2}H_{5}). The data set was most consistent with a pair of disulfides, the second presumably having formed after the initial LC-MS+ analysis. Interestingly, the region of PetC most accessible to collisionally activated dissociation was its transmembrane region (Fig. 3, shaded), 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 c-heme attached at Cys-66/Cys-69. To achieve an optimal match of measured precursor mass to that calculated and improve sequence coverage, we also altered one N-terminal Glu to Gln (amidation) (Fig. 3B and Table II). Although this analysis provided strong evidence that the protein identification and processing was accurate (P Score, 1.06e−21) with the entire sequence of the protein covered, coverage of individual bonds was quite low (17 of 288). In the case of cytochrome b (PetB; Fig. 3C), 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 (P Score, 3.4e−28). Once again coverage of individual bonds was less extensive (20 of 213) such that it was not possible to confirm modification of Cys-35 by c-heme rather than Cys-43. This example also needed an altered N-terminal Glu to Gln (amidation) to achieve an optimal match of measured precursor mass to that calculated and improve sequence coverage. Confirmation of the proposed point amidations for PetA and PetB will require further analysis beyond the scope of this study. Note that in all examples of covalent c-heme modification we used a delta mass of 615.16947 Da. This was confirmed for the standard bovine cytochrome c (Table II and bottom-up peptide experiments not shown) and results due to oxidation of the heme iron in the electrospray source to Fe(III), thereby giving the heme a 1+ charge without it carrying a proton. The mass of this proton is subtracted from the mass of heme (heme b; 616.177295) because the software used for interpretation assumes all charges are due to addition of H^{+} (22). The four small subunits of the *Nostoc* cytochrome b_{6f} complex were all found to be unmodified after translation as each one retained its initiating formylmethionine residue with the formyl group intact (Table II). Note that retention of the formyl group is considered as a modification for data processing.

Murine Voltage-dependent Anion Channel (VDAC1)—Top-down mass spectrometry was used to investigate modification of a recombinant form of murine mitochondrial VDAC1 with a potential inhibitor, Ro 68-3400 (23). Conditions were adjusted such that the protein became predominantly modified with a single Ro 68-3400 molecule (see supplemental Fig. S1). Precursor and product ions from the unmodified form matched the known sequence of the recombinant form of the protein within 3-ppm mass accuracy on the precursor and an...
Membrane Protein Post-translational Modifications by FT-MS

Overall P Score of 6.57e–13 (Table II). Analysis of the singly modified form of the protein was also performed with the specific goal of determining whether all the modifications were at a single site consistent with a specific inhibitor. Based on the structure of the molecule, we predicted modification of either of the two Cys residues within the sequence at positions 127 and 232 (139 or 244 with His tag). First, the available data from the singly modified precursor was matched to the unmodified protein sequence (Fig. 4A), achieving product b- and y-ion matches that overlapped (shaded light blue). Such an observation is usually an indication that the complete sequence is covered and correct, but in this case, where the difference between precursor and sequence is 274 Da due to the presence of the bound inhibitor, the data must be examined further to explain the discrepancy. The observation of overlapping b- and y-ions could arise in two ways by either neutral loss, whereby some product ions that carried the modification lost it such that their measured mass corresponded to the unmodified ion, or the presence of mixed populations with modifications at both Cys residues (24). The neutral loss problem can be solved by relying solely upon product ions that carry the modification. The data were consequently fitted with either the modification at Cys-127 (Fig. 4B) or at Cys-232 (Fig. 4C). When placed at Cys-127 a number of y-ions appear to support non-modification at Cys-232 (shaded light blue), although these ions do not carry the modification and could result due to neutral loss (Fig. 4B). y-ions that carry the modification (shaded dark blue) were consistent with modification of either Cys residue. When the modification is placed at Cys-232, a number of y-ions that carry the modification support this assignment (shaded dark blue), providing strong evidence that Cys-232 is indeed modified (Fig. 4C). Several of these ions that carry the modification can also be seen without the modification in the assignments for Cys-127 (Fig. 4B), providing support for the idea that neutral loss is contributing to the problem. Because no ions that carry the modification support localization at Cys-127, we cannot make a firm conclusion as to whether this site is modified (substoichiometrically) or not. Subsequent experiments have shown that several sites on VDAC1 can be modified by Ro 68-3400 (data not shown), leading us to conclude that this molecule is in fact quite nonspecific such that further experiments are unwarranted. Nevertheless, the data presented establish the accessibility of integral transmembrane porins to analysis by top-down mass spectrometry and highlight some practical features that complicate interpretation of top-down mass spectrometry data.

