Proteomics on Full-Length Membrane Proteins Using Mass Spectrometry†

Johannes le Coutre,*,‡ Julian P. Whitelegge,§,|| Adrian Gross,⊥ Eric Turk,@ Ernest M. Wright,@ H. Ronald Kaback,‡ and Kym F. Faull||,#

Howard Hughes Medical Institute, Departments of Physiology and of Microbiology and Molecular Genetics, Molecular Biology Institute, Department of Chemistry and Biochemistry, The Pasarow Mass Spectrometry Laboratory, Jules Stein Eye Institute, Department of Physiology, Department of Psychiatry and Biobehavioral Sciences, and Neuropsychiatric Institute, University of California, Los Angeles, California 90095

Received January 24, 2000

ABSTRACT: A general technique has been developed that allows rapid mass spectrometric analysis of full-length membrane proteins [Whitelegge, J. P., le Coutre, J., et al. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 10695−10698]. Using in-line HPLC electrospray ionization mass spectrometry (LC−MS), different native and recombinant bacterial membrane proteins of up to 61 kDa are characterized. Mass spectrometric data of four entirely different membrane proteins from three bacterial organisms, two transporters, a channel, and a porin protein are presented. In addition to determination of the molecular mass with an accuracy of ±0.01%, the technique monitors alkylation or oxidation of single Cys residues and errors in deduced amino acid sequences. Finally, using in-line LC−MS, unknown proteins can be identified from solubilized *Escherichia coli* membranes without prior purification.

Membrane proteins are notoriously resistant to structure determination by X-ray crystallography, and other high-resolution techniques generally cannot be applied to this class of proteins. Recent advances in molecular biology and biochemistry have led to rapid progress in understanding structure−function relationships for some membrane proteins, and in about a dozen cases, structures at atomic resolution have been determined. Nevertheless, the level of understanding of these proteins is almost inversely proportional to the roles they play in cell metabolism and viability. As an increasing number of genomes are sequenced and an equally growing number of membrane proteins are identified, this discrepancy will likely increase. Moreover, in the postgenomic era, with proteomics emerging as a new field directed at the structural and functional assessment of the entire complement of proteins in a given organism, the limitations in working with large membrane proteins will become even more disadvantageous.

The electrospray ionization-mass spectrometry (ESI-MS)[1] spectrum of a native or recombinant full-length protein provides a unique and precise image of its covalent state (*1*, *2*). However, the ability to perform mass spectrometric analysis on full-length membrane proteins has been hampered by the amphipathic nature of these proteins and the resulting inability to ionize them sufficiently in the absence of salt or detergent. Thus, until recently, mass spectrometry has mostly been carried out with protein fragments generated by either enzymatic or chemical cleavage (*3*, *4*). By using liquid chromatography−mass spectrometry (LC−MS) with an appropriate solvent system in combination with ESI-MS, these problems have been resolved (*1*, *2*). Highly accurate determination of the molecular mass of a purified, full-length protein with a precision exceeding 0.01% is the most straightforward method for assessing purity and homogeneity. Furthermore, the technique can be used to detect alterations such as oxidation or post-translational modification. Also, selective chemical modification of specific residues can be used to gain mechanistic information about the protein of interest. Finally, in organisms where the entire genome has been sequenced, chromatographic separation of crude protein extracts and subsequent ESI-MS can identify new proteins without prior purification.

In this paper, several of these issues are addressed independently to demonstrate the broad applicability of the technique: (i) controlled alkylation of a Cys residue with 2-(4′-maleimidylanilino)naphthalene-6-sulfonic acid (MIANS) in the lactose permease from *Escherichia coli* and development of a simple substrate binding assay for the wild-type protein, (ii) determination of spin labeling efficiency in the *Streptomyces lividans* K^+ channel KcsA, (iii) detection of errors deduced amino acid sequences and determination

† The W. M. Keck Foundation provided support toward instrument purchase. The work was also supported in part by National Institutes of Health Grants R01 DK51131-01 (to H.R.K.), DK44602 (to E.M.W.), and GM58568 (to A.G.).

