A High-throughput Approach for Subcellular Proteome

IDENTIFICATION OF RAT LIVER PROTEINS USING SUBCELLULAR FRACTIONATION COUPLED WITH TWO-DIMENSIONAL LIQUID CHROMATOGRAPHY TANDEM MASS SPECTROMETRY AND BIOINFORMATIC ANALYSIS*

Xiao-Sheng Jiang‡, Hu Zhou‡, Lei Zhang‡, Quan-Hu Sheng‡, Su-Jun Li‡§, Long Li‡, Pei Hao§, Yi-Xue Li§, Qi-Chang Xia‡, Jia-Rui Wu‡, and Rong Zeng‡¶

Four fractions from rat liver (a crude mitochondria (CM) and cytosol (C) fraction obtained with differential centrifugation, a purified mitochondrial (PM) fraction obtained with nycodenz density gradient centrifugation, and a total liver (TL) fraction) were analyzed with two-dimensional liquid chromatography tandem mass spectrometry analysis. A total of 564 rat proteins were identified and were bioinformatically annotated according to their physico-chemical characteristics and functions. While most extreme alkaline ribosomal proteins were identified in the TL fraction, the C fraction mainly included neutral enzymes and the PM fraction enriched alkaline proteins and proteins with electron transfer activity or oxygen binding activity. Such characteristics were more apparent in proteins identified only in the TL, C, or PM fraction. The Swiss-Prot annotation and the bioinformatic prediction results proved that the C and PM fractions had enriched cytoplasmic or mitochondrial proteins, respectively. Combination usage of subcellular fractionation with two-dimensional liquid chromatography tandem mass spectrometry was proved to be a high-throughput, sensitive, and effective analytical approach for subcellular proteomics research. Using such a strategy, we have constructed the largest proteome database to date for rat liver (564 rat proteins) and its cytosol (222 rat proteins) and mitochondrial fractions (227 rat proteins). Moreover, the 352 proteins with Swiss-Prot subcellular location annotation in the 564 identified proteins were used as an actual subcellular proteome dataset to evaluate the widely used bioinformatic tools such as PSORT, TargetP, TMHMM, and GRAVY. Molecular & Cellular Proteomics 3:441–455, 2004.

The proteome of any cell, tissue, or organism is a complex mixture of proteins that span a wide range of size, relative abundance, acidity/basicity, and hydrophobicity. The separation of the protein mixture into organelles or other multiprotein complex fractions prior to a proteomics analysis is usually the first step to increase the probability of detecting low-copy-number proteins (1–5). Subcellular fractionation and purification of organelles provide attractive additions to protein separation techniques commonly used in proteomic analysis. There has been a tendency to focus on subcellular proteomes concerning specific subcellular compartments and macromolecular structures of the cell (6). Subcellular proteomics research cannot only provide information about subcellular location of certain protein and imply its function (2), but also tell us the whole-protein components of the specific subcellular fraction (organelle or other multiprotein complex) (3) and then help understand their structures (4) and biological functions (5).

Chronologically, the most widely used method for the complex protein mixtures prior to mass spectrometry (MS)¹ analysis is two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) (7–9), followed by enzymatic digestion of the separated protein spots. 2D-PAGE is capable of detecting more than 10,000 protein spots theoretically (10) and gives valuable additional information on experimental pl and Mr values of the proteins to help in protein identification (11). Moreover, the visible separation of protein isoforms on 2D-PAGE gels is often the result of post-translational protein modifications (12). Although the resolving power of 2D-PAGE is excellent, the identification of individual spot requires a second analytical step such as MS. The extraction, digestion, and analysis of each spot from 2D-PAGE are a tedious and time-consuming process. Automation of this process requires expensive