Raw data (Thermo, .raw) and peak lists (ProSightPC, .puf) were uploaded to Tranche. They may be accessed at https://proteomecommons.org/dataset.jsp?id=74238.

DISCUSSION

General Applicability of Top-down FT-MS to Integral Membrane Proteins—In previous work, we established the feasibility of top-down FT-MS for studies of integral membrane proteins and their transmembrane domains and emphasized the power of hybrid linear ion trap FT-MS to yield precursor and product ion mass accuracy of <5 ppm (see the Introduction). In this study, we have further explored the capabilities and limitations of hybrid FT-MS systems with respect to a range of integral membrane proteins of both polytopic α-helical bundle and porin motifs with masses up to 35 kDa and carrying a variety of post-translational modifications. It is clear that top-down high resolution mass spectrometry is generally applicable to studies of membrane proteins, provided chromatographic conditions are defined that are compatible with electrospray ionization. Deeper proteome coverage requires multidimensional separations, and the use of a “gel-free” approach has been recently described for top-down proteomics (25). The gel-free approach uses a continuous elution gel system for a size-based separation and liquid fractions for top-down analysis. Such a system might be suitable for integral membrane proteins if conditions can be defined for their transfer from aqueous detergent micellar solution to an aqueous-organic mixture compatible with their solubility and ESI. Alternatively, there are advantages to using non-denaturing conditions in the first dimension with respect to membrane proteins, and the use of size exclusion chromatography sep-

Fig. 4. Top-down mass spectrometry of integral transmembrane porin, murine VDAC1 modified with Ro 68-3400. A, ion assignments, no modification included. VDAC1 carrying one molecule of Ro 68-3400 was analyzed by top-down mass spectrometry. CAD was performed on the 32-charge precursor at m/z 1014. The ion assignments shown were obtained by matching the product ion peak list from the singly modified VDAC1 to the unmodified sequence (with a consequent precursor delta mass of 274 Da). A set of overlapping b- and y-ions (shaded light blue; ions do not carry the modification) indicate that the data support full sequence coverage of unmodified VDAC1, alerting us to potential problems with the data set at least with respect to ions that do not carry the modification. Ions that do not carry the modification can arise from ions that do carry the modification due to neutral loss and are consequently ambiguous with respect to useful localization of a modification. Only product ions that carry the modification should be used to localize the modification site. B, ion assignments, Cys-127 modification. With the modification placed at Cys-127, a set of product y-ions appears (shaded dark blue; ions carry the modification) that localizes the modification to the C-terminal segment of the protein that contains both Cys residues. A number of unmodified product ion assignments (shaded light blue) support modification of Cys-127 but should be regarded with skepticism based upon the interpretation of A. According to the crystal structure, Cys-127 protrudes into the bilayer. C, ion assignments, Cys-232 modification. With the modification placed at Cys-232, a new set of y-ions appears that carries the modification (shaded dark blue) and excludes Cys-127. Thus, assuming only Cys residues can be modified, there is firm evidence to support the presence of Cys-232 modification. It remains possible that a substoichiometric population modified at Cys-127 is present, but this cannot be confirmed with the available ion assignments. According to the crystal structure, Cys-232 protrudes into the pore channel.
arations of integral membrane protein complexes as the first of two dimensions in the context of top-down proteomics has been demonstrated (26). The "blue native" gel is probably the gold standard for separation of integral membrane protein complexes and might be the best option if ways can be devised to release complexes from the gel. This was achieved using electroelution with subsequent MS of the intact mitochondrial complex 1 using the laser-induced liquid bead ion desorption ionization technique, although it was noted that the St1 subunit was displaced from the complex apparently during running of the gel (27).