* To whom correspondence should be addressed: HHMI/UCLA, 6-720 MacDonald Research Laboratories, Box 951662, Los Angeles, CA 90095-1662. E-mail: lecoutre@hhmi.ucla.edu. Phone: (310) 206-5055. Fax: (310) 206-8623.

‡ Howard Hughes Medical Institute, Departments of Physiology and of Microbiology and Molecular Genetics, and Molecular Biology Institute.

§ Department of Chemistry and Biochemistry.

|| The Pasarow Mass Spectrometry Laboratory.

⊥ Jules Stein Eye Institute.

@ Department of Physiology.

Department of Psychiatry and Biobehavioral Sciences and Neuropsychiatric Institute.

[1] Abbreviations: DDM, *n*-dodecyl β-D-maltopyranoside; ESI-MS, electrospray ionization mass spectrometry; MALDI-TOF MS, matrix-assisted laser desorption time-of-flight mass spectrometry; MIANS, 2-(4′-maleimidylanilino)naphthalene-6-sulfonic acid, sodium salt; TDG, β-D-galactopyranosyl 1-thio-β-D-galactopyranoside; NEM, *N*-ethylmaleimide.

10.1021/bi000150m CCC: $19.00
Published on Web 03/21/2000

of the correct N-terminus in a Na^+/glucose symporter from *Vibrio parahaemolyticus* (vSGLT), and (iv) identification of proteins from a crude *E. coli* membrane extract.

EXPERIMENTAL PROCEDURES

Protein Preparation and Purification. All purified membrane proteins were produced as recombinant species in *E. coli* and isolated by using a poly-His affinity tag and nickel chelate chromatography (*2*, *5*).

Total E. coli Membrane Protein Extract. An *E. coli* membrane fraction, obtained by shearing cells in a French pressure cell and centrifugation (*2*), was washed with urea [5 M urea, 50 mM KP_i (pH 7.5), 0.5 mM Pefabloc, and 10 mM β-mercaptoethanol (BME)], stirred for 30 min on ice, and centrifuged for 2 h at 170000g_{max}. The supernatant was discarded; the membranes were resuspended in cold 50 mM KP_i (pH 7.2), 0.2 M NaCl, 10 mM BME, and 0.25 mg/mL phospholipid (3:1 w/w phosphatidylethanolamine/phosphatidylglycerol) (1.0 mL of buffer/g of cells), and *n*-dodecyl β-D-maltopyranoside (DDM) was added to a final concentration of 2% while stirring on ice for 45 min. After removal of insoluble material by centrifugation at 150000g_{max} for 30 min, the supernatant was subjected to the sample preparation procedure for ESI-MS described below.

Alkylation of Purified Proteins. MIANS labeling of purified lac permease used for ESI-MS was achieved by incubating purified protein in DDM for 10 min at room temperature at the indicated MIANS concentration. Cys residues in KcsA were spin-labeled with (1-oxy-2,2,5,5-tetramethylpyrrolinyl-3-methyl)methanethiosulfonate (RI) as described previously (*5*). Spin-labeling results in a mass increment of 184.28 Da per label.

Sample Preparation for ESI-MS. Typically, aliquots of purified membrane proteins in DDM were precipitated with $CHCl_3$/MeOH, essentially as described previously (*6*). An aliquot (100 μL) of the aqueous protein solution at a protein concentration of 1−2 mg/mL was diluted 1:3 (v/v) with MeOH and mixed briefly. $CHCl_3$ (100 μL) was added, and mixing yielded a single phase. Phase separation was accomplished by addition of 200 μL of water and vigorous mixing. The phases were separated by centrifugation (10000g_{max} for 2 min), yielding a precipitate at the interface. The bulk of the upper aqueous methanol phase was then carefully removed, and methanol (300 μL) was added. After being mixed, the sample was centrifuged (10000g_{max} for 1 min), and the pellet was dried for 1−2 min with the tube inverted prior to dissolving the sample in 50 μL of 90% formic acid. To avoid protein formylation, samples were then immediately subjected to HPLC (*2*).

HPLC. For final purification, HPLC was used prior to ESI-MS in an in-line setup. Purified and solubilized proteins were separated from low-molecular mass impurities by isocratic elution (0.25 mL/min) through a size exclusion column (SUPER SW 2000, TosoHaas, 4 μm, 300 mm × 4.6 mm). A mixture of degassed $CHCl_3$, MeOH, and 1% aqueous formic acid (4:4:1 v/v/v) was used as the solvent. The column was equilibrated extensively and tested with blank injections of 90% formic acid prior to experiments.

A polystyrene/divinylbenzene column (PLRP/S, Polymer Labs) was used for separation of the DDM membrane extract. Gradient elution using 1% aqueous formic acid as solvent A and 1% formic acid in CH_3CN as solvent B was applied. Starting at 5% B for the first 5 min, the gradient increased to 100% B after 55 min. All chromatographic separations were performed at 40 °C using a modified ABI120A dual-syringe pump machine equipped with a post-detector splitter for back-pressure regulation and fraction collection.

Cyanogen Bromide Cleavage and MALDI-TOF MS of HPLC Fractions. Selected HPLC fractions were dried in a vacuum centrifuge and resuspended in an equal amount of a saturated CNBr solution in pure formic acid (*7*). After addition of 40% water (v/v), the solution was incubated for 3 h at room temperature. The sample was dried again and analyzed by standard MALDI-TOF MS procedures.

ESI-MS. ESI-MS analysis was performed using a Perkin-Elmer Sciex API III+ triple-quadrupole instrument operating in the positive ion mode as described previously (*1*). An orifice potential of 75 V was found to be optimal for all proteins studied here.

Fluorescence Spectroscopy. Fluorescence measurements were performed on a SLM 8000C spectrofluorometer essentially as described in ref *8*. Briefly, time courses of MIANS labeling were recorded at an emission wavelength of 415 nm on purified wild-type lac permease at a concentration of 200 μg/mL in 50 mM KP_i (pH 7.5) and 0.016% DDM with an excitation wavelength of 330 nm. To demonstrate substrate binding, β-D-galactopyranosyl 1-thio-β-D-galactopyranoside (TDG) was added to a final concentration of 20 mM from a stock solution in water (870 mM).

RESULTS AND DISCUSSION

MIANS Alkylation and Ligand Binding by Lactose Permease. The lactose permease from *E. coli*, encoded by the lacY gene, is a polytopic membrane transport protein that catalyzes the coupled stoichiometric translocation of H^+ and galactosides as a monomer (*9*). The permease which is composed of 417 amino acid residues is a highly flexible protein (see ref *10*) containing 12 helical segments that traverse the membrane in zigzag fashion with both the N- and C-termini on the cytoplasmic side (*11*). Cys148 in helix V which is protected against alkylation by ligand is a component of the substrate binding site in addition to the irreplaceable residues Glu126 in helix IV and Arg144 in helix V (*12*, *13*). Typically, substrate binding is studied by inhibition of Cys148 alkylation with either NEM or MIANS in a functional protein devoid of all other Cys residues (*8*, *13*−*15*).

The measured molecular mass of unlabeled wild-type permease formylated at the N-terminus is 47 356 Da which is in excellent agreement with the calculated molecular mass of 47 355 Da (*2*) (Figure 1). At a low MIANS concentration (4 μM), most of the protein remains unlabeled, and only a small peak at 47 752 Da reflecting an increment in the molecular mass of 396 Da (the calculated molecular mass of the MIANS adduct is 393.38 Da) indicates labeling of Cys148 in a small fraction of the population of molecules. At this MIANS concentration, preincubation of the sample with 20 mM TDG fully protects against labeling, and only the unlabeled species is observed. At higher MIANS concentrations (8 μM), the pattern is similar, but a larger fraction of the sample is labeled in the absence of substrate and full protection is still observed. Labeling with 40 μM

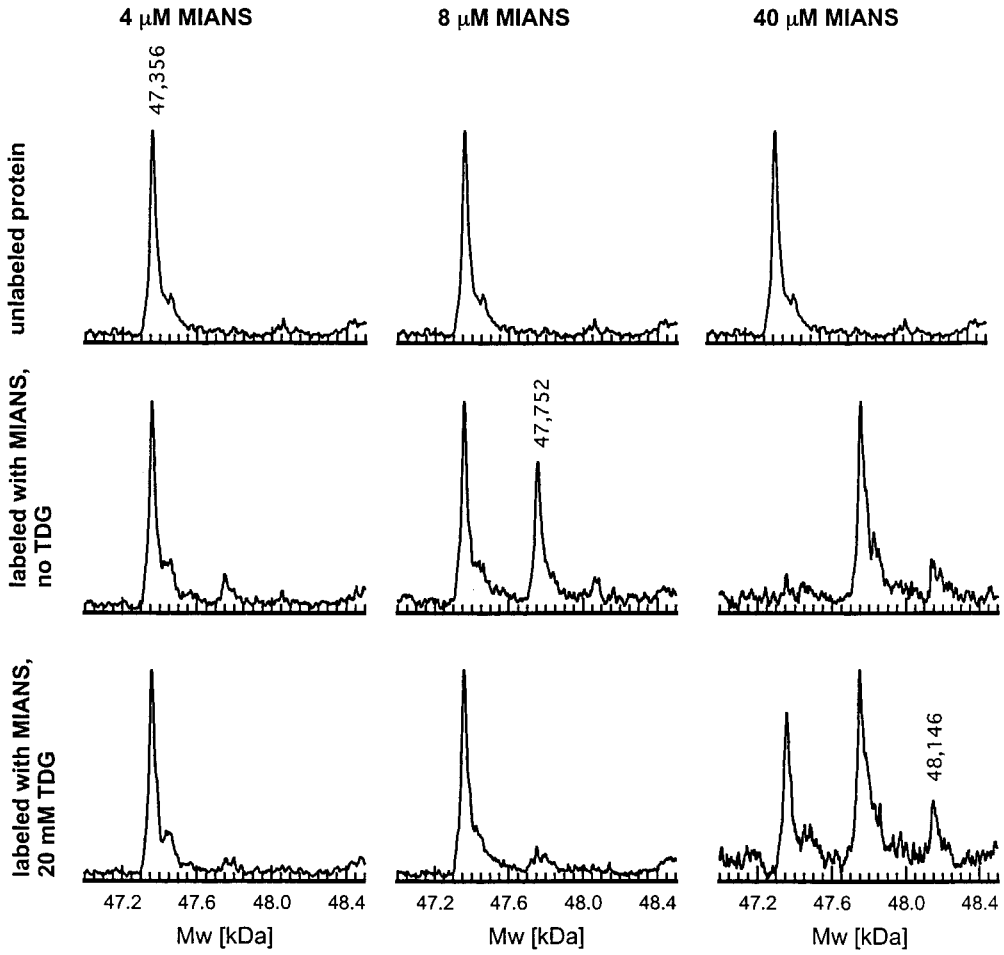

FIGURE 1: Mass peaks of wild-type lactose permease and labeling with MIANS. For clarity, the mass peaks of the unlabeled species are shown in the upper row (M_w = 47 356 Da). Increments in the molecular mass of 393 Da indicate covalent attachment of a MIANS label. MIANS labeling with increasing concentrations is shown in the second row, demonstrating partial labeling with a single fluorophor (M_w = 47 752 Da) at MIANS concentrations of 4 and 8 μM and a majority of the population remaining unlabeled. At a MIANS concentration of 40 μM, the unlabeled species is completely depleted in favor of a dominating population with a single label and a small fraction of molecules with two MIANS labels (M_w = 48 146 Da). Preincubation of the sample with 20 mM TDG, as shown in the bottom row, leads to complete protection against MIANS labeling with final concentrations of 4 and 8 μM. However, only partial protection is observed against single and double alkylation at a MIANS concentration of 40 μM.

MIANS results in a complete loss of the unlabeled population, and the majority of the molecules exhibits a molecular mass of 47 752 Da, indicating reaction with a single MIANS label. However, at this MIANS concentration, a small fraction is doubly labeled as indicated by the peak at 48 146 Da, and preincubation with TDG only partially protects against labeling. The pattern observed is identical to that found for wild-type permease labeled with NEM, where Cys148 and Cys355 are alkylated (*2*, *11*).

A time course of fluorescence labeling was recorded at a MIANS concentration of 4 μM (Figure 2). As shown, addition of TDG completely inhibits MIANS labeling with the protein either in DDM or reconstituted into proteoliposomes (not shown). Therefore, despite the presence of eight native Cys residues in wild-type permease, as indicated by both ESI-MS and fluorescence, only Cys148 is labeled under these conditions.

Spin-Labeling KcsA. KcsA is a small prokaryotic K^+ channel of known structure from the cytoplasmic membrane of *S. lividans* (*16*, *17*). The channel is assembled from four identical subunits in a 4-fold symmetrical fashion. Each subunit consists of 160 amino acids and forms two membrane-spanning α-helices. In the assembled channel, eight transmembrane helices separate a central water-filled pathway for K^+ ions from the interior of the lipid bilayer. A variety of biophysical investigations such as EPR, FTIR, and CD spectroscopy have been employed (*5*, *18*, *19*) to test flexibility and solvent accessibility of the protein. For EPR spectroscopy, it is important to ensure that the spin-label that reacts with a specific Cys residue in the protein in an efficient manner. Moreover, a recent site-directed spin-labeling study (*5*) shows that labeling efficiency in KcsA is a direct function of the structure of the protein. A typical ESI-MS spectrum of purified wild-type KcsA shows molecular ions carrying 10−25 positive charges (Figure 3). The measured molecular mass is given in the upper inset as 19 275 Da, which compared to the calculated molecular mass of 19 274.39 Da for the non-*N*-formylated protein demonstrates an accuracy of better than 0.01%. Site-specific replacement of Phe114 with Cys alters the measured molecular mass (19 231 Da), and spin-labeling of the mutant results in a homogeneous population of molecules with a molecular mass of 19 417 Da. ESI-MS not only monitors alkylation efficiency and sample homogeneity for individual mutants as shown here

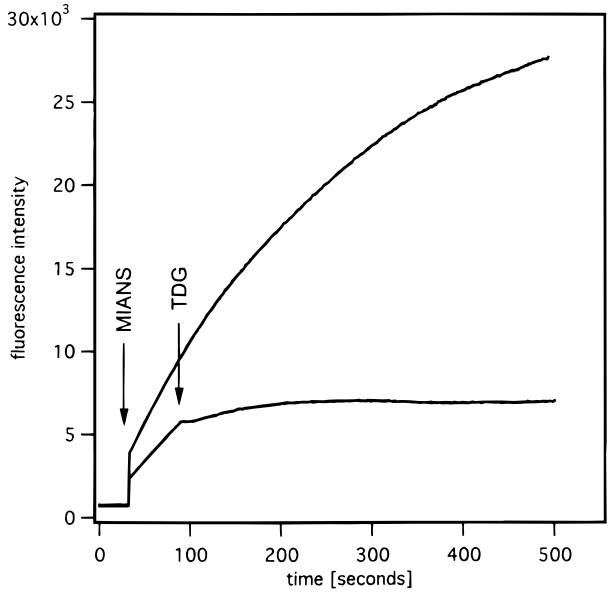

FIGURE 2: Time course of MIANS labeling of wild-type lactose permease. Addition of MIANS at a final concentration of 4 μM to 200 μg/mL purified wild-type lac permease in 50 mM KP_i (pH 7.5) and 0.016% DDM induces a steady increase in fluorescence (upper trace) due to continuous labeling of Cys148. Addition of TDG at a final concentration of 20 mM, as indicated, fully blocks further labeling (bottom trace).

but also should be able to scan entire transmembrane segments for accessibility to different reagents for the purpose of obtaining information about secondary structure.

Primary Structure of vSGLT from V. parahaemolyticus. Sodium cotransporters belong to the superfamily of membrane proteins that catalyze solute accumulation against a concentration gradient coupled to the downhill transport of Na^+ (*20*). A Na^+/glucose symporter (vSGLT) has been cloned from *V. parahaemolyticus* that is closely related to the human Na^+/glucose cotransporter SGLT1. The full-length protein was reported to contain 530 amino acids starting with the motif MVFAIYVAII at the N-terminus and a calculated molecular mass of 57 417.09 Da (*21*). The protein from the original clone purified with an engineered six-His tag connected via the 10-amino acid linker peptide VLYKSGGSPG to the C-terminus was subjected to ESI-MS (Figure 4). The observed molecular mass of 60 680 Da is not in agreement with the published amino acid sequence, since the molecular mass of the published sequence plus the C-terminal linker and the six-His tag is calculated to be 59 186 Da. Thus, it appears the predicted N-terminus is incorrect due to a DNA sequencing error (E. Turk, manuscript in preparation), and the correct N-terminus of the recombinant protein is formylated MSNIEHGLSFIDI with a calculated molecular mass of 60 671.67 Da.

Identification of Membrane Proteins without Purification. With proteomics as an emerging field and the increasing demand for high-throughput techniques, it is important to obtain high-resolution mass spectrometric information directly from protein mixtures (*22*). Therefore, an extract of *E. coli* membranes solubilized in DDM was subjected to in-line reverse phase HPLC prior to ESI-MS, and membrane proteins were eluted with an increasing gradient of CH_3CN (Figure 5A). The polystyrene/divinylbenzene stationary phase proved to be effective in separation of membrane proteins in the presence of formic acid (*1*). Although complete

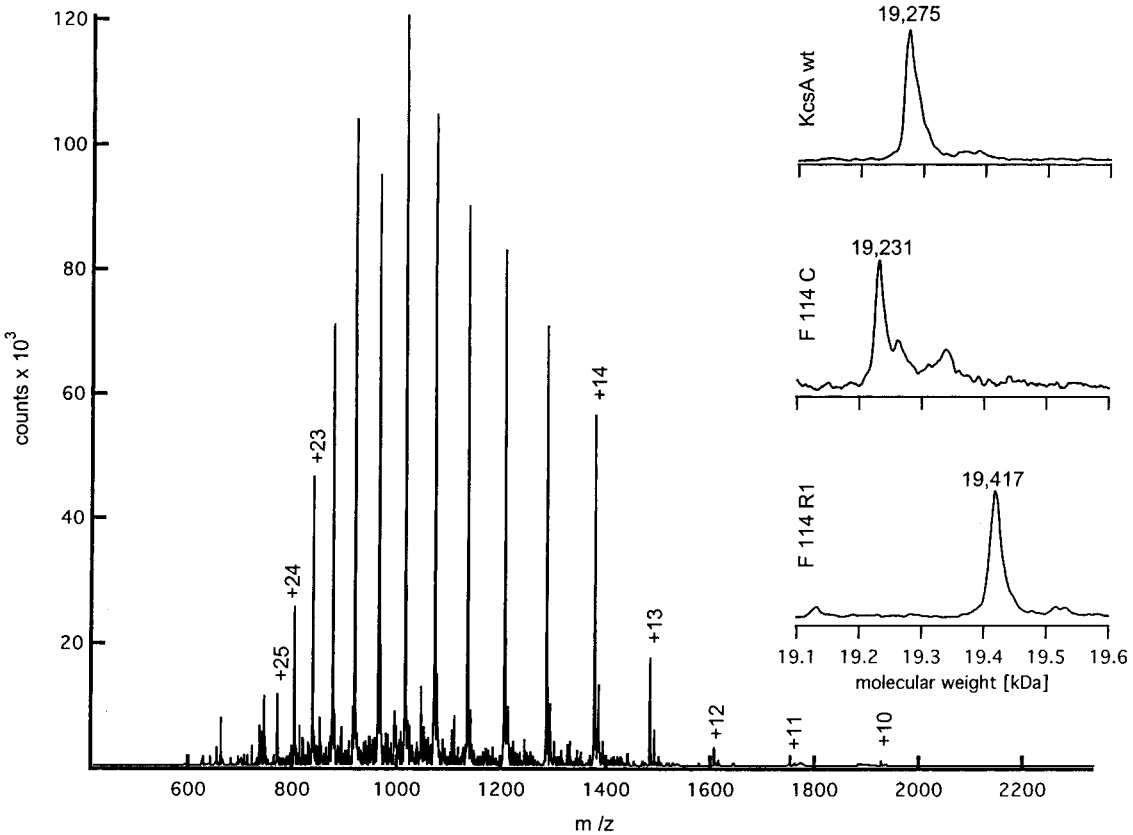

FIGURE 3: Spin-labeling KcsA. A typical ESI-MS spectrum is shown for wild-type KcsA with ions carrying positive charges from 10 to 25. The insets show the reconstructed mass peaks for the wild-type protein, the F114C mutant, and the F114C mutant labeled with a spin-label. A single peak with a measured mass increment of 186 Da indicates uniform and complete labeling of the protein.

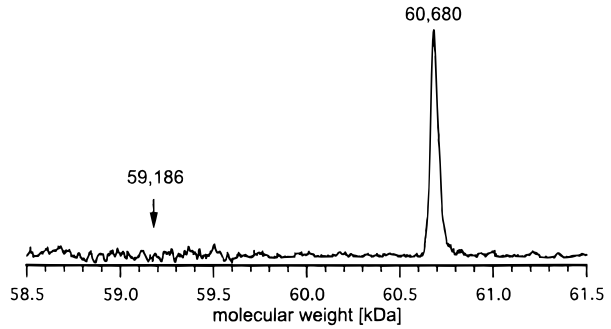

FIGURE 4: Primary structure of vSGLT. Mass peak of purified vSGLT with a molecular mass of 60 680 Da. The arrow indicates the molecular mass of the gene product assuming the previously published deduced amino acid sequence (see the text).

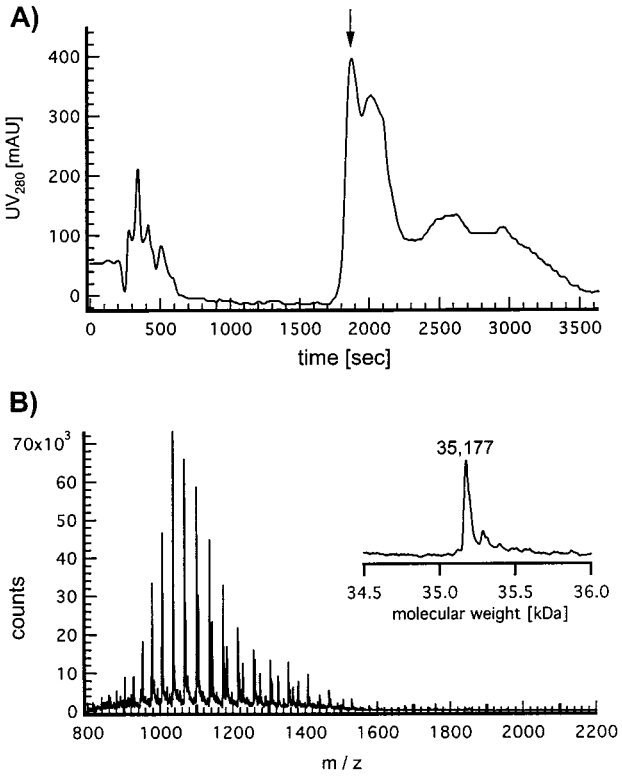

FIGURE 5: Identification of OmpA from a crude membrane extract. (A) Elution profile of solubilized *E. coli* membrane proteins using an increasing gradient of acetonitrile on a reverse phase HPLC column. Between 200 and 500 s, the flow-through of unbound protein after multiple injections is observed. A relatively sharp peak elutes at 1900 s (arrow) prior to elution of the majority of the sample. (B) ESI-MS spectrum of the indicated peak in panel A. The reconstructed mass peak is shown in the inset with a molecular mass of 35 177 Da corresponding to OmpA (see the text for further discussion).

separation of the proteins is not achieved, the elution profile exhibits a relatively sharp and symmetric peak at 1900 s. On-line ESI-MS of this peak reveals a single molecular species with a molecular mass of 35 177 Da (Figure 5B). To identify this protein, a database search with a stringency of 0.1% deviation from the molecular mass was performed in the SWISSPROT database using the ExPASy bioinformatics platform (www.expasy.ch) for *E. coli* membrane proteins. Unambiguously, the protein is identified as the outer membrane protein A precursor, OmpA (P02934), the most abundant protein in the *E. coli* outer membrane. Moreover, to verify the identification, the eluate of the presumed OmpA HPLC peak was cleaved with CNBr, and the fragments were subjected to matrix-assisted laser desorption ionization (MALDI) MS. Although formylated during cleavage, all six expected OmpA fragments are clearly identified (see the Supporting Information).

As demonstrated here, chromatography combined with ESI-MS can separate an individual membrane protein from a crude mixture and measure its molecular mass which can be used subsequently to identify the protein with a database search provided the sequence is correct and the protein has not been post-translationally modified. With the use of tandem mass spectrometry (MS/MS) (*22*, *23*), it should be possible in the future to expand the approach and to fragment and identify the separated membrane protein in the mass spectrometer.

ACKNOWLEDGMENT

We thank Jenny Lee, Lawrence Lee, Susannah Marshall, and Olivia Kim for technical assistance.

SUPPORTING INFORMATION AVAILABLE

Identification of the 35 177 Da membrane protein obtained through HPLC from a crude mixture using cyanogen bromide cleavage and MALDI-TOF MS on the fragments derived thereof. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Whitelegge, J. P., Gundersen, C. B., and Faull, K. F. (1998) *Protein Sci. 7*, 1423−1430.
2. Whitelegge, J. P., le Coutre, J., Lee, J. C., Engel, C. K., Privé, G. G., Faull, K. F., and Kaback, H. R. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 10695−10698.
3. Schindler, P. A., Van Dorsselaer, A., and Falick, A. M. (1993) *Anal. Biochem. 213*, 256−263.
4. Ball, L. E., Oatis, J. E., Jr., Dharmasiri, K., Busman, M., Wang, J., Cowden, L. B., Galijatovic, A., Chen, N., Crouch, R. K., and Knapp, D. R. (1998) *Protein Sci. 7*, 758−764.
5. Gross, A., Columbus, L., Hideg, K., Altenbach, C., and Hubbell, W. L. (1999) *Biochemistry 38*, 10324−10335.
6. Wessel, D., and Flügge, U. I. (1984) *Anal. Biochem. 138*, 141−143.
7. Witkop, B. (1961) *Adv. Protein Chem. 16*, 221−321.
8. Wu, J., and Kaback, H. R. (1994) *Biochemistry 33*, 12166−12171.
9. Kaback, H. R., and Wu, J. (1999) *Acc. Chem. Res. 32*, 805−813.
10. le Coutre, J., Narasimhan, L. R., Patel, C. K., and Kaback, H. R. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 10167−10171.
11. Frillingos, S., Sahin-Tóth, M., Wu, J., and Kaback, H. R. (1998) *FASEB J. 12*, 1281−1299.
12. Venkatesan, P., and Kaback, H. R. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 9802−9807.
13. Sahin-Tóth, M., le Coutre, J., Kharabi, D., le Maire, G., Lee, J. C., and Kaback, H. R. (1999) *Biochemistry 38*, 813−819.
14. Frillingos, S., and Kaback, H. R. (1996) *Biochemistry 35*, 3950−3956.
15. Sahin-Tóth, M., Akhoon, K. M., Runner, J., and Kaback, H. R. (2000) *Biochemistry* (in press).
16. Schrempf, H., Schmidt, O., Kümmerlen, R., Hinnah, S., Müller, D., Betzler, M., Steinkamp, T., and Wagner, R. (1995) *EMBO J. 14*, 5170−5178.
17. Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., and MacKinnon, R. (1998) *Science 280*, 69−77.

18. le Coutre, J., Kaback, H. R., Patel, C. K. N., Heginbotham, L., and Miller, C. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 6114−6117.
19. Cortes, D. M., and Perozo, E. (1997) *Biochemistry 36*, 10343−10352.
20. Turk, E., and Wright, E. M. (1997) *J. Membr. Biol. 159*, 1−20.
21. Sarker, R. I., Okabe, Y., Tsuda, M., and Tsuchiya, T. (1996) *Biochim. Biophys. Acta 1281*, 1−4.
22. Jensen, P. K., Pasa-Tolic, L., Anderson, G. A., Horner, J. A., Lipton, M. S., Bruce, J. E., and Smith, R. D. (1999) *Anal. Chem. 71*, 2076−2084.
23. McLafferty, F. W., Kelleher, N. L., Begley, T. P., Fridriksson, E. K., Zubarev, R. A., and Horn, D. M. (1998) *Curr. Opin. Chem. Biol. 2*, 571−578.

BI000150M