A method for the comprehensive proteomic analysis of membrane proteins

Christine C. Wu[1]*, Michael J. MacCoss[1]*, Kathryn E. Howell[2], and John R. Yates III[1,3†]

Published online 14 April 2003; doi:10.1038/nbt819

We describe a method that allows for the concurrent proteomic analysis of both membrane and soluble proteins from complex membrane-containing samples. When coupled with multidimensional protein identification technology (MudPIT), this method results in (i) the identification of soluble and membrane proteins, (ii) the identification of post-translational modification sites on soluble and membrane proteins, and (iii) the characterization of membrane protein topology and relative localization of soluble proteins. Overlapping peptides produced from digestion with the robust nonspecific protease proteinase K facilitates the identification of covalent modifications (phosphorylation and methylation). High-pH treatment disrupts sealed membrane compartments without solubilizing or denaturing the lipid bilayer to allow mapping of the soluble domains of integral membrane proteins. Furthermore, coupling protease protection strategies to this method permits characterization of the relative sidedness of the hydrophilic domains of membrane proteins.

Membrane proteins carry out many essential cellular functions[1]. However, their hydrophobic nature makes them notoriously difficult to study, keeping them consistently underrepresented in proteomic analyses[2]. Traditionally, proteomic analyses involve gel electrophoresis followed by mass spectrometry[2]. Unfortunately, this approach is limited by the difficulty of solubilizing and resolving the membrane proteins on gels; moreover, identification of gel-separated proteins by mass spectrometry is tedious even with current advances in automation. Shotgun methods provide useful alternatives to gels, whereby proteins are first digested into more complex peptide mixtures that are then analyzed directly by liquid chromatography coupled to mass spectrometry (LC-MS)[3]. Three different shotgun methods have been applied recently to analyze enriched membrane fractions using detergents[4], organic solvents[5,6], or organic acids[3] in the presence of trypsin[4–6] or CNBr[3] to solubilize membrane proteins. These methods, though effective, are limited to protein identification.

A principal objective of proteomics is the systematic identification of all proteins expressed in a cell or tissue, but comprehensive insight into protein function also requires the identification of covalent modifications. Most shotgun approaches focus on protein phosphorylations using chemical derivatizations[7–9] or affinity strategies for phosphopeptide enrichment[10]. A more versatile approach that is not restricted to the identification of a single type of modification uses multiple protease cleavages to increase the protein sequence coverage by identified peptides[11,12]. Though successful in moderately complex mixtures of soluble proteins (<300 proteins), this method is problematic when applied to whole-cell lysates and membrane-containing samples. Therefore, a more efficient strategy is required for the analysis of covalent modifications on both soluble and membrane proteins in complex membrane-containing samples.

High-pH conditions have been frequently used for biochemical fractionation of soluble proteins from integral proteins embedded in membrane bilayers[3,13,14]. Electron micrographs from the original method evaluation revealed that high pH prevents the resealing of membrane structures after mechanical agitation, favoring the presence of membrane 'sheets' with free edges[13]. Detergents, organic solvents, and organic acids solubilize membrane proteins, resulting in the loss of information about native topology[3–6]. In contrast, high pH disrupts sealed membrane structures without denaturing the lipid bilayer or extracting the integral membrane proteins[1]. A combination of high pH and nonspecific proteolytic digestion permits cleavage of the soluble domains from integral membrane proteins while preserving native topology and allowing for global mapping of hydrophilic domains. Digestion with nonspecific enzymes also increases the total number of unique and overlapping peptides produced, because the peptide cleavage sites and their corresponding proximities to the lipid bilayer are not limited by protein sequence. Furthermore, combining this method with classical protease protection strategies permits characterization of relative cellular localizations of soluble proteins as well as the orientation of membrane proteins[15,16].

We describe a method using high pH and proteinase K (hpPK) that is optimized specifically for the global analysis of both membrane and soluble proteins from membrane-containing samples. High pH favors the formation of membrane sheets, while proteinase K cleaves exposed hydrophilic domains of membrane proteins (Figs. 1, 2). Subtilisin and elastase have been reported to have sufficient activities in simple mixtures[11], but their activities are substantially diminished when applied to complex membrane-containing samples (data not shown). Proteinase K, however, is extremely robust and often results in the complete digestion of proteins to dipeptides. Fortuitously, high pH attenuates the enzyme's activity to

[1]*Department of Cell Biology, The Scripps Research Institute, La Jolla, CA 92037.* [2]*Department of Cellular and Structural Biology, University of Colorado Health Sciences Center, Denver, CO 80206.* [3]*Department of Proteomics and Metabolomics, Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121-1125.* **These authors contributed equally to this work.* †*Corresponding author (jyates@scripps.edu).*

