Toward the Complete Membrane Proteome

HIGH COVERAGE OF INTEGRAL MEMBRANE PROTEINS THROUGH TRANSMEMBRANE PEPTIDE DETECTION*⑤

Frank Fischer‡§, Dirk Wolters§¶, Matthias Rögner‡, and Ansgar Poetsch‡∥

To attain a comprehensive membrane proteome of two strains of *Corynebacterium glutamicum* (L-lysine-producing and the characterized model strains), both sample pretreatment and analysis methods were optimized. Isolated bacterial membranes were digested with trypsin/cyanogen bromide or trypsin/chymotrypsin, and a complementary protein set was identified using the multidimensional protein identification technology (MudPIT). Besides a distinct number of cytosolic or membrane-associated proteins, the combined data analysis from both digests yielded 326 integral membrane proteins (~50% of all predicted) covering membrane proteins both with small and large numbers of transmembrane helices. Also membrane proteins with a high GRAVY score (Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132) were identified, and basic and acidic membrane proteins were evenly represented. A significant increase in hydrophobic peptides with distinctly higher sequence coverage of transmembrane regions was achieved by trypsin/chymotrypsin digestion in an organic solvent. The percentage of identified membrane proteins increased with protein size, yielding 80% of all membrane proteins above 60 kDa. Most prominently, almost all constituents of the respiratory chain and a high number of ATP-binding cassette transport systems were identified. This newly developed protocol is suitable for the quantitative comparison of membrane proteomes and will be especially useful for applications such as monitoring protein expression under different growth and fermentation conditions in bacteria such as *C. glutamicum*. Moreover with more than 50% coverage of all predicted membrane proteins (including the non-expressed species) this improved method has the potential for a close-to-complete coverage of membrane proteomes in general. *Molecular & Cellular Proteomics* 5:444–453, 2006.

About 20–30% of all genes in an organism code for integral membrane proteins. Integral membrane proteins are involved in central cellular processes and form the major protein class for drug targets. According to their structure they can be classified either as α-helical or as β-barrel proteins. Although the latter can be studied easily by proteomic techniques, α-helical membrane proteins still remain an analytical challenge. Due to their location in the lipid bilayer, membrane proteins are amphipathic, and the targets for tryptic cleavage, lysine and arginine, are mainly absent in transmembrane (TM)[1] helices and only found in the hydrophilic part of the protein. The size of exposed hydrophilic domains varies among integral membrane proteins from large (*e.g.* epidermal growth factor receptor) to small (*e.g.* rhodopsin). Because most integral membrane proteins are of low abundance and show a high GRAVY score (1), all proteomic protocols usually involve prefractionation steps for enrichment of the membrane fraction. Examples include organelle separation by free-flow electrophoresis (2) and removal of membrane-associated proteins by chaotropes (3) or alkaline pH washes (4). In the next step, membranes are either treated directly with protease or solubilized with detergents, and the intact proteins are further separated before digestion. However, even after enrichment, a high percentage of the identified proteins are usually not membrane proteins regardless of the applied proteomic approach. In addition, only a small fraction (usually less than 30%) of all predicted membrane proteins are detected. In conclusion, despite their high importance and consequentially a considerable effort in method development (5), current approaches aiming at obtaining a comprehensive coverage of the membrane proteome are far from satisfactory. The 2DE technique is unsuitable for the separation of integral membrane proteins mainly due to protein aggregation during the IEF step. For this reason only membrane proteins with a low GRAVY score and only one to two TM helices are detected. In comparison, the combination of SDS-PAGE and LC-ESI MS/MS has been applied with more success, but problems like protein insolubility and the loss of hydrophobic peptides, which prevent protein identification, still remain (6). These can be overcome by a variety of protein digestion strategies using isolated intact membranes that are followed by peptide separation either in one or two dimensions (*e.g.* multidimensional protein identification technology (MudPIT)). Although cleavage with CNBr/trypsin (7) in formic acid or proteinase K in aqueous buffer (8) almost exclusively yields peptides of exposed hydrophilic domains from membrane

---

[1] The abbreviations used are: TM, transmembrane; IMP, integral membrane protein; SIMPLE, specific integral membrane peptide level enrichment; MudPIT, multidimensional protein identification technology; ABC, ATP-binding cassette; 2D, two-dimensional; 2DE, two-dimensional gel electrophoresis; HFBA, heptafluorobutyric acid; nLC, nanoflow LC.

---

proteins (e.g. loops), digestion in 60% methanol also permits the analysis of peptides from transmembrane helices (9).

We developed a modified integral membrane protein enrichment and cleavage procedure and used it for the analysis with MudPIT (7). The membranes were first incubated with trypsin in aqueous buffer to remove hydrophilic domains and membrane-associated proteins followed by chymotrypsin/trypsin digestion in 60% methanol to release peptides originating from transmembrane helices. Alternatively intact membranes were treated with CNBr/trypsin to identify membrane proteins by exposed domains (7).

For benchmarking of the newly developed protocol, the Gram-positive actinomycete *Corynebacterium glutamicum* discovered by Kinoshita *et al.* (10) was used. It is exploited for the biotechnological production of amino acids and nucleotides (11) with the most important products being the amino acids L-glutamate (1,000,000 tons/year) and L-lysine (560,000 tons/year), which are mainly used as a flavor enhancer and food supplement, respectively. Both in corynebacteria (12) and the related mycobacteria, the cell wall has been shown to be an efficient permeability barrier. It prevents the penetration of antibiotics into the bacterium, and is therefore one of the reasons for the virulence of pathogenic mycobacteria. Because *C. glutamicum* does not contain L-lysine- and L-arginine-degrading enzymes, amino acid export processes across the cell envelope play an important role in reducing their high intracellular concentrations during amino acid production (13). Such transport proteins are among the roughly 660 annotated integral membrane proteins and represent about 22% of all 3099 proteins forming the proteome of *C. glutamicum* (14, 15). In contrast, previous analysis of the membrane proteome of *C. glutamicum* by 2D electrophoresis failed to identify membrane integral proteins, and only cytosolic and water-soluble membrane-associated proteins were detected (16, 17). Because IEF is unsuitable for the separation of hydrophobic integral membrane proteins (6, 18), we have recently developed a method that replaces the IEF step of 2D electrophoresis by ion exchange chromatography. Preliminary results with this 2D protein separation approach enabled the separation and identification of more than 170 proteins, including 50 integral membrane proteins (19). Although this separation of intact polypeptides is highly suitable for the screening of protein modifications and for the quantitative comparison of proteomes, the coverage of membrane proteins is still unsatisfactory.

In this communication we show that a total of 326 integral membrane proteins (*i.e.* more than 50% of all predicted) could be identified by applying a complementary cleavage strategy for an L-lysine-producing strain of *C. glutamicum*. The analysis was comprehensive with respect to membrane protein pI and GRAVY and enabled the detection of hydrophilic as well as highly hydrophobic peptides. This new method is easy to perform and universally applicable for all membranes. Furthermore the introduction of a new tryptic predigest specifi-

cally enriches TM helices, thereby significantly improving the identification of highly hydrophobic integral membrane proteins by MS analysis.

EXPERIMENTAL PROCEDURES

Materials—Materials used and their suppliers were as follows: formic acid, sodium chloride, sodium hydrogen phosphate, magnesium chloride, and ammonia (all from Mallinckrodt Baker); ammonium carbonate and ammonium bicarbonate (Fluka); potassium chloride and manganese chloride (Riedel-de Haen); potassium hydrogen phosphate, methanol, and cyanogen bromide (Merck); protease inhibitor for bacterial cells (Sigma); sequencing grade trypsin (Promega, Madison, WI); sequencing grade chymotrypsin and DNase I (Roche Diagnostics); and heptafluorobutyric acid (Pierce). SPEC Plus PT C_{18} solid-phase extraction pipette tips were obtained from Ansys Diagnostics (Lake Forest, CA).

Bacterial Strains, Growth, and Cell Lysis—*C. glutamicum* cells from the L-lysine-producing strain DM 1698 (derived from ATCC 21527) were obtained from a large scale fermentation supplied by the Degussa AG (Halle, Germany). The bacteria were sedimented by centrifugation at 6000 × *g* for 15 min and washed with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). The washed cells were resuspended in disintegration buffer (PBS containing additional 20 mM $MgCl_2$, 10 mM $MnCl_2$, 200 units/ml DNase I, protease inhibitor mixture for bacterial cells (Sigma)) at a concentration of 4 ml of buffer/g of wet cells. The cells were disrupted by French pressure cell treatment (40,000 psi pressure cell with a volume of 35 ml, Thermo Spectronic, Rochester, NY) with the unbroken cells being subsequently sedimented twice by centrifugation at 6000 × *g* at 4 °C. The membranes were pelleted by ultracentrifugation at 100,000 × *g* at 4 °C for 30 min.

MudPIT—A Dual Gradient System HPLC pump (Dionex, Amsterdam, The Netherlands) including a Famos autosampler and Switchos valve unit was connected to a Finnigan LTQ ion trap mass spectrometer (Thermo Electron Corp., San Jose, CA). The LTQ was operated via instrument method files of Xcalibur to acquire a full MS scan between 350 and 2000 *m/z* followed by full MS/MS scans of the three most intensive ions from the preceding MS scan. The heated desolvation capillary was set to 180 °C. The relative collision energy for collision-induced dissociation was set to 35%, and dynamic exclusion was enabled with a repeat count of 1 and a 3-min exclusion duration window.

The cleaned samples were loaded onto a triphasic microcapillary column. The fritless 100-µm capillary column was first packed with 12 cm of Eclipse XDB C_{18} (Hewlett Packard, Palo Alto, CA), then with 4 cm of a strong cation exchange material (Whatman, Clifton, NJ), and finally with 3 cm of the reverse-phase material (Hewlett Packard). The column flow rate was set to 0.15–0.25 µl/min, and a spray voltage of 1.8 kV was used.

The three buffer solutions used for the chromatography were 5% ACN, 0.012% HFBA, 0.5% acetic acid (buffer A); 80% ACN, 0.012% HFBA, 0.5% acetic acid (buffer B); and 1 M ammonium acetate, 5% ACN, 0.012% HFBA (buffer C).

2, 4, 7, 10, 12, 15, 20, 25, 30, 40, 50, 75, and 100% buffer C were used to displace fractions from the cation exchanger onto the reverse-phase material. Finally a 20-min salt wash with 100% buffer C was performed. Up to a salt concentration of 12% buffer C, peptides were eluted with a linear gradient from 0 to 100% buffer B in 180 min; higher salt concentration steps were followed by a linear gradient of buffer B in 160 min.

nLC-ESI—nLC-ESI analysis was performed on a Finnigan LCQ ion trap mass spectrometer (Thermo Electron Corp.). The instrumental setup for the liquid chromatography was identical to that in the MudPIT experiment with the exception that the fritless 75-µm column

was packed only with 10 cm of the reverse-phase material Eclipse XDB C_{18}. The desalted peptide mixture was loaded onto the reverse-phase column under high pressure, and peptides were eluted by a linear acetonitrile gradient from 0 to 100% buffer B within 220 min directly into the mass spectrometer. The instrument method files of Xcalibur are identical to the LTQ settings, and the MS/MS data interpretation was performed by the SEQUEST algorithm.

SEQUEST Analysis—The SEQUEST algorithm was used for MS/MS data interpretation. To obtain reliable protein identification, only peptides with a ΔCn score above 0.1 were used. In addition, a peptide had to be tryptic, chymotryptic, or a fragment of a CNBr cleavage to be accepted, and the cross-correlation scores of single, double, and triple charged peptides had to be greater than 1.8, 2.5, and 3.5, respectively. As modifications the oxidation of methionine and the formation of homoserine lactone after CNBr cleavage were permitted.

Sample Preparation for MudPIT—Following a protocol of Washburn et al. (7) for 2D LC-ESI analysis, isolated membranes from *C. glutamicum* (protein content, 600 μg) were first washed with ammonium carbonate (0.1 M, pH 11) and then dissolved in 1 ml of 90% formic acid followed by digestion with CNBr for 24 h at room temperature.

Thereafter the sample volume was reduced to about 200 μl, and the solution was adjusted to pH 8.6 using solid ammonium carbonate and 2 M urea. Subsequently the sample was digested overnight with trypsin (1:100, w/w). In the following sections this method is named MudPIT_class.

For another type of sample preparation the isolated membranes were digested with trypsin overnight in a 25 mM ammonium bicarbonate buffer, pH 8.6, at 37 °C. The next day the membranes were pelleted by centrifugation at 100,000 × *g*. The membranes were washed twice with ice-cold deionized water. After the washing steps the membrane pellet was resuspended in 60% methanol and 40% 25 mM ammonium bicarbonate buffer by sonication for 2 × 10 min. The pH was adjusted to 8.6 with ammonium hydroxide. Subsequently trypsin and chymotrypsin (each 1:100, w/w) were added to the sample. The proteolysis was performed overnight at 37 °C. In the following sections this method is named specific integral membrane peptide level enrichment (SIMPLE).

To remove undigested proteins and membrane fragments 3-kDa cut-off filters (Microcon YM-3, Millipore) were used. Both samples were adjusted to pH 2.0 with acetic acid and used for a MudPIT experiment.

Sample Preparation for nLC-ESI—Isolated membranes of 600 μg of protein were digested overnight with trypsin (1:100, w/w) in 25 mM ammonium bicarbonate buffer at 37 °C. Membranes were then sedimented by ultracentrifugation at 100,000 × *g* for 30 min. After washing with ice-cold 0.1 M ammonium carbonate buffer, pH 11.0, and deionized water the membrane pellet was resuspended in three different solvents: (a) 60% methanol (in 25 mM ammonium bicarbonate, pH 8.6), (b) 2 M urea (in 25 mM ammonium bicarbonate, pH 8.6), and (c) 25 mM ammonium bicarbonate buffer (pH 8.6). Subsequently trypsin and chymotrypsin (3 μg each) were added to the samples, and proteolysis was carried out overnight at 37 °C. After cleavage, the membranes were sedimented by centrifugation at 100,000 × *g* for 20 min. The peptide-containing supernatant was desalted by solid-phase extraction pipette tips and used for LC-ESI analysis.

RESULTS

Sample Prefractionation—The aim of our work was to establish a new method for the identification of membrane proteins in combination with MudPIT. This should finally help to characterize the composition of the membrane proteome of *C. glutamicum*.

Due to the low abundance of membrane proteins in comparison with other proteins, prefractionation is indispensable and vital. Previous studies have reported the effect of a carbonate washing procedure of isolated membranes at alkaline pH, which results in the formation of membrane sheets and the removal of soluble and peripheral membrane proteins (4, 20). This approach is also compatible with a LC-MS/MS analysis (8). Following a slightly modified protocol of Wu et al. (8) (see "Experimental Procedures"), we observed, however, that the carbonate-washed, isolated membrane fraction of *C. glutamicum* contained mainly soluble or peripheral proteins (69%) and only few integral membrane proteins (IMPs) (21%). To achieve a better enrichment of IMPs, we evaluated three washing procedures: (a) carbonate washing in combination with sonication for the release of soluble proteins that could be captured in vesicles, (b) NaBr washing (high salt) to remove soluble proteins by reducing the electrostatic interactions, and (c) incubation with the detergent octyl β-glucopyranoside to inhibit vesicle formation (21) during the washing procedure. In all three cases the washing efficiency was verified by SDS-PAGE and LC-MS/MS measurements (data not shown), and all results were unsatisfactory. Even the best result (with octyl β-glucopyranoside), although increasing the percentage of secreted proteins (defined as proteins with signal peptides) from 10 to 26%, did not increase the percentage of IMPs (data not shown). In summary these results suggest that a vesicle formation was not the reason for the inefficient removal of the contaminating proteins.

Because all tested washing procedures did not show the expected enrichment of the IMPs, we concluded that the unique cell wall composition of the actinomycetes prevented a satisfactory removal of membrane-associated proteins. We therefore modified our approach by predigesting all soluble proteins with trypsin, assuming that the resulting peptides have weaker interactions with the cell wall than the undigested proteins (Fig. 1, *step 1*). Alternatively we also evaluated a preliminary carbonate washing procedure of the membranes (Fig. 1, *step 2*) that is outlined below. Although this predigestion with trypsin also attacks the soluble parts of integral membrane proteins, their integral parts are protected by the lipid bilayer from proteasomal action. Subsequently these integral parts should be cleaved by a protease mixture (trypsin/chymotrypsin), and the membrane proteins may be identified by peptides originating from TM helices. After predigestion with trypsin, 136 (73%) soluble proteins and 23 (12%) integral membrane proteins were identified in the supernatant by LC-MS/MS. To identify the integral membrane peptides of the membrane fraction, digests with trypsin/chymotrypsin were tested in three different solvents (Fig. 1, *steps 1a–1c*): (a) 25 mM ammonium bicarbonate, (b) 2 M urea (in 25 mM ammonium bicarbonate), and (c) 60% methanol (in 25 mM ammonium bicarbonate).

Fig. 2a shows the relative amount of all proteins that have been classified according to the various digestion strategies
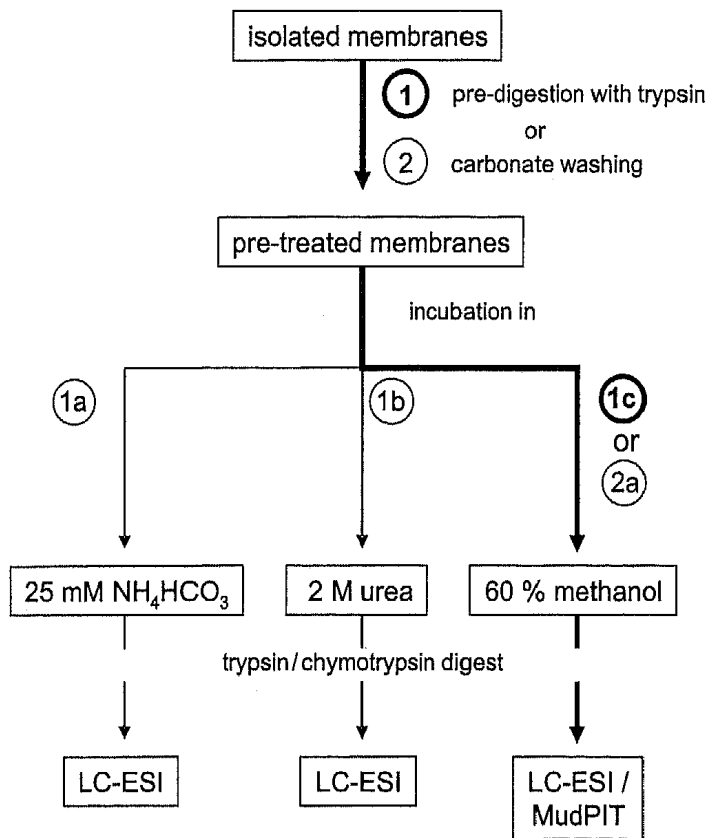
FIG. 1. **Overview on the methods evaluated for a better identification of IMPs.** Isolated membranes of *C. glutamicum* were either predigested with trypsin (*step 1*) or washed with ammonium carbonate (pH 11) to remove soluble proteins (*step 2*). Subsequently the treated membranes were sedimented by ultracentrifugation, and the resulting pellets were dissolved in 25 mM ammonium bicarbonate (*a*), 2 M urea (in 25 mM ammonium bicarbonate) (*b*), or 60% methanol (in 25 mM ammonium bicarbonate) (*c*). Each of these treatments was followed by a trypsin/chymotrypsin digest, and the generated peptide mixture was then analyzed by LC-ESI-MS/MS. The optimal procedure that was used for the MudPIT analysis is highlighted with *thick arrows*.

FIG. 2. **Isolated membranes of *C. glutamicum* pretreated with trypsin (according to Fig. 1, *step 1*) were resuspended in three different solvents: 1) 25 mM ammonium bicarbonate, 2) 2 M urea (in 25 mM ammonium bicarbonate), and 3) 60% methanol (in 25 mM ammonium bicarbonate).** After digestion with a trypsin/chymotrypsin mixture, each sample was analyzed by LC-MS/MS followed by protein identification via the SEQUEST algorithm. *a*, identified proteins were classified into secreted proteins, soluble proteins, and integral membrane proteins for all three solvents. *b*, number of proteins identified on the basis of peptides with predicted TM regions. In the best case, 29 proteins were identified with at least one transmembrane peptide in organic solvent (60% MeOH). In the aqueous solvents, only 15 (2 M urea) or 12 (NH_4HCO_3) proteins were identified with a TM peptide. The *axis* on the *right side* shows the average number of amino acids in identified TM peptides, *i.e.* peptides located in the lipid bilayer. *øAA*, average number of amino acids.

that were subsequently analyzed by LC-MS/MS (according to Fig. 1). After treatment with 25 mM ammonium bicarbonate, only 30% (49 proteins) of all identified proteins were IMPs, similar to the treatment of 2 M urea after which only 25% (47 proteins) of all identified proteins were IMPs. The similar number of identified membrane proteins in both assays indicates that there is no specific solubilization effect of urea, an observation that is also supported by the fact that the percentages of identified soluble and secreted proteins were almost equal.

In contrast, after digestion in 60% methanol, more than 50% (75 proteins) of all identified proteins were IMPs. Using the predigestion with trypsin for all three samples (see Fig. 1), the organic solvent assay yielded a dramatic increase of identified IMPs and an increased coverage of their TM regions. In parallel, the number of contaminating soluble proteins was reduced dramatically (below 50%). Comparison of these data with carbonate-washed membranes (Fig. 1, *step 2*) showed that predigestion with trypsin also contributed considerably to an enrichment of IMPs. A carbonate-washed
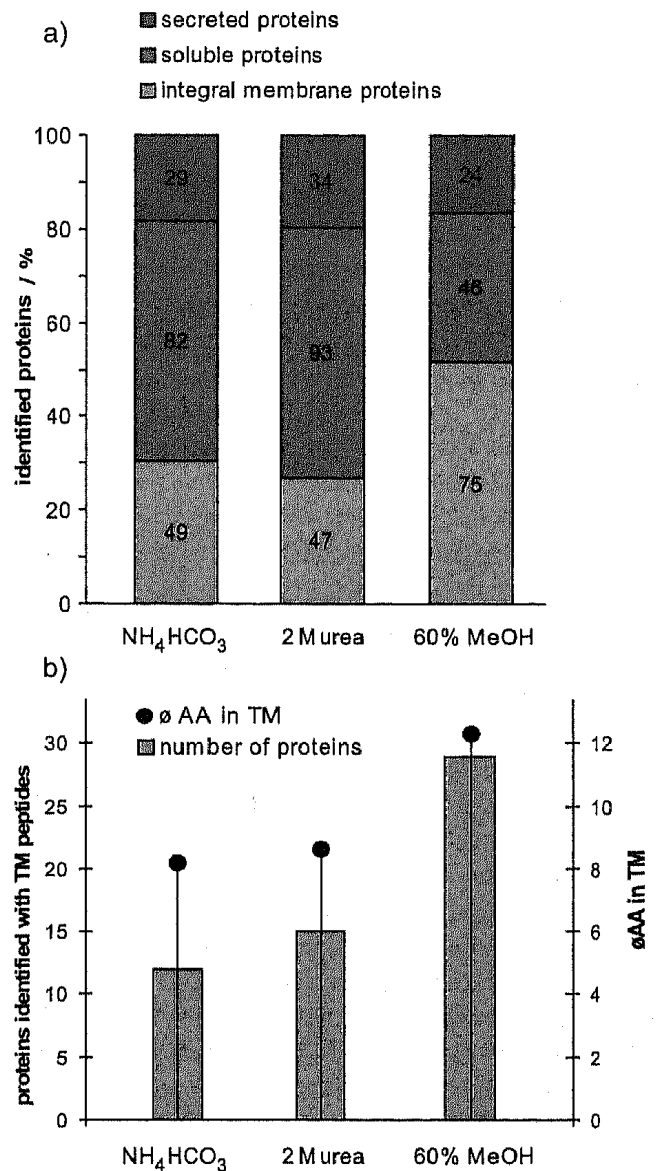
membrane fraction, which was digested with trypsin/chymotrypsin in 60% methanol followed by LC-MS/MS (see Fig. 1, step 2*a*), resulted in only 32% IMPs and 56% soluble proteins. In contrast, over 50% IMPs (Fig. 2*a*) were identified after the newly established predigestion sample preparation method (Fig. 1*c*). In conclusion, the predigestion with trypsin led to an

enrichment of IMPs, and protein digestion in organic solvents leads to a higher coverage of integral membrane proteins, especially of transmembrane helices (22). Fig. 2b shows that digestion in the organic solvent yielded a 2-fold increase in identified proteins with at least one TM peptide (defined as peptides that cover at least partially predicted TM domains). In addition, these peptides covered more integral membrane parts of the proteins. On average, after digestion in organic solvent, 12 amino acids of the identified peptides could be localized in the lipid bilayer, whereas after pretreatment in the two aqueous solvents, transmembrane helices were only covered by about eight amino acids. We observed that methanol did not solubilize membranes; therefore, we postulate that the trypsin and alcohol pretreatment destabilizes the membrane, ameliorating protease access to membrane integral helices.

MudPIT—In this study we compared two different membrane prefractionation methods in combination with two different digestion assays followed by MudPIT analysis. We could identify 274 IMPs when using the newly established membrane preparation method (see Fig. 1c) and 202 IMPs when using the modified protocol of Washburn *et al.* (7); in total we identified 326 different IMPs. Among these proteins, 124 could only be identified in the SIMPLE experiment, and 52 could only be identified in the MudPIT_class experiment (see "Experimental Procedures"). Cyanogen bromide cleavage in formic acid can result in different protein modifications, such as the oxidation of methionine and the formation of homoserine lactone, which must also be considered during peptide identification. According to published results (23) formylation of 17% of all Ser and of 5% of all Thr residues may occur too. If we included this modification in SEQUEST searches, we did indeed identify a few modified peptides, but this fact did not affect the total number of identified proteins.

Bioinformatic analysis could reveal whether there was a selective bias for integral membrane proteins possessing particular physiochemical characteristics, *e.g.* very hydrophobic proteins, proteins with many transmembrane helices, or very basic proteins. Fig. 3a shows a virtual 2D gel (visualized by JVirGel; www.jvirgel.de) of the identified membrane proteins in the two different MudPIT experiments. It is apparent that with these methods proteins covering a wide pI range and even very basic proteins (up to a pI of 12) can be identified. It is also obvious that the mass of a protein is not decisive for its identification, although a higher number of small proteins (<20 kDa) were identified in the SIMPLE experiment (*red spots*). Also by plotting protein size against the GRAVY score it is evident (Fig. 3b) that the SIMPLE approach (*red triangles*) yields significantly more small, hydrophobic proteins than the MudPIT_class approach (*blue dots*). A comparison of protein and peptide GRAVY scores between the SIMPLE and MudPIT_class experiments is shown in Table I. In total, the average GRAVY score of all identified membrane proteins in the SIMPLE experiment was 0.36, indicating that a high number of very hydrophobic proteins were detectable by this approach. In con-
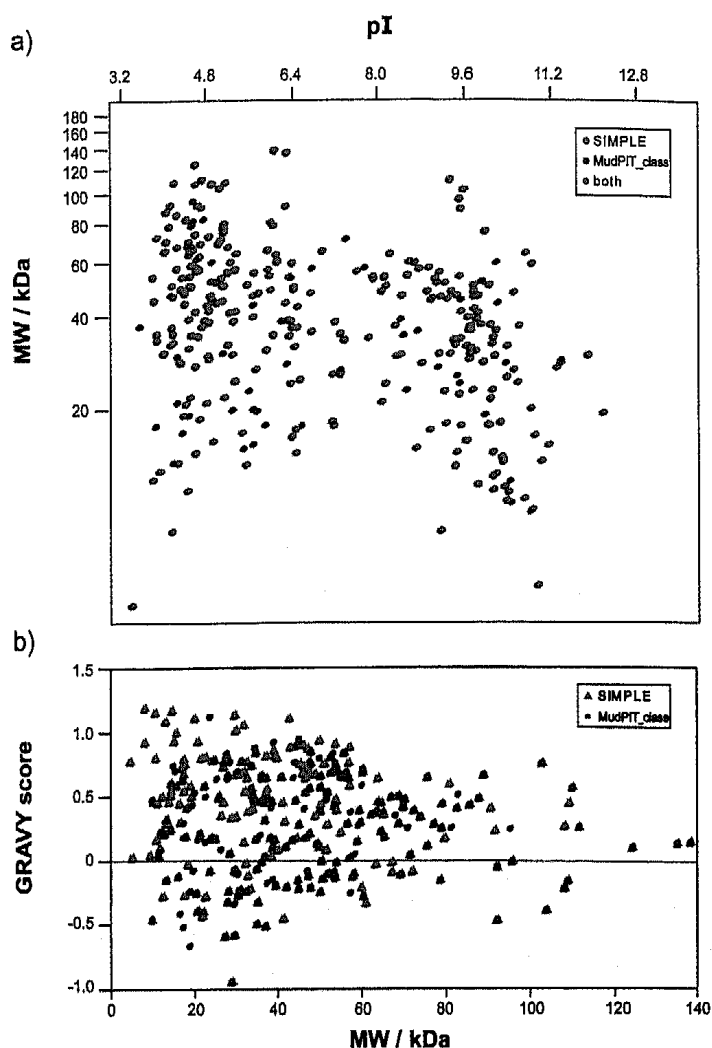

FIG. 3. *a*, visualization of IMPs identified by the MudPIT experiments on a virtual 2D gel using the program JVirGel (www.jvirgel.de). The *red spots* show the identified proteins in the SIMPLE experiment, and the *blue spots* show those IMPs that were identified in the MudPIT_class experiment. Colored in *green* are all IMPs that were found in both MudPIT experiments. *b*, relationship between molecular mass and GRAVY scores of the identified IMPs in the MudPIT experiments. A higher number of small, hydrophobic proteins were detected in the SIMPLE experiment. Recent studies have shown that proteins with GRAVY scores above 0.1 are not detectable by 2DE.

TABLE I

Comparison of the MudPIT_class and the SIMPLE approach; the SIMPLE method enables the detection of more hydrophobic proteins with more TM peptides

	MudPIT_class	SIMPLE
∅[a] GRAVY score of identified IMPs	0.28	0.36
∅ TM peptide GRAVY score of the identified IMPs	0.06	0.67
∅ number of amino acids in the identified TM peptides	6.2	11.0
IMPs only identified with TM peptide	5	35

[a] ∅, average.

trast, the average GRAVY score in the MudPIT_class experiment was only 0.28. The average GRAVY score of all identified TM peptides was about 10-fold higher in the SIMPLE than in
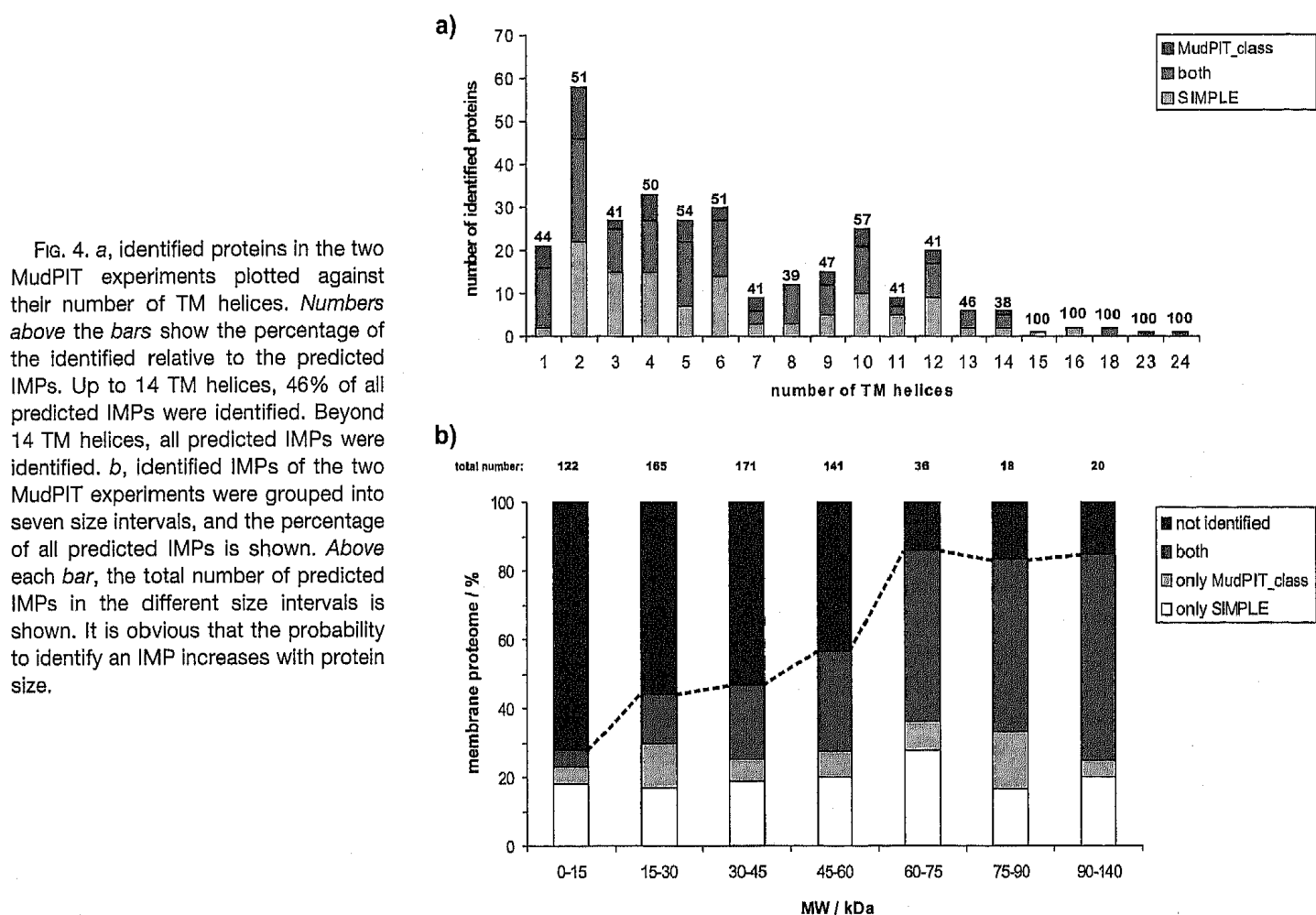
a)

b)

FIG. 4. *a*, identified proteins in the two MudPIT experiments plotted against their number of TM helices. *Numbers above* the *bars* show the percentage of the identified relative to the predicted IMPs. Up to 14 TM helices, 46% of all predicted IMPs were identified. Beyond 14 TM helices, all predicted IMPs were identified. *b*, identified IMPs of the two MudPIT experiments were grouped into seven size intervals, and the percentage of all predicted IMPs is shown. *Above* each *bar*, the total number of predicted IMPs in the different size intervals is shown. It is obvious that the probability to identify an IMP increases with protein size.

the MudPIT_class approach (0.67–0.06); similarly the average number of amino acids in peptides covering TM domains was much higher (11 in the SIMPLE *versus* 6.2 in the MudPIT-_class experiment). Also the number of proteins that were identified by at least one TM peptide increased upon implementation of the newly established sample preparation method (SIMPLE, 35 proteins; MudPIT_class, five proteins). In total, 135 peptides of predicted TM regions corresponding to 82 proteins were identified by the SIMPLE approach, whereas only 32 peptides from 28 proteins were detected by the MudPIT_class experiment. In summary, the SIMPLE experiment helped to identify a higher number of hydrophobic proteins, and the coverage of membrane-embedded parts of IMPs increased.

In Fig.4*a*, the number of identified proteins is plotted against their predicted number of TM helices. On average, 46% of all predicted proteins in the range of 1–14 TM helices (298 proteins) and 100% of all proteins with more than 15 TM helices (seven proteins) were identified. To deduce whether protein size has an impact on identification, all identified membrane proteins were grouped into seven size classes. Fig. 4*b* shows that in the high molecular mass range (60–140 kDa) over 80% of all predicted membrane proteins were identified, but only ~30% in the low molecular mass range (1–15 kDa) were identified. This suggests that MudPIT favors the identification

of high molecular weight membrane proteins possibly due to the fact that the digestion of small IMPs yields only very few peptides (in some cases none), significantly reducing the chance to identify those proteins.

Major Components of the Membrane Proteome—Predicted and identified integral membrane proteins were sorted into functional groups according to the genome annotation from DNA Data Bank of Japan (DDBJ). The biggest group of predicted coding sequences are hypothetical proteins, representing about 50% of all integral membrane proteins, whereas only 33% of all cytosolic proteins belong to this group. Among the predicted membrane proteins, 18% are transporters and substrate binding proteins, 7% are unclassified proteins, 6% are part of the cell envelope, and 5% are proteins involved in energy metabolism with the remainder being spread among the other groups. Fig. 5 shows that the functional classification yields very similar results for both the predicted and the identified membrane proteome with transport and substrate binding proteins (22%) and proteins of the energy metabolism (7%) being slightly overrepresented in the latter. 39% of all identified integral membrane proteins were annotated as hypothetical. Several factors may be responsible for this deviation from the predicted number of 50%. 1) Many hypothetical membrane proteins are small and contain only one transmembrane helix. 2) Although a small size hampers their detection
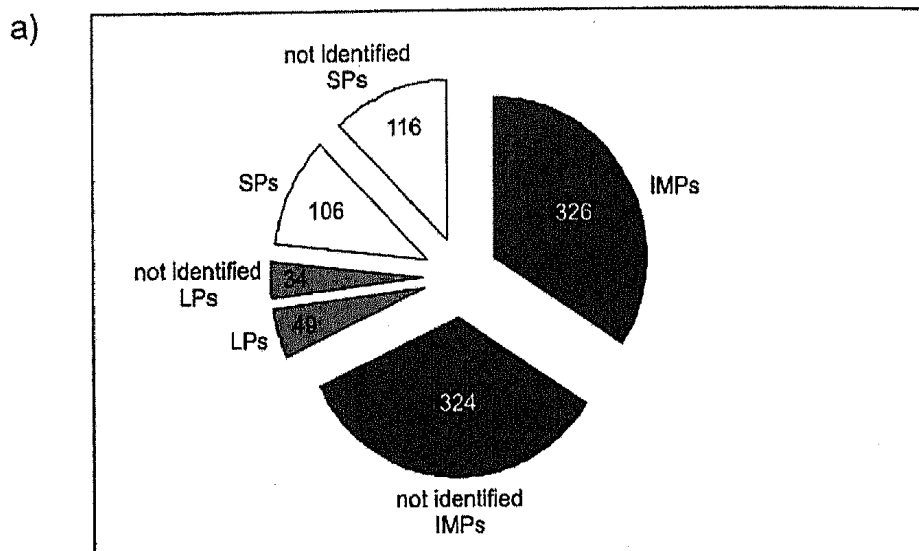
a)

FIG. 5. *a*, overview of all identified IMPs, lipoproteins (*LPs*), and secreted proteins (*SPs*) in the combined MudPIT approach. In all three categories approximately 50% of all predicted proteins were identified. *b*, functional classification (according to DDBJ) of all IMPs identified by the MudPIT experiments in comparison with all predicted IMPs for *C. glutamicum*. The percentage of IMPs listed in the table is calculated relative to the amount of experimentally identified IMPs (*i.e.* 326) and the amount of all predicted IMPs (*i.e.* 650), respectively. In both cases the largest group are hypothetical proteins. For most categories, the percentage between predicted and identified IMPs is similar.

b)

Functional category (DDBJ)	# IMPs - MudPIT/predicted	% IMPs (by MudPIT)	% IMPs (predicted)
Amino acid biosynthesis	3/4	0.8	0.5
Biosynthesis of cofactors, prosthetic groups, and carriers	2/5	0.6	0.6
Cell envelope	18/47	5.0	5.5
Cellular processes	17/26	4.7	3.1
Central intermediary metabolism	10/20	2.8	2.3
DNA metabolism	1/1	0.3	0.1
Energy metabolism	24/39	6.7	4.6
Fatty acid and phospholipid metabolism	5/13	1.4	1.5
Hypothetical proteins	142/417	39.4	48.9
Protein fate	14/31	3.9	3.6
Protein synthesis	1/2	0.3	0.2
Purines, pyrimidines, nucleosides, and nucleotides	2/3	0.6	0.4
Regulatory functions	7/9	1.9	1.1
Signal transduction	4/7	1.1	0.8
Transcription	1/3	0.3	0.4
Transport and binding proteins	82/153	22.8	18
Unclassified	20/56	5.6	6.6
Unknown function	7/16	1.9	1.9

by MS, the presence of only one TM segment leaves a chance for a false annotation by the prediction software. 3) Also the gene identification software may have falsely predicted coding sequences or the proteins may be expressed at a low level, preventing an identification and functional assignment.

To benchmark the capability of our new analysis method and to check for the absence or presence of membrane proteins, we chose ABC transporters as model compounds that contain at least one integral membrane and one soluble subunit. For eight tripartite or tetrapartite ABC transporters, at least three subunits were found, and for 13 tripartite or dipartite transporters, at least two subunits were found. In five cases the permease subunit was not identified, and in two cases the periplasmic component was not identified, whereas the ATPase subunit was detected for all 13 proteins. In conclusion, the membrane-intrinsic permease subunits of ABC transporters appear to be more difficult to detect than their soluble interaction partners. Similarly a previous proteome study of *Halobacterium salinarum* reports an even more obvious underrepresentation of permease subunits (6). Among the ABC proteins with two or three identified subunits are the glutamate permease (Glu), the ABC type phosphate transport

system (Pts), the *sn*-glycerol 3-phosphate transport system (Ugp), and transporters for which the exact substrate is unknown such as metal, sugar, multidrug, and dipeptide/oligopeptide/nickel transporters. For eight multisubunit ABC transporters none of the subunits could be found. Nevertheless the total number of identified ABC transporters was rather high, suggesting that the majority are expressed under the cultivation conditions. Other experiments are required to quantify these various ABC transporters. Additional transporters that could be identified under these conditions include the lysine exporter; the phosphotransferase systems for glucose, fructose, sucrose, and UDP-*N*-acetylmuramyl pentapeptide; the maltose permease; and the low affinity ammonium transporter AmtB.

We also focused on respiratory chain proteins because they are essential for survival and should be rather abundant. Indeed we could identify NADH (Ndh), malate (Mqo), pyruvate (Poxb), D-/L-lactate (Dld/Ldh), glycerol 3-phosphate (Glpd), L-proline (Puta), and succinate (Sdha, Sdhb, and Sdhcd) dehydrogenases as well as nitrate reductase (Narg, Narh, and Nari), cytochrome-*c*-*aaa₃* oxidase complex (Qcrb, Qcra, Qcrc, Ctaf, Ctac, Ctad, and Ctaf; except Ctae), and the complete

ATP synthase complex. Notably no subunit of the cytochrome *bd* menaquinol oxidase was detected, consistent with the absence of its biogenesis protein, *cyd*, and previous reports showing an expression of this enzyme only under microaerobic or copper-deficient conditions (24).

DISCUSSION

Overcoming Problems to Analyze IMPs in Proteomics — The MudPIT technology was first used to characterize the proteome of *Saccharomyces cerevisiae* by Washburn *et al.* (7). In this study nearly 20% of the predicted transmembrane proteins of *S. cerevisiae* were identified. In a second study (Wu *et al.* (8)), membranes from liver and brain cells were digested with the nonspecific proteinase K at high pH to create overlapping peptides of 6–20 amino acids. With this approach, ~28% of the identified proteins contained TM domains. In our investigations, the quality of the MS/MS spectra after proteinase K digest was much inferior to the spectra of peptides after tryptic digestion. In addition, the nonspecific proteinase K digestion leads to a dramatic increase in the number of peptides; they have to be separated in several chromatographic steps, which complicates the study of a complex proteome. For this reason we used a mixture of the specific proteases trypsin and chymotrypsin, which generates a smaller number of peptides. Peptides of previous studies mainly represented either loop regions or soluble subunits of membrane protein complexes. In contrast, TM regions were generally missing, being either inaccessible for the proteases or having no cleavable residues for these proteases. In conclusion, the identification of proteins with small exposed hydrophilic domains is hampered by this approach. The procedure presented in this study, *i.e.* the use of chymotrypsin in 60% methanol, allows cleaving of hydrophobic regions of IMPs, which are mainly TM helices. However, our aim was not focussed on solvent optimization for better digestion of membrane proteins, although other authors had shown that solvents might have an interesting effect on digestion (25).[2] In these latter studies the authors have shown that methanol and acetonitrile allow similar sequence coverage when using trypsin digestion. Blonder *et al.* (9) reported the digestion of IMPs in 60% methanol with trypsin generating large peptides (*e.g.* complete TM helices), but mostly they were hard to detect and often showed poor MS/MS spectra. Using a mixture of trypsin/chymotrypsin in 60% methanol in our studies, many proteins could be identified via chymotryptic peptides. In summary, by our modified procedure we could achieve a comprehensive coverage of the predicted membrane proteome of *C. glutamicum*. This is mainly due to our newly developed membrane digestion protocol, which yielded a significantly higher number of small, hydrophobic proteins. It is apparent from a comparison of SIMPLE and MudPIT_class results that there was an increased coverage of TM regions of the identified IMPs. Another reason for the difficulties in detecting low abundance membrane proteins may have been an insufficient removal of high abundance soluble proteins. They have been substantially removed by our tryptic predigestion, which prevents co-elution and possibly masking of peptides from integral membrane proteins. Another major reason for the improved coverage was the possibility to identify IMPs via their TM regions. This enhances the probability to identify small hydrophobic IMPs because not all peptides produce MS/MS spectra of high quality. More hydrophobic peptides were found in the SIMPLE approach with digestion occurring also in the TM domains. These hydrophobic peptides are easier to detect because they elute later in LC, resulting in much less overlap with contaminating peptides from soluble proteins. Recent studies showed that the peptide signal intensity of a precursor ion in a LC-ESI experiment increases with the hydrophobicity of the peptide as a result of a better ionization efficiency of nonpolar residues (26). They predicted that hydrophobic peptides occur more on the surface of the electrospray droplets where their nonpolar residues can be desolvated, whereas the polar peptides remain in the droplet interior. For this reason we conclude that the critical step is the enrichment and separation of hydrophobic peptides and not their actual detection by LC-ESI. Our results clearly show that under our experimental conditions the extreme hydrophobic nature of TM peptides can be exclusively exploited to improve peptide separation and protein identification.

Toward the Complete Membrane Proteome — Recently Gu *et al.* (27) compared a 2DE approach with a shotgun approach (one-dimensional gel followed by LC-MS/MS) to identify the membrane proteome of the related *Mycobacterium tuberculosis* after washing the cell membranes with a high pH buffer. Although only two integral membrane proteins could be identified after 2DE, 78 were found with the shotgun approach. However, these 78 IMPs represent only 10% of all identified proteins, and this number is still quite small in comparison with the number of IMPs detected in our MudPIT experiments. A possible reason for the comparatively low number of IMPs from the mycobacteria may be their cell wall composition, which prevents efficient removal of membrane-associated proteins at high pH. This is supported by our observation that washing of the membranes from *C. glutamicum* (19) at high pH is not very efficient.

In conclusion, a substantial removal of contaminating proteins is mandatory for obtaining a comprehensive membrane proteome analysis. In contrast to existing methods (*e.g.* high pH, detergent wash), the predigestion with trypsin developed by us for a specific enrichment of IMPs should also be applicable for other organisms, *i.e.* have general significance.

Although the predigest of membranes with trypsin followed by a digestion with trypsin/chymotrypsin in 60% methanol

---

[2] M. Strader, D. L. Tabb, W. J. Hervey IV, C. Pan, and G. B. Hurst (2005) Poster titled "Organic solvents improve trypsin digestion efficiency and specificity for digesting limited protein sample amount" presented at the 53rd American Society for Mass Spectrometry Conference, San Antonio, Texas (June 5–9, 2005).

favors the identification of IMPs with large hydrophobic domains, some IMPs were exclusively detected with the Mud-PIT_class approach, which favors IMPs with large soluble domains. By combining both datasets for the final analysis overall 326 IMPs were identified with 52 resulting specifically from the classic approach, 124 resulting exclusively from the new approach, and 150 being identified from both approaches.

In addition to the IMPs, the combined approach yielded 49 lipoproteins and 106 secreted proteins in the membrane fraction, representing more than 50% of all predicted lipo- and secreted proteins (see Fig. 5). The fact that many membrane protein complexes contain soluble subunits (*e.g.* Sdha and Sdhb) may have contributed to an overestimation of the cytosolic proteins in our membrane fraction. Although the classification of peripheral membrane proteins by bioinformatics is difficult and a reliable number of currently identified peripheral membrane proteins could not be given, they certainly cannot be treated as soluble proteins and should be regarded as either part of the membrane proteome or as a separate class of proteins associated with the membrane. In doing so, the relative amount of IMPs would certainly be higher than in our estimation.

Another fact should also be considered. Statistically we identified about 50% of the predicted membrane proteins from *C. glutamicum* during L-lysine production. There is, however, growing evidence that under no circumstances will all proteins be expressed. For instance, Ghaemmaghami *et al.* (28) demonstrated that only ~80% of all predicted ORFs in *S. cerevisiae* were expressed during log-phase growth. In addition, membrane proteins are much more regulated than cytosolic proteins. It is known, for instance, that under our cultivation conditions only certain import and export systems (*e.g.* AmtB and amino acid exporters such as LysE) are expressed in *C. glutamicum.*[3] Keeping this in mind, the coverage of IMPs by our approach should actually be much higher than the claimed 50%, which is based on the theoretical number of existing, not expressed, genes of IMPs.

In conclusion, we believe that a comprehensive functional characterization of the *C. glutamicum* membrane proteome has been obtained. However, the described strategy does not circumvent the limitation that protein identification relies on the presence of known sequences in a database. Using other strategies, *e.g.* "*de novo*" sequencing of the measured (and yet to be assigned) MS/MS spectra, may be useful to identify proteins that have not been annotated (29). The surprisingly high number of identified transporters and components of the respiratory chain suggests a complex metabolic network, which requires further analysis by the application of quantitative and comparative proteomics. We are certain that the new approach presented in this study will permit the complete functional analysis of similar complex organisms like other bacteria but also allows a more comprehensive analysis of membranes from the more complex eukaryotic cells.

---

[3] A. Burkowski, personal communication.

REFERENCES

1. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132
2. Zischka, H., Weber, G., Weber, P. J., Posch, A., Braun, R. J., Buhringer, D., Schneider, U., Nissum, M., Meitinger, T., Ueffing, M., and Eckerskorn, C. (2003) Improved proteome analysis of Saccharomyces cerevisiae mitochondria by free-flow electrophoresis. *Proteomics* **3**, 906–916
3. Xiong, Y., Chalmers, M. J., Gao, F. P., Cross, T. A., and Marshall, A. G. (2005) Identification of Mycobacterium tuberculosis H37Rv integral membrane proteins by one-dimensional gel electrophoresis and liquid chromatography electrospray ionization tandem mass spectrometry. *J. Proteome Res.* **4**, 855–861
4. Fujiki, Y., Hubbard, A. L., Fowler, S., and Lazarow, P. B. (1982) Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. *J. Cell Biol.* **93**, 97–102
5. Wu, C. C., and Yates, J. R. (2003) The application of mass spectrometry to membrane proteomics. *Nat. Biotechnol.* **21**, 262–267
6. Klein, C., Garcia-Rizo, C., Bisle, B., Scheffer, B., Zischka, H., Pfeiffer, F., Siedler, F., and Oesterhelt, D. (2005) The membrane proteome of Halobacterium salinarum. *Proteomics* **5**, 180–197
7. Washburn, M. P., Wolters, D., and Yates, J. R., III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247
8. Wu, C. C., MacCoss, M. J., Howell, K. E., and Yates, J. R. (2003) A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.* **21**, 532–538
9. Blonder, J., Conrads, T. P., Yu, L. R., Terunuma, A., Janini, G. M., Issaq, H. J., Vogel, J. C., and Veenstra, T. D. (2004) A detergent- and cyanogen bromide-free method for integral membrane proteomics: application to Halobacterium purple membranes and the human epidermal membrane proteome. *Proteomics* **4**, 31–45
10. Kinoshita, S., Udaka, S., and Shimono, M. (1959) Amino acid fermentation. I. Production of L-glutamic acid by various microorganisms. *J. Gen. Appl. Microbiol.* **3**, 193–205
11. Liebl, W. (1992) in *The Prokaryotes* (Balows, A., Trueper, H. G., Dworkin, M., Harder, W., and Schleifer, K.-H., eds) Vol. 2, 2nd Ed., pp. 1157–1171, Springer, New York
12. Eggeling, L., and Sahm, H. (2001) The cell wall barrier of Corynebacterium glutamicum and amino acid efflux. *J. Biosci. Bioeng.* **92**, 201–213
13. Vrljic, M., Sahm, H., and Eggeling, L. (1996) A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum. *Mol. Microbiol.* **22**, 815–826
14. Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B. J., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Kramer, R., Linke, B., McHardy, A. C., Meyer, F., Mockel, B., Pfefferle, W., Puhler, A., Rey, D. A., Ruckert, C., Rupp, O., Sahm, H., Wendisch, V. F., Wiegrabe, I., and Tauch, A. (2003) The complete Corynebacterium glutamicum ATCC 13032 genome se-

quence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J. Biotechnol.* **104,** 5–25

15. Ikeda, M., and Nakagawa, S. (2003) The Corynebacterium glutamicum genome: features and impacts on biotechnological processes. *Appl. Microbiol. Biotechnol.* **62,** 99–109

16. Hermann, T., Pfefferle, W., Baumann, C., Busker, E., Schaffer, S., Bott, M., Sahm, H., Dusch, N., Kalinowski, J., Puhler, A., Bendt, A. K., Kramer, R., and Burkovski, A. (2001) Proteome analysis of Corynebacterium glutamicum. *Electrophoresis* **22,** 1712–1723

17. Schaffer, S., Weil, B., Nguyen, V. D., Dongmann, G., Gunther, K., Nickolaus, M., Hermann, T., and Bott, M. (2001) A high-resolution reference map for cytoplasmic and membrane-associated proteins of Corynebacterium glutamicum. *Electrophoresis* **22,** 4404–4422

18. Santoni, V., Molloy, M., and Rabilloud, T. (2000) Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21,** 1054–1070

19. Schluesener, D., Fischer, F., Kruip, J., Rogner, M., and Poetsch, A. (2005) Mapping the membrane proteome of Corynebacterium glutamicum. *Proteomics* **5,** 1317–1330

20. Molloy, M. P., Herbert, B. R., Slade, M. B., Rabilloud, T., Nouwens, A. S., Williams, K. L., and Gooley, A. A. (2000) Proteomic analysis of the Escherichia coli outer membrane. *Eur. J. Biochem.* **267,** 2871–2881

21. Rigaud, J. L., Pitard, B., and Levy, D. (1995) Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins. *Biochim. Biophys. Acta* **1231,** 223–246

22. Blonder, J., Goshe, M. B., Moore, R. J., Pasa-Tolic, L., Masselon, C. D., Lipton, M. S., and Smith, R. D. (2002) Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **1,** 351–360

23. Duewel, H. S., and Honek, J. F. (1998) CNBr/formic acid reactions of methionine- and trifluoromethionine-containing lambda lysozyme: probing chemical and positional reactivity and formylation side reactions by mass spectrometry. *J. Protein Chem.* **17,** 337–350

24. Bott, M., and Niebisch, A. (2003) The respiratory chain of Corynebacterium glutamicum. *J. Biotechnol.* **104,** 129–153

25. Russell, W. K., Park, Z. Y., and Russell, D. H. (2001) Proteolysis in mixed organic-aqueous solvent systems: applications for peptide mass mapping using mass spectrometry. *Anal. Chem.* **73,** 2682–2685

26. Cech, N. B., and Enke, C. G. (2001) Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.* **20,** 362–387

27. Gu, S., Chen, J., Dobos, K. M., Bradbury, E. M., Belisle, J. T., and Chen, X. (2003) Comprehensive proteomic profiling of the membrane constituents of a *Mycobacterium tuberculosis* strain. *Mol. Cell. Proteomics* **2,** 1284–1296

28. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003) Global analysis of protein expression in yeast. *Nature* **425,** 737–741

29. Lu, B., and Chen, T. (2004) Algorithms for de novo peptide sequencing using tandem mass spectrometry. *BIOSILICO* **2,** 85–90