Coverage of Transmembrane Domains—Bond cleavage by collisionally activated dissociation did not appear to be correlated with the presence or absence of transmembrane domains with some accessible to dissociation and others remaining quite impenetrable. Although this parallels the results of even the most successful bottom-up studies of membrane proteins, there is an important difference, namely that the top-down approach covers regions that are resistant to dissociation through intact mass measurement and product ions that include these regions (see Fig. 1, for example). Thus, when product ions including N and C termini overlap and precursor mass is in agreement with that calculated for the primary structure, it is concluded that full sequence coverage is achieved with the structure of regions poorly accessible to dissociation inferred from genomic translation. Such coverage is adequate until it is necessary to locate a novel modification when strategies must be devised to improve bond coverage in appropriate regions. Alternative dissociation mechanisms such as ECD (28, 29) and/or MS^3 (30) or middle-down approaches (31) are available in this respect. Previously, it was demonstrated that gas-phase thermal excitation was necessary for efficient ECD (activated ion ECD) of the c-subunit of F_0 of the ATP synthase (8-kDa with two transmembrane -helices; Ref. 17). We are currently exploring conditions for efficient ECD and activated ion ECD of the proteins described in this study and looking forward to the application of electron transfer dissociation to integral membrane proteins.

Localization of Post-translational Modifications—This study focused on the ability of top-down CAD experiments to precisely localize post-translational modifications. Although most of the experiments successfully localized known and unknown post-translational modifications, the precision of the localization was not always unambiguous. Furthermore, in two examples, it was necessary to introduce a single hypothetical amidation to best explain the available data, clearly an undesirable situation. In the example of exogenous modification of the porin VDAC1 with a small molecule, we discussed ambiguities that arise due to technical challenges such as neutral loss and heterogenous modification sites. Firm conclusions could only be drawn from product ions that carried the modification, limiting the completeness with which the experiment could be interpreted. In hind site after extensive data analysis, we conclude that there are not enough data to draw firm conclusions on primary structure assignment and that we need to return for further experiments as described in the previous section. The utility of the top-down technique would benefit immensely from on-the-fly data interpretation so that immediate experiments could be performed to address incomplete data sets with termination of the experiment only when full primary structure is confirmed. On-the-fly data interpretation and real time, data-driven instrument control will undoubtedly appear in the future as software and hardware solutions catch up with the true analytical power of high resolution mass spectrometry.

Acknowledgment—We thank Professor Joseph Loo for access to the LTQ-FT, which was purchased with National Institutes of Health support (Grant S10 RR023045).

This work was supported, in whole or in part, by National Institutes of Health Grants R21 RR025811, U19 AI067769, PO1 HL0111 and GM-32383 (WAC).

This article contains supplemental Fig. S1.

To whom correspondence should be addressed: The Pasarow Mass Spectrometry Laboratory, The NPI-Semel Inst., David Geffen School of Medicine, University of California Los Angeles, 760 Westwood Plaza, Los Angeles, CA 90024. Tel.: 310-794-5156; Fax: 310-206-2161; E-mail: jpw@chem.ucla.edu.

REFERENCES


Zabrouskov, V., and Whitelegge, J. P. (2007) Increased coverage in the transmembrane domain with activated-ion electron capture dissociation for top-down Fourier-transform mass spectrometry of integral membrane proteins. J. Proteome Res. 6, 2205–2210