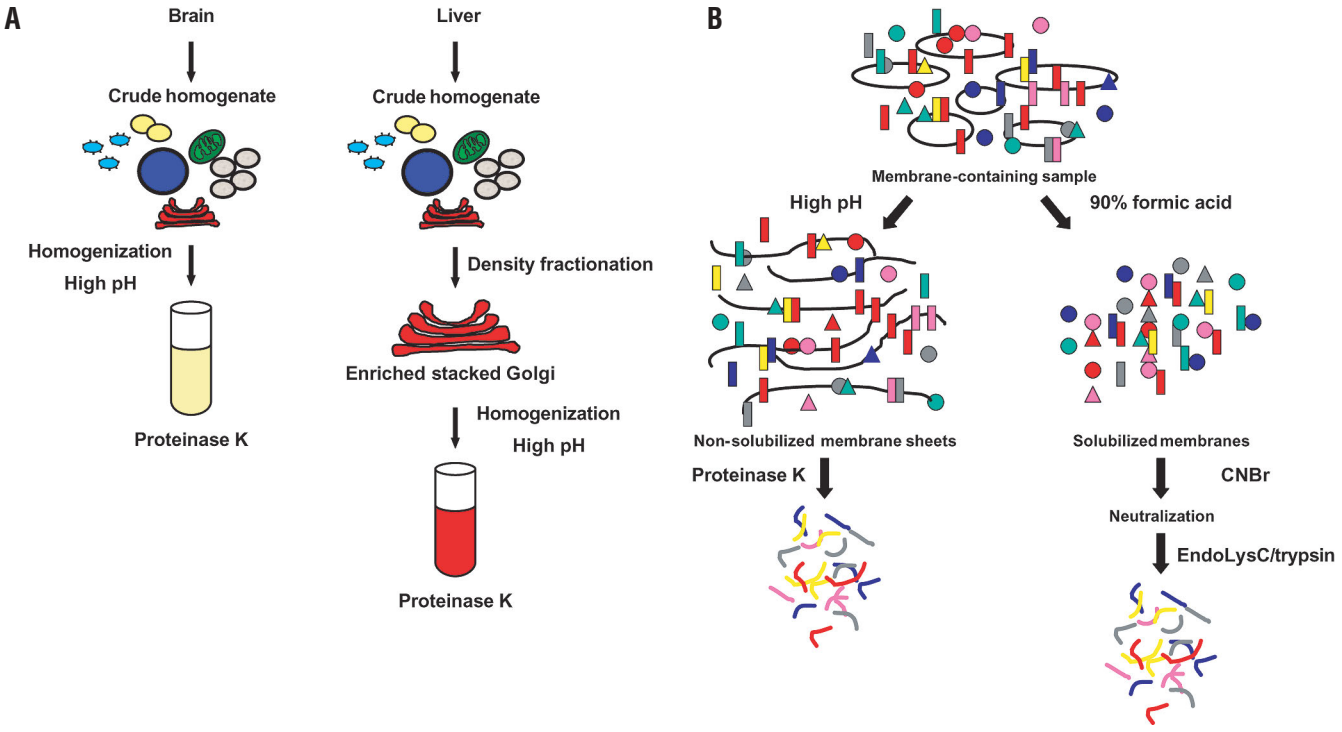

Figure 1. Application of the hpPK method to complex membrane-containing samples. (A) At left, crude unfractionated rat brain homogenates containing total cellular membranes and cytosol of all cell types within the dissected tissue are homogenized at high pH and incubated with proteinase K. At right, a stacked Golgi fraction is enriched from crude rat liver homogenates, homogenized at high pH, and incubated with proteinase K. (B) At left, treatment with high pH disrupts sealed membrane compartments to produce unsealed 'membrane sheets' with free edges. The activity of proteinase K is attenuated under these conditions, permitting the cleavage of accessible protein domains on nonsolubilized membrane sheets. At right, membranes are solubilized and proteins are digested with multiple proteolytic steps using the earlier reported formic acid–CNBr method.

levels at which peptides (6- to 20-mers) optimal for analysis by μLC/μLC-MS/MS are produced[3,17]. Here we report concurrent shotgun analysis of (i) protein identifications, (ii) post-translational modifications (PTMs), and (iii) membrane protein topology from mammalian tissue samples.

Results and discussion

Membrane protein identifications. To test the limits of complexity, we analyzed unfractionated brain homogenates which contained total cellular membranes and cytosol from all cell types present (Fig. 1A). The peptide mixture produced by hpPK treatment of the brain homogenate was analyzed by MudPIT[3], and 1,610 proteins were identified at >95% confidence[18] with a two-peptide minimum (the complete list is available in Supplementary Table 1 online). Of these, 454 proteins (28.2%) were predicted to have transmembrane domains (TMDs) using TMHMM 2.0 (ref. 19). Although many of the identified membrane proteins have only 1 TMD, most have 2–23 TMDs (54.4% of total membrane proteins) (Fig. 3A). Global genomic analyses predict that 20–30% of all open reading frames encode integral membrane proteins[20]. The percentage of identified membrane proteins in brain homogenate (28.2%) seems to approximate these predictions and leads us to conclude that the proportion of soluble and membrane proteins identified using the hpPK method is reflective of the composition of the sample.

Post-translational modifications. The success of the nonspecific protease approach of MacCoss *et al.* is related directly to the extent of redundant sequence coverage[11]. Increased redundancy reduces ambiguity in the assignment of a covalent modification to a specific residue[11]. The ability to obtain high sequence coverage for proteins identified in brain homogenate tested the upper limits of our methodology. Notably, even though the majority of the proteins had <20% sequence coverage (Fig. 3B), 79 modifications were still identified on 51 proteins (24 were assigned to 18 membrane proteins) (see Supplementary Table 2 online for full listing with confidence values)[18]. The modifications include 22 phosphorylations (16 correlated with predictions by NetPhos 2.0 prediction software[21]), 14 monomethylations, 39 dimethylations, and 4 trimethylations or acetylations.

It should be noted that assignment of these modifications is based on measurements of mass shifts in the tandem mass spectra. Therefore, confidence in the assignment is dependent on the mass accuracy of the instrument and localization of the mass shift to a specific residue. Even though all reported modifications (Supplementary Table 2) have mass shifts localized to their respective residue within the spectrum, the possibility remains that they are artifactual. Therefore, modification sites should always be confirmed using an alternate assay[12].

Comprehensive proteomic analysis of selected proteins. Membrane proteins (STX2, AQP4, GRID2, and ATP2b2) and soluble proteins (NVP2, or VSNL1, and CALM1) were chosen to demonstrate the comprehensive nature of proteomic information attained for individual proteins within the context of a shotgun analysis of unfractionated brain homogenate. STX2 and AQP4 are displayed in Figure 4 and NVP2/VSNL1, GRID2, CALM1, and ATP2b2 are available in Supplementary Figure 1 online.

STX2 (syntaxin 2) is a SNARE protein that plays a central role in exocytosis of synaptic vesicles[22]. It has one TMD at its C-terminal end (Fig. 4A). Peptides detected redundantly cover 54.2% of the protein sequence. Three peptides identify a phosphorylation on Ser14 with 99.7% confidence. This previously reported phosphorylation site is developmentally regulated in rat brain and segregates discrete domains on axonal plasma membranes to control synaptic

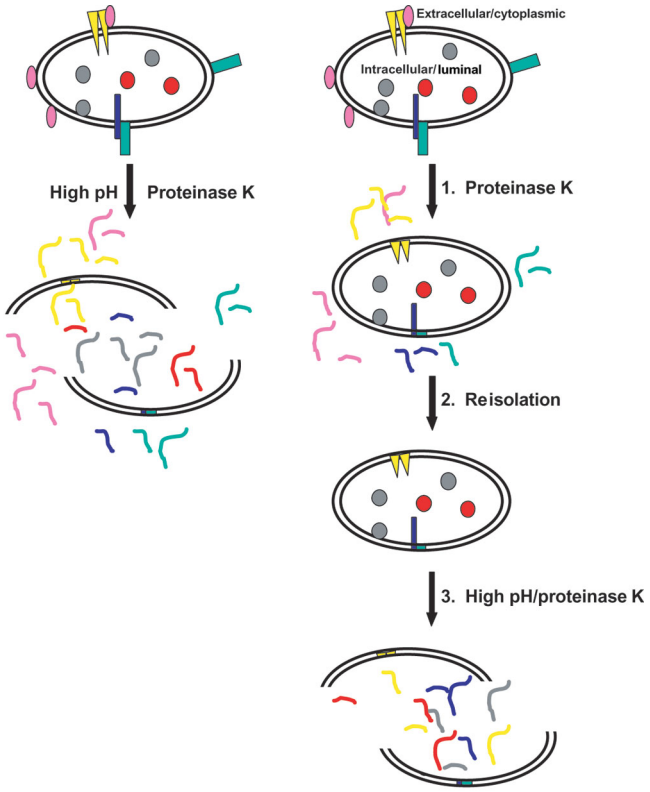

Figure 2. Characterization of membrane protein topology and relative protein localization. Application of the hpPK method results in the mapping of soluble domains on integral membrane proteins. At left, when sealed membrane compartments are agitated at high pH, unsealed membrane sheets are enriched and provide access to luminal and external domains and proteins. At right, through a protease protection approach, the method can be separated into sequential steps resulting in characterization of membrane protein topology and relative protein localization. Samples containing sealed membrane compartments are digested with proteinase K to remove all externally accessible protein domains. The membranes are reisolated and homogenized at high pH to unseal the membranes and enrich for membrane sheets. Subsequent digestion with proteinase K cleaves previously protected domains of membrane proteins and protected soluble luminal proteins.

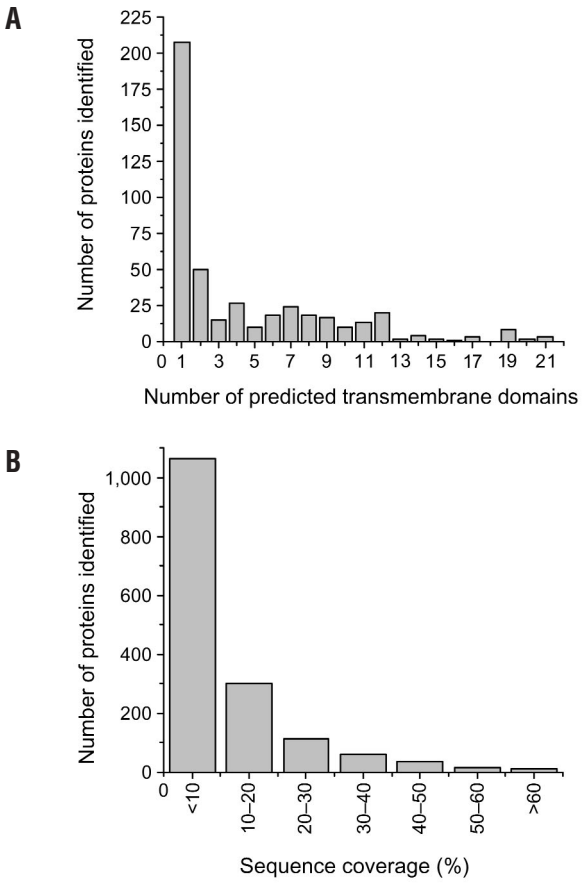

Figure 3. Distribution of transmembrane domains and total sequence coverage in identified brain proteins. (A) The prediction software TMHMM 2.0 was used to predict the number of TMDs in each of the identified brain proteins. The distribution of predicted TMDs on membrane proteins identified is plotted. (B) The percentage of total identified proteins is plotted against the percentage sequence coverage. Most of the identified brain proteins had <20% sequence coverage.

fusion[22]. The tandem mass spectrum of the nontryptic +2 phosphopeptide KDS*DDEEEVVHVD (S* designates modified residue) shows a characteristic loss of -H_2PO_4 $[M-98]^{2+}$ from the precursor ion during fragmentation. Fragment ions representative of the entire phosphopeptide sequence are present in both the b- and y-ion series, resulting in an unambiguous modification assignment.

AQP4 (aquaporin 4) is the predominant water channel in brain and is expressed only on the basolateral surface of epithelial cells[23,24]. It has six predicted TMDs spaced throughout the protein (Fig. 4B). Although coverage for the entire protein sequence is only 26.3%, the localized redundant sequence coverage remains high (45.7%) on the C-terminal soluble domain. Five identified peptides overlap into the C-terminal-most predicted TMD (residues shown in red), suggesting that the domain is exposed and not embedded within the bilayer as predicted. Reported phosphorylations of Ser180 and Ser276 control its mechanism of action[24] and cell surface localization[23], respectively, and were not detected in our analysis owing to a lack of coverage. However, three peptides identify a previously unreported phosphorylation on Ser285 with 98.0% confidence. The tandem mass spectrum of the +2 peptide DNRS*QVETEDLILKPGVV represents a classical phosphoserine tandem mass spectrum with the base peak fragment ion resulting from the loss of –H_2PO_4. The spectrum, though dominated by a single fragment ion, matches the phosphorylated peptide DNRS*QVETEDLILKPGVV with a SEQUEST XCorr of 3.24.

Global analysis of protein topology and localization. Protease protection assays have traditionally been used to determine the sidedness of proteins relative to a sealed membrane compartment[15,16]. These experiments typically involve the complete digestion of exposed domains or proteins on the outside of the sealed compartment in the presence or absence of detergents, with subsequent detection of the protected domains or proteins by gel electrophoresis, western blotting, or both[25–28]. A global protease protection assay was conducted to analyze protein topology and relative localization by differentially digesting exposed and protected domains of membrane proteins using the hpPK method separated into three distinct steps (Fig. 2, right). Proteinase K is added to membrane-containing samples at neutral pH in step 1. Both luminal soluble proteins and luminal domains of integral membrane proteins are protected from proteolytic digestion by the sealed lipid bilayer. Digested intact membrane structures are reisolated in step 2. Treating with the one-step hpPK in step 3 permits unsealing of membranes, digestion of luminal domains or proteins, and analysis of peptides by MudPIT. It is important to note that the relative sidedness of membrane protein is reflective of its cellular localization. For example, cell surface membrane proteins are found in the membranes of the secretory

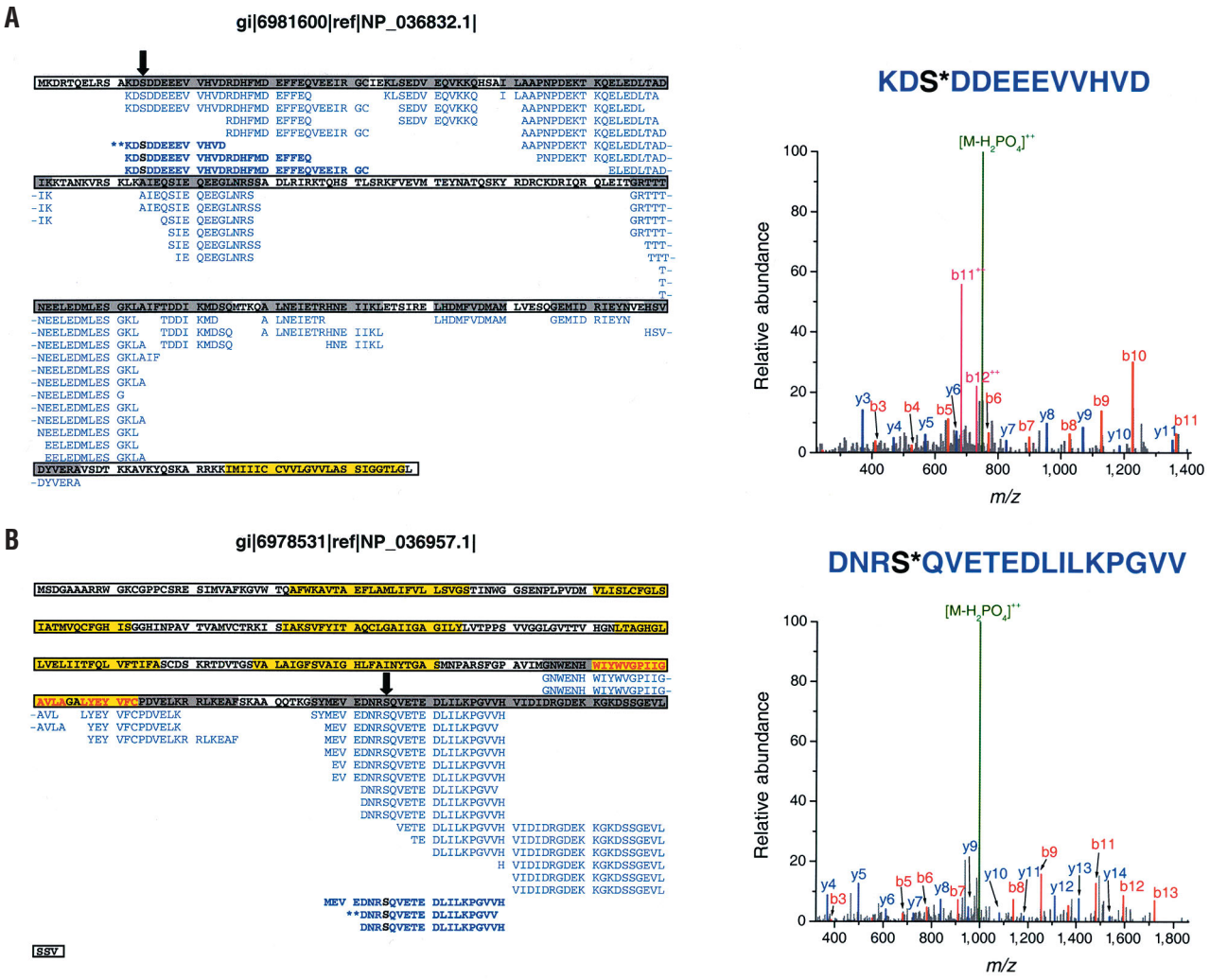

Figure 4. Comprehensive characterization of individual proteins from unfractionated rat brain homogenates. Crude brain homogenates were digested with proteinase K at high pH and analyzed by MudPIT. Both membrane and soluble proteins were identified, and several post-translational modifications were assigned. Overall topology was characterized with the mapping of soluble domains on membrane proteins. (A) Characterization of STX2. (B) Characterization of AQP4. In both (A) and (B), protein sequences of identified proteins are displayed on the left (boxed sequence). Regions shaded in gray indicate coverage by identified peptides. Regions shaded in yellow indicate predicted TMDs. Peptides detected are displayed below the protein sequence in blue text. Modified peptides are displayed in bold blue text, and modification sites are indicated on the protein sequence with an arrow. Spectra are displayed on the right for modified peptides tagged with **. Spectra are annotated using Roepstorff and Fohlman nomenclature[38]. Four additional examples can be found in Supplementary Figure 1 online.

pathway *en route* to the plasma membrane. A hydrophilic domain of a membrane protein found on the luminal side while trafficking through the Golgi becomes the extracellular domain at the cell surface. The exclusive use of enriched membrane fractions minimizes these ambiguities for localization.

In a proteome-wide protease protection experiment using enriched liver Golgi membranes (Fig. 1A)[29], soluble and membrane proteins were analyzed concurrently using two proteomic analyses. In the first, the sample was treated with the one-step hpPK method (Fig. 2, left) to identify peptides from all soluble proteins and all hydrophilic domains of membrane proteins. In the second analysis, the sample was prepared using the three-step protease protection strategy (Fig. 2, right) to identify peptides only from protected protein domains. A comparison of the results of these two analyses facilitated deductions about relative protein topology and protein localization. Most bona fide resident Golgi membrane proteins were found in the correct predicted orientation. The rat liver Golgi proteome is beyond the scope of this paper and will be presented elsewhere. However, to illustrate the application of our method, results for three proteins will be discussed here: SIAT1 (data in Fig. 5), and RPN1 and NUCB2 (data available in Supplementary Figure 2 online).

SIAT1 (sialyltransferase 1) is a resident Golgi type II membrane protein involved in the terminal sialylation of N-linked carbohydrate groups of glycoproteins and glycolipids[30]. It was identified with peptides covering 46.9% of the protein sequence (Fig. 5). Use of the one-step hpPK method (Fig. 2, left) allowed detection of peptides over the entire C-terminal soluble domain. One of these peptides, a +2 HLNEGTDEDIY*LFGK, identified a phosphorylation on Tyr388 with 90.0% confidence. The tandem mass spectrum has prominent fragment ions covering most of the peptide sequence, and fragment ions b10, b11, and y5 localize the phosphorylation to the tyrosine residue. Because the majority of the SIAT1 protein sequence is located on the luminal side of the *trans*-Golgi membrane, protease-protected peptides were detected for most of the protein (orange).

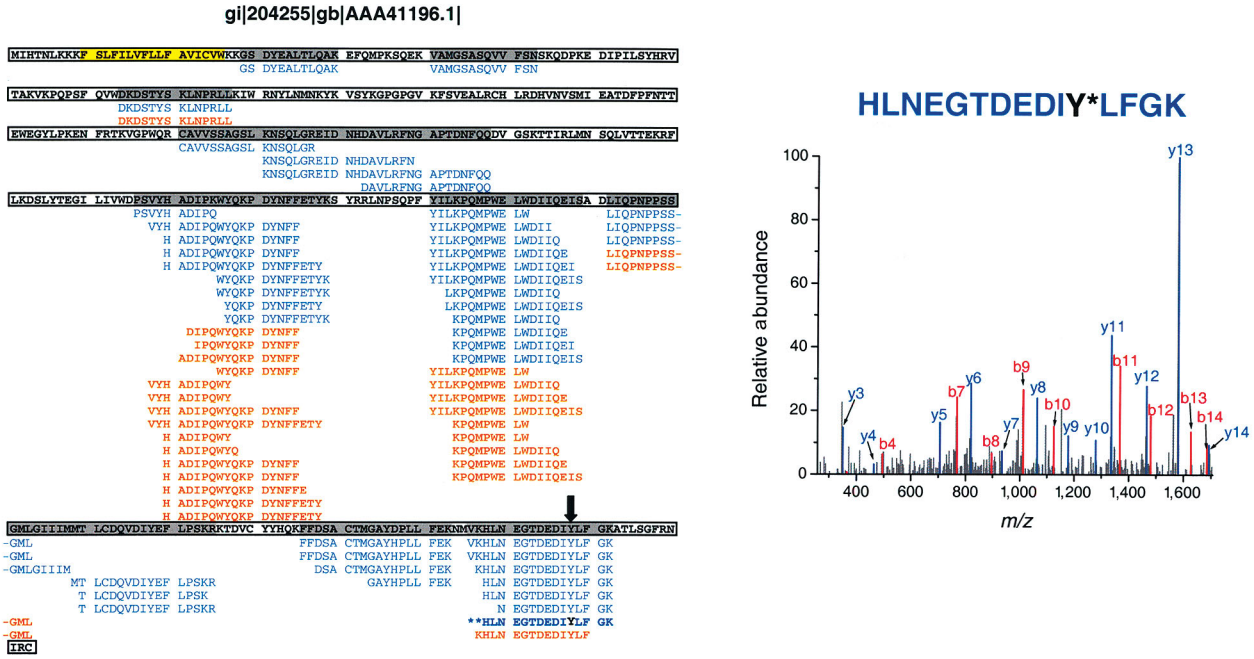

Figure 5. Comprehensive characterization of a Golgi membrane protein from a global protease protection analysis. Global protease protection was conducted on enriched Golgi fractions by digestion with proteinase K, reisolation, and hpPK treatment. Two experiments (Fig. 2) were carried out and analyzed by MudPIT. Both membrane and soluble proteins were identified, and some post-translational modifications were assigned. Overall topology and relative localization was characterized with the mapping of detected peptides onto the protein sequence. The protein sequence of *trans*-Golgi membrane protein SIAT1 is displayed on the left (boxed text). Regions shaded in gray indicate coverage by identified peptides. Regions shaded in yellow indicate predicted TMDs. Peptides detected are displayed below the protein sequence in blue text. The modified peptide is displayed in bold blue text and its spectrum is displayed on the right. Modified tyrosine residue is indicated on the protein sequence with an arrow. Protected peptides detected from the protease protection experiment (Fig. 2, left) are displayed below the protein sequence in bold orange text.

Conclusions

The hpPK method was designed specifically to facilitate shotgun proteomic analysis of membrane proteins in complex biological samples. This methodology was used to analyze unfractionated brain homogenates and resulted in the identification of 1,610 proteins with the predicted native cellular proportions of soluble to membrane proteins (71.8% and 28.2%). Other methods have also been optimized for the identification of membrane proteins[3–5]. However, these methods solubilize the membrane proteins and use specific proteases and thus lack the versatility to provide information on covalent modifications and membrane protein topology.

The hpPK method is rapid and robust, and capable of both identifying covalent modifications and characterizing membrane protein topologies. However, these results are acquired only if protein sequence coverage is high, with redundant overlapping peptides. This is typically the case when a protein is abundant or enriched[31], and it therefore seemed unlikely that any modifications would be detected in a sample as complex as tissue homogenate. In fact, a total of 79 modifications were identified on highly abundant proteins. Notably, several members of the seven membrane-spanning receptor family were also identified, including the serotonin, GABA, and glutamate receptors and several G-protein-coupled receptors (GPCRs)[32]. The coverage on these proteins was understandably low in the context of the total brain homogenate and therefore did not result in the acquisition of information on covalent modifications. However, future applications of this method on highly enriched membrane samples should result in more comprehensive data. Although the methodology is not quantitative as presented here, it can readily be coupled with existing quantitative proteomic approaches such as metabolic labeling[33,34] and chemical derivatization[6,35].

Experimental protocol

Materials. Male Sprague-Dawley rats (~250 g) were purchased from Harlan (Indianapolis, IN). Methods involving animals were approved by the institutional Animal Research Committee (accredited by the American Association for Accreditation of Laboratory Animal Care). All chemicals were purchased from Sigma (St. Louis, MO) unless otherwise noted. Enzymes were purchased from Roche Applied Science (Indianapolis, IN).

Sample preparation. *Brain homogenate:* Rats were killed by halothane inhalation, and brains were removed and dissected. Region-specific tissue (cortex, cerebellum, striatum/hypothalamus, hippocampus) was homogenized at a 1:10 ratio (tissue to buffer) in ice-cold homogenization buffer (100 mM K_2HPO_4/KH_2PO_4, pH 6.7, 5 mM $MgCl_2$, 250 mM sucrose) using 30 strokes in a tight-fitting Dounce homogenizer. The homogenate was centrifuged at low speed (3,000 × *g* for 15 min at 4 °C). The supernatant (crude tissue homogenate) was collected and adjusted to 1 mg/ml with homogenization buffer for proteomic analysis. Protein concentration was by Lowry H Protein Assay (Bio-Rad, Hercules, CA).

Fractionated Golgi: Enriched Golgi fractions were prepared from livers of cycloheximide-treated rats using a method described earlier[29].

Sample digestion. Tissue homogenates (500 µg protein) were diluted twofold with 100 mM K_2HPO_4/KH_2PO_4, pH 6.7, and pelleted by centrifuging at 18,000 × *g* at 4 °C for 30 min. After resuspension at 1 mg/ml in 200 mM Na_2CO_3, pH 11, with five passes through an insulin syringe, the pellet was incubated on ice for 1 h. The membrane sample was then adjusted to 8 M urea and reduced and alkylated as reported before[3]. Proteinase K (5 µg) was added to the sample and incubated at 37 °C for 3 h in a Thermomixer (Brinkmann, Westbury, NY). An additional aliquot of proteinase K (5 µg) was added and incubated at 37 °C for 1.5 h. The reaction is quenched with formic acid (5% final concentration) and microcentrifuged at 18,000 × *g* at 4 °C for 15 min to remove particulates.

Fractionated Golgi: Golgi membranes were prepared using the method described earlier for brain homogenates. Additional Golgi samples were prepared as described by Washburn *et al*[3].

Protease protection. Proteinase K at a 1:100 mass-to-mass ratio of enzyme to substrate was added to 1 mg/ml Golgi membranes in 100 mM K_2HPO_4/KH_2PO_4, pH 6.7, 5 mM $MgCl_2$, 250 mM sucrose, and incubated on a rotator at 4 °C for 30 min. The sample was diluted two-fold with 100 mM K_2HPO_4/KH_2PO_4, pH 6.7 and layered on top of a sucrose cushion (100 mM K_2HPO_4/KH_2PO_4, pH 6.7, 5 mM $MgCl_2$, 250 mM sucrose). Membranes were reisolated by microcentrifuging at 18,000 × *g* for 30 min at 4 °C. The membrane pellet was resuspended at ~1 mg/ml in 200 mM Na_2CO_3, pH 11, and prepared as reported earlier under Sample Digestion.

Multidimensional protein identification technology (MudPIT). The protein digest was pressure-loaded onto a fused-silica capillary desalting column containing 5 cm of 5-µm Polaris C18-A material (Metachem, Ventura, CA) packed into a 250-µm inner diameter (i.d.) capillary with a 2-µm filtered union (UpChurch Scientific, Oak Harbor, WA). The desalting column was washed with buffer containing 95% water, 5% acetonitrile, and 0.1% formic acid (all vol/vol). The desalted proteins were then eluted onto the rear end of a triphasic chromatography column using 20% water, 80% acetonitrile, and 0.1% formic acid. The triphasic column consisted of a 100-µm i.d. capillary with a 5-µm pulled tip and was packed in the following order from the tip: (i) 7 cm 5-µm Aqua C18 material (Phenomenex, Ventura, CA), (ii) 3 cm 5-µm Partisphere strong cation exchanger (Whatman, Clifton, NJ), and (iii) 3 cm 5-µm hydrophilic interaction chromatography material (PolyLC, Columbia, MD).

Once loaded with the peptide digests, the column was placed inline with an Agilent 1100 quaternary HPLC (Palo Alto, CA) and analyzed using a modified 12-step separation described earlier[3]. The buffer solutions used were 5% acetonitrile–0.1% formic acid (buffer A), 80% acetonitrile–0.1% formic acid (buffer B), and 500 mM ammonium acetate–5% acetonitrile–0.1% formic acid (buffer C) (all vol/vol). Step 1 consisted of a 100-min gradient from 0 to 100% buffer B. Steps 2–11 had the following profile: 3 min of 100% buffer A, 2 min of *x*% buffer C, a 10-min gradient from 0 to 15% buffer B, and a 97-min gradient from 15% to 45% buffer B. The 2-min buffer C percentages (*x*) were 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, respectively, for the 12-step analysis. For the final step, the gradient contained 3 min of 100% buffer A, 20 min of 100% buffer C, a 10-min gradient from 0 to 15% buffer B, and a 107-min gradient from 15% to 70% buffer B.

As peptides eluted from the microcapillary column, they were electrosprayed directly into an LCQ-Deca mass spectrometer (ThermoFinnigan, Palo Alto, CA) with the application of a distal 2.4 kV spray voltage. A cycle of one full-scan mass spectrum (400–1,400 *m*/*z*) followed by three data-dependent MS/MS spectra at a 35% normalized collision energy was repeated continuously throughout each step of the multidimensional separation. Application of mass spectrometer scan functions and HPLC solvent gradients were controlled by the Xcaliber data system (ThermoFinnigan, Palo Alto, CA).

Analysis of tandem mass spectra. MS/MS spectra were analyzed using the following software analysis protocol. 2to3 determined the charge state (+2 or +3) of multiply charged peptide spectra and deleted poor-quality spectra. Each MS/MS spectrum after 2to3 was searched against the RefSeq protein database (rat, mouse, human sequences) using SEQUEST[36]. DTASelect selected peptide sequences from +1, +2, and +3 charged peptide precursors with normalized SEQUEST XCorr scores >0.3 (ref. 18) and $\Delta C_n > 0.1$. To minimize false positives, only proteins with two or more peptides exceeding the peptide filters were considered. DTASelect then assembled the peptide sequences into proteins and removed redundant protein sequences[37]. For example, if ten different peptides identified a gene locus and three of the ten were also present in another gene locus, only the locus with the greater number of peptides was listed, and the subset locus was removed. If all ten peptides were identified in two gene loci, both loci were listed but only counted as single protein identifications. The resulting protein list was used to create a subset database to expedite SEQUEST differential modification searches. The MS/MS spectra were then re-searched four times against the subset database to consider modifications of: (i) +80 on STY (phosphorylation), (ii) +14 on KRH (methylation), (iii) +28 on KR (dimethylation), and (iv) +42 on K (trimethylation, acetylation, or carbamylation). Carbamylations can occur when urea is used as a denaturant, and because a mass shift of +43 Da (carbamylation) cannot be distinguished from +42 Da (trimethylation or acetylation), seemingly real identifications of trimethylations or acetylations of lysine residues could potentially be artifactual. All searches were parallelized and done on a Beowulf computer cluster consisting of 34 1.2-GHz Athlon computer processing units. No enzyme specificity was considered for any search.

The MS/MS spectra for the modified peptides were manually evaluated using criteria reported earlier[17]. Modified peptide spectra exceeding these criteria were re-searched using SEQUEST against the NCBI nonredundant protein database. Confidence for modifications was estimated from overlapping modified peptides as described elsewhere[18].

Note: Supplementary information is available on the Nature Biotechnology website.

Acknowledgments

The authors gratefully acknowledge financial support from the American Cancer Society PF-03-065-01-MGO (C.C.W.) and the National Institute of Health grants F32DK59731 (M.J.M.), RO1-GM42629 (K.E.H) and R33 CA81665 and RR11823 (J.R.Y.).

Competing interests statement

The authors declare that they have no competing financial interests.

Received 2 November 2002; accepted 11 February 2003

1. Alberts, B. *et al. Molecular Biology of the Cell* (Garland Science, New York, 2002).
2. Santoni, V., Molloy, M. & Rabilloud, T. Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054–1070 (2000).
3. Washburn, M.P., Wolters, D. & Yates, J.R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247 (2001).
4. Han, D.K., Eng, J., Zhou, H. & Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **19**, 946–951 (2001).
5. Blonder, B. *et al.* Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography–tandem mass spectrometry. *J. Proteome Res.* **1**, 351–360 (2002).
6. Goshe, M.B., Blonder, B. & Smith, R.D. Affinity labeling of highly hydrophobic integral membrane proteins for proteome-wide analysis. *J. Proteome Res.* **2**, 153–161 (2003).
7. Zhou, H., Watts, J.D. & Aebersold, R. A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **19**, 375–378 (2001).
8. Oda, Y., Nagasu, T. & Chait, B.D. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19**, 379–382 (2001).
9. Goshe, M.B. *et al.* Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analysis. *Anal. Chem.* **73**, 2578–2586 (2001).
10. Ficarro, S.B. *et al.* Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **20**, 301–305 (2002).
11. MacCoss, M.J. *et al.* Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. USA* **99**, 7900–7905 (2002).
12. Cheeseman, I.M. *et al.* Phospho-regulation of kinetochore–microtubule attachments by the aurora kinase ipl1p. *Cell* **111**, 163–172 (2002).
13. Howell, K.E. & Palade, G.E. Hepatic Golgi fractions resolved into membrane and content subfractions. *J. Cell Biol.* **92**, 822–832 (1982).
14. Taylor, R.S. *et al.* Proteomics of rat liver Golgi complex: minor proteins are identified through sequential fractionation. *Electrophoresis* **21**, 3441–3459 (2000).
15. Blobel, G. & Sabatini, D.D. Controlled proteolysis of nascent polypeptides in rat liver cell fractions. I. Location of the polypeptides within ribosomes. *J. Cell Biol.* **45**, 130–145 (1970).
16. Sabatini, D.D. & Blobel, G. Controlled proteolysis of nascent polypeptides in rat liver cell fractions. II. Location of the polypeptides in rough microsomes. *J. Cell Biol.* **45**, 146–157 (1970).
17. Link, A.J. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676–682 (1999).
18. MacCoss, M.J., Wu, C.C. & Yates, J.R., III. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal. Chem.* **74**, 5593–5599 (2002).
19. Moller, S., Croning, M.D.R. & Apweiler, R. Evaluation of methods for the prediction of membrane-spanning regions. *Bioinformatics* **17**, 646–653 (2001).
20. Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7**, 1029–1038 (1998).
21. Blom, N., Gammeltoft, S. & Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351–1362 (1999).
22. Foletti, D.L., Lin, R., Finley, M.A. & Scheller, R.H. Phosphorylated syntaxin 1 is localized to discrete domains along a subset of axons. *J. Neurosci.* **20**, 4535–4544 (2000).
23. Madrid, R. *et al.* Polarized trafficking and surface expression of the AQP4 water channel are coordinated by serial and regulated interactions with different clathrin–adaptor complexes. *EMBO J.* **20**, 7021 (2001).
24. Zelenina, M., Zelenin, S., Bondar, A.A., Brismar, H. & Aperia, A. Water permeability of aquaporin-4 is decreased by protein kinase C and dopamine. *Am. J. Physiol. Renal Physiol.* **283**, F309–F318 (2002).
25. Sprong, H. *et al.* UDP-galactose:ceramide galactosyltransferase is a class I integral membrane protein of the endoplasmic reticulum. *J. Biol. Chem.* **237**, 25880–25888 (1998).
26. Ring, G. & Eichler, J. Characterization of inverted membrane vesicles from the halophilic archaeon *Haloferax volcanii*. *J. Membr. Biol.* **183**, 195–204 (2001).
27. Kawano, J. *et al.* CALNUC (nucleobindin) is localized in the Golgi apparatus in insect cells. *Eur. J. Cell Biol.* **79**, 16167–16173 (2000).
28. Morel-Huaux, V.M. *et al.* The calcium-binding protein p54/NEFA is a novel luminal resident of medial Golgi cisternae that trafficks independently of mannosidase II. *Eur. J. Cell Biol.* **81**, 87–100 (2002).
29. Taylor, R.S., Jones, S.M., Dahl, R.H., Nordeen, M.H. & Howell, K.E. Characterization of the Golgi complex cleared of proteins in transit and examination of calcium uptake activities. *Mol. Biol. Cell* **8**, 1911–1931 (1997).
30. Wen, D.X., Svensson, E.C. & Paulson, J.C. Tissue-specific alternative splicing of the β-galactoside α2,6- sialyltransferase gene. *J. Biol. Chem.* **267**, 2512–2518 (1992).
31. Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
32. Pierce, K.L., Premont, R.T. & Lefkowitz, R.J. Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.* **3**, 639–650 (2002).
33. Oda, Y., Huang, K., Cross, F.R., Cowburn, D. & Chait, B.T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA* **96**, 6591–6596 (1999).
34. Washburn, M.P., Ulaszek, R., Deciu, C., Schieltz, D.M. & Yates, J.R., III. Analysis of quantitative proteomic data generated *via* multidimensional protein identification technology. *Anal. Chem.* **74**, 1650–1657 (2002).
35. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (2000).
36. Eng, J.K., McCormack, A.L. & Yates, J.R., III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
37. Tabb, D.L., McDonald, W.H. & Yates, J.R., III. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**, 21–26 (2002).
38. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601 (1984).