¹ The abbreviations used are: MS, mass spectrometry; 2D-PAGE, two-dimensional polyacrylamide gel electrophoresis; IMP, integral membrane proteins; MDLC, multidimensional liquid chromatography; 2D-LC, two-dimensional liquid chromatography; MS/MS, tandem mass spectrometry; DALPC, direct analysis of large protein complexes; MudPIT, multidimensional protein identification technology; CM, crude mitochondria; PM, purified mitochondria; C, cytosol; TL, total liver; TM, transmembrane; GRAVY, grand average hydropobicity; GO, Gene Ontology; SCX, strong cation exchange; RP, reversed phase; ER, endoplasmic reticulum.
robotics to cut out and process the spots. The 2D-PAGE has other limitations, concerning the detection of low-abundance proteins, hydrophobic proteins, and proteins with extreme size and charge values (9, 13, 14). In particular, the analysis of integral membrane proteins (IMPs) remains a critical challenge, although new detergents have been designed to enhance membrane protein solubility for analysis by 2D-PAGE (15, 16).

An alternative approach to 2D-PAGE, multidimensional liquid chromatography (MDLC), has emerged to directly interface protein and peptide separations to mass spectrometers. Giddings demonstrated that the overall peak capacity of multidimensional separations is the product of the peak capacities in each independent dimension only if the separation dimensions are orthogonal and components separated in one dimension remain separated in any additional separation dimension (17). To achieve this criterion, column chromatography with independent stationary phases has been coupled together in different system combinations such as LC-capillary electrophoresis or capillary isoelectric focusing-LC (18–21). One strategy employing multidimensional liquid separations for protein identification in protein complexes is direct analysis of large protein complexes (DALPC) (22) by multidimensional protein identification technology (MudPIT) (20), which combines MDLC with electrospray ionization tandem mass spectrometry (MS/MS). The MDLC method integrates a strong cation exchange (SCX) resin and a reversed-phase resin in a biphasic column. This strategy was employed to the proteome of Saccharomyces cerevisiae strain BJ5460 and identified 1,484 proteins of which 131 have three or more predicted transmembrane domains (23). Directly identifying proteins from complexes bypasses the potential limitations of gel electrophoresis, including protein insolubility, limited fractionation ranges, and limited recoveries of material. MudPIT was demonstrated a dynamic range of 10,000 to 1 between the most abundant and least abundant proteins/peptides in a complex peptide mixture (20). In addition, DALPC provides a highly automatic system and rapid process for repeated analysis of protein complexes.

Bioinformatics is an integral part of proteomics research including MS data analysis and interpretation, analysis and storage of the gel images to databases, gel comparison, and advanced methods to study e.g. protein co-expression, protein-protein interactions, as well as metabolic and cellular pathways (24). With experimentally verified information on protein function lagging far behind, computational methods are needed for reliable and large-scale functional annotation of proteins. Prediction of in vivo fates of proteins such as function (25), subcellular location (26–29), modification (28), hydrophobicity, and membrane protein structure has become an even more important theme of bioinformatics. The PSORT and TargetP programs have been popular tools used for protein subcellular location prediction. The PSORT program can localize proteins in 17 different subcellular localizations (10), with a newer, retrained version called PSORT 2 that uses a slightly different decision algorithm and integrates a number of pre-existing prediction programs as well as calculated characteristics such as overall amino acid composition within a unified framework (29, 30). The TargetP predictor has more limited prediction scope than PSORT and only classifies proteins to secretary, mitochondrial, chloroplast proteins (for plant only), and others. The method looks for N-terminal sorting signals by feeding the outputs from SignalP, ChloroP, and an analogous mitochondrial predictor into a “decision neural network” that makes the final choice between the different compartments (29, 31). The grand average hydrophobicity (GRAVY) values determined according to Kyte and Doolittle (32) provide an image of the hydrophobicity of the whole protein. GRAVY values usually vary in the range ±2. Positive scores indicate hydrophobicity and negative scores indicate hydrophilicity. The TMHMM program (33) was based on a hidden Markov model and was used to predict the theoretical transmembrane (TM) domains. The TMHMM program was claimed to predict 97–98% of TM helices correctly and was applied to a number of proteome data (33).

The rat is a useful, widely used animal model for biological and toxicity studies. Rat liver is one of the most important organs involved in physiological, pathological, and toxicological activities. Proteomic research on rat liver has great significance. At present, many 2D-PAGE databases of rat liver (34, 35) or its subcellular fractions such as mitochondria (36, 37), Golgi complex (38, 39), cytosol (35), nuclear pore complex (40), and mitochondrial ribosome (41) have been established. After 2D-PAGE separation of proteins followed by in-gel enzymatic digestion and MS identification, 113 unique proteins were identified from 163 protein spots among 5,222 protein spots in 78 2D-PAGE gels for rat liver (34); 273 unique proteins were identified from the total liver and cytosolic fractions, 20% of which were detected only in the cytosol fraction (35); 192 unique proteins were identified from the mitochondrial fraction (37); and 47 unique proteins were identified from the Golgi complex fraction (39). Though 2D-PAGE separations produced hundreds or even thousands of resolved spots, only a few of them have been correlated to proteins. New strategy should be applied in proteomics, especially subcellular proteomics, research.

In this work, we provided a high-throughput strategy for subcellular proteomics research: identification of proteins from subcellular fractions using 2D-LC-MS/MS followed by bioinformatics annotation. Such a strategy was applied to rat liver subcellular proteome research. The four fractions from rat liver: crude mitochondria (CM) and cytosol (C) fractions obtained with differential centrifugation, a purified mitochondrial (PM) fraction obtained with nycodenz density gradient centrifugation, and a total liver (TL) fraction, were analyzed with 2D-LC-MS/MS analysis. A total of 564 rat proteins were identified and bioinformatically annotated according to their physicochemical characteristics such as molecular mass, pl,
hydrophobicity, and TM domain, subcellular location annotated in Swiss-Prot database or predicted by PSORT or TargetP, and function family categorized from universal Gene Ontology (GO) annotation terms. This strategy has proved to be a high-throughput, sensitive, effective, and largely unbiased analytical approach for subcellular proteomics research.

EXPERIMENTAL PROCEDURES

Materials

Water was prepared using a Milli-Q system (Millipore, Bedford, MA). Nycodenz, formic acid, guanidine hydrochloride, sodium orthovanadate (Na4VO4, and sodium fluoride (NaF) were obtained from Sigma (St. Louis, MO). Acetonitrile with high-performance LC grade was obtained from Fisher (Fair Lawn, NJ). Trypsin (sequencing grade) was obtained from Roche (Mannheim, Germany). EDTA, EGTA, and phenylmethylsulfonyl fluoride were purchased from Amresco (Solon, OH). Adult male Sprague-Dawley rats were purchased from Shanghai Laboratory Animal Center (Jiu-Ting, Shanghai, China).

Differential Centrifugation Separation of Rat Liver Subcellular Fractions

Subcellular fractionation of rat liver was performed according to the procedure of Ayako and Fridovich (42) with minor modifications. Briefly, Sprague-Dawley rats were sacrificed and the livers were promptly removed and placed in ice-cold homogenization buffer consisting of 200 mM mannitol, 50 mM sucrose, 1 mM EDTA, 0.5 mM EGTA, and a mixture of protease inhibitor (1 mM phenylmethylsulfonyl fluoride) and phosphatase inhibitors (0.2 mM Na4VO4, 1 mM NaF) and 10 mM Tris-HCl at pH 7.4. After mincing with scissors and washing to remove blood, the livers were homogenized in a Potter-Elvejhem homogenizer with a Teflon piston, using 10 ml of the homogenization buffer per 2 g of tissue. Centrifugation at successively higher speeds at 4 °C yielded the following fractions: crude nuclear fraction at 1,000 × g for 10 min; mitochondria at 15,000 × g for 15 min; and microsomes at 144,000 × g for 90 min. The final supernatant was the cytosol fraction. Each successive pellet was washed three times with the homogenization buffer. The centrifuges used were the Himac CR 21G high-speed refrigerated centrifuge and Himac CP 80MX preparative ultracentrifuge, both from Hitachi Koki Co. Ltd. (Tokyo, Japan).

Purification of Rat Liver Mitochondria through Nycodenz Density Gradient Centrifugation

The procedures recommended by Nycomed Pharma and Invitrogen Life Technologies were followed (42). Nycodenz was dissolved to 50% (w/v) in 5 mM Tris-HCl, pH 7.4, containing 1 mM EDTA, 0.5 mM EGTA and a mixture of protease inhibitor and phosphatase inhibitors as above. This stock solution was diluted with buffer containing 0.25 M sucrose, 5 mM Tris-HCl, and 1 mM EDTA, 0.5 mM EGTA, and a mixture of protease inhibitor and phosphatase inhibitors at pH 7.4. The crude mitochondrial pellets obtained from differential centrifugation were suspended in 12 ml of 25% nycodenz and placed on the following discontinuous nycodenz gradients: 5 ml of 34% and 8 ml of 30%, and this was topped off with 8 ml of 23% and finally 3 ml of 20%. The sealed tubes were centrifuged for 90 min at 52,000 × g at 4 °C. The bands of particles seen after centrifugation have been identified by Nycomed Pharma and Invitrogen Life Technologies as follows: nuclei at the 40/50% interface; peroxisomes at the 34/40% interface; mitochondria at the 25/30% interface, lysosomes at the 15/20% interface, and Golgi membranes at the 10/15% interface (42). The band at the 25/30% interface was collected and diluted with the same volume homogenization buffer and then centrifuged at 15,000 × g for 20 min.

Protein Preparation

For preparation of the total protein extract of rat liver (TL), rat liver tissue (1.0 g) was suspended in 10 ml of lysis buffer consisting of 8 M urea, 4% 3-(3-cholamidopropyl)dimethylammonio)-1-propanesulfonate (CHAPS), 65 mM dithiothreitol, 40 mM Tris. The suspension was homogenized for ~1 min, sonicated at 100 W for 30 s, and centrifuged at 25,000 × g for 1 h. The supernatant contained the TL proteins solubilized in the isoelectric focusing-compatible agents.

2D-LC-MS/MS

Orthogonal 2D-LC-MS/MS was performed using a ProteomeX work station (Thermo Finnigan, San Jose, CA). The system was fitted with a SCX column (320 μm ID × 100 mm, DEV SCX, Thermo Hypersil-Keystone) and two C18 reversed-phase (RP) columns (180 μm × 100 mm, BioBasic® C18, 5 μm, Thermo Hypersil-Keystone). The salt steps used were 0, 25, 50, 75, 100, 150, 200, 400, and 800 mM NH4Cl synchronized with nine 140-min RP gradients. RP solvents were 0.1% formic acid in either water (A) or acetonitrile (B). The setting of the LCQ Deca Xplus ion-trap mass spectrometer is as follows: one full MS scan was followed by three MS/MS scans on the four datasets were correlated with the combined hu-
man, mouse, and rat database or only the rat database extracted from a nonredundant database comprised of Swiss-Prot, GenPept, and PIR entries and a six-way translation of dbEST downloaded from

Molecular & Cellular Proteomics 3.5 443
NCBI on March 1, 2003, using the program SEQUEST running on a DEC Alpha work station (43). SEQUEST results were interpreted using a conservative criteria set according to Washburn et al. (23) with minor modification. Briefly, all accepted results had a Cn of 0.1 or greater, a value shown to lead to high confidence in a SEQUEST result (43, 44). A singly charged peptide must be tryptic, and the cross-correlation score (Xcorr) had to be at least 1.8. Tryptic or partially tryptic peptides with a charge state of +2 must have a cross-correlation score of at least 2.2. Triply charged trypsin or partially tryptic peptides with a +3 charge state were accepted if the Xcorr was ≥3.7. When a protein was identified by four or more unique peptides possessing SEQUEST scores that passed the above criteria, no visual assessment of spectra was conducted and the protein was considered present in the mixture. When a protein was identified by three or fewer unique peptides possessing SEQUEST scores that passed the above criteria, at least one of SEQUEST results was visually assessed using criteria described previously to confirm or deny the presence of a protein (22, 44). That is to say, protein identifications based on mass spectra correlating to one or more unique tryptic peptides were considered as valid identifications. Single peptides that alone identify a protein were manually validated after meeting the following criteria. First, the SEQUEST cross-correlation score must be ≥1.8 for a +1 tryptic peptide or ≥2.2 for a +2 tryptic peptide or ≥3.7 for a +3 tryptic peptide. Second, the MS/MS spectrum must be of good quality with fragment ions clearly above baseline noise. Third, there must be some continuity to the b or y ion series. Fourth, the y ions that correspond to a proline residue should be intensive ions. Fifth, unidentified, intense fragment ions either correspond to +2 fragment ions or the loss of one or two amino acids from one of the ends of the peptide. After going through this process, we are fairly confident of protein identification.

**RESULTS**

**High-throughput Identification of Rat Liver Proteins with 2D-LC-MS/MS**

The MS/MS spectra acquired from equivalent normalized aliquots of the four respective subcellular fractions were searched against the combined human, mouse, and rat non-redundant database or rat database extracted from Swiss-Prot, GenPept, and PIR entries and a six-way translation of dbEST using the program SEQUEST running on a DECA work station. Table I shows the results using different database and Xcorr filters, which indicates the database and Xcorr filters could affect the search results significantly. In order to avoid false-positive hits, first we used data only from rat database to eliminate the low confidence of identification resulted from the nonspecificity of the database; second, we referred to the parameters reported in previous studies and applied more strict criteria for peptide identification (≥1.8 for +1 tryptic peptide, ≥2.2 for +2 tryptic peptide, and ≥3.7 for +3 tryptic peptide) than those in most reported work (22, 47); third, we manually checked the mass spectra of the identified peptides using different filters, which showed the current criteria could give good quality of the MS/MS spectra. This resulted in highly confident identification of a total of 564 unique rat proteins (2,042 unique peptides) in the four fractions. Of these, 350 proteins (1,130 unique peptides) were characterized in the TL fraction, 267 proteins (796 unique peptides) in the CM fraction, 222 proteins (661 unique peptides) in the C fraction, and 227 proteins (901 unique peptides) in the PM fraction (Table I). At the same time, 46 proteins were characterized in all the four fractions and 93 proteins were only characterized in the TL fraction, 53 only in the C fraction, 55 only in the CM fraction, and 61 only in the PM fraction (Fig. 1A). Compared with the TL fraction, the subcellular fractionation (C, CM, and PM) provided additional identification of 214 (37.9%) proteins. Table II shows the number of proteins identified according to 1, 2, 3, and >3 unique peptides. About 50% proteins in each fraction were identified from a unique peptide in a single run with 300 μg of proteins, consistent with

**Table I**

| Fraction | Xcorr ≥1.8 (+1), 2.2 (+2), 3.7 (+3); ΔCn ≥0.1 | Xcorr ≥1.5 (+1), 2.0 (+2), 2.5 (+3); ΔCn ≥0.1 |
|----------|---------------------------------------------|---------------------------------------------|
|          | Total peptides | Unique peptides | Unique proteins | Total peptides | Unique peptides | Unique proteins |
| Rat protein database | | | | | | |
| C        | 3,989 | 661 | 222 | 5,451 | 1,050 | 453 |
| CM       | 2,449 | 796 | 267 | 3,671 | 1,236 | 542 |
| PM       | 4,633 | 901 | 227 | 5,933 | 1,143 | 314 |
| TL       | 4,617 | 1,130 | 350 | 5,450 | 1,289 | 420 |
| Total    | 15,689 | 2,042 | 564 | 20,505 | 3,044 | 1,118 |
| Human, mouse, and rat protein database | | | | | | |
| C        | 4,938 | 1,040 | 692 | 7,847 | 2,360 | 1,803 |
| CM       | 7,292 | 1,419 | 775 | 11,124 | 3,076 | 2,131 |
| PM       | 5,650 | 1,258 | 579 | 7,707 | 1,974 | 1,097 |
| TL       | 5,045 | 1,352 | 767 | 7,059 | 2,088 | 1,278 |
| Total    | 22,925 | 3,261 | 1,816 | 33,737 | 7,192 | 4,979 |
of the protein identification using current identification filters and criteria. It is noted that most of the proteins identified by a single peptide are also only found in a certain fraction, which indicates those proteins are low-abundant components in the cell and enriched after fractionation. On the other hand, proteins identified by >3 unique peptides account for 56% in the common proteins found in all fractions, which should be high-abundant proteins (Table II). Only 300 μg of proteins were loaded for 2D-LC-MS/MS, and only data from single run of 22 h were used in this work. The 564 proteins present the largest database of rat liver proteome so far. Fig. 1, B and C present the distribution of various physicochemical characteristics and functions of the 564 proteins. Compared with the traditional 2D-PAGE method, the 2D-LC-MS/MS strategy presents a number of data with very rapid speed and limited sample consumption. Most importantly, the heart of "omics" is based on the large datasets. This shotgun assay we used overcomes the low yields of unique gene products identified in 2D-PAGE resulting from multiple spots corresponding to a single gene product (35, 37), which can provide large-scale and extensive data for further bioinformatic analysis and evaluation of the subcellular fractionation in our work.

**Physicochemical Characteristics of the Identified Proteins**

The 564 identified proteins were classified according to different physicochemical characteristics such as molecular mass, pl, hydrophobicity (GRAVY value), and TM domain predicted by TMHMM. The protein distribution patterns of the four fractions were compared with the results from the above characteristics (Fig. 2).

In the present work, more proteins with molecular mass above 100 kDa and below 10 kDa were observed than using 2D gel separation. The smallest and the largest molecular mass obtained are 1.3 kDa and 532.0 kDa, respectively. For the 564 proteins, 419 (74.3%) proteins distribute among 10–60 kDa molecular mass intervals, which are compatible with general 1D-PAGE or 2D-PAGE, while there are 18 (3.2%) proteins with mass <10 kDa and 48 (8.5%) proteins with mass >100 kDa, beyond the general 1D-PAGE or 2D-PAGE separation limits (Fig. 2A). It is more interesting that in the proteins only found in a certain fraction, the proteins with mass <10 kDa and >100 kDa are dramatically increased, which indicate those proteins also are low-abundant components and enriched after fractionation.

Regarding the pl distribution, the 564 proteins distribute across a wide pl range (3.6–12.3). A total of 493 proteins (87.4%) distribute among pl 4–10 intervals but 14 (2.5%) proteins have pl < 4.3 and 65 (11.5%) proteins have pl > 10, also beyond the 2D-PAGE separation capability. Interestingly, the pl distribution patterns are very different within the four respective fractions (Fig. 2B). The TL fraction has a similar pl distribution pattern to the total 564 proteins, most of which distribute among pl 4–9 except for a higher percentage

---

**Fig. 1.** A total of 564 rat proteins identified from the four fractions of rat liver. A shows the percentage of the proteins identified from the each fraction and overlapping of different fractions. For example, TL represents the proteins identified only in the TL fraction; TL/C means in TL and C fractions but not in CM or PM fractions; TL/C/CM indicates in TL, C, and CM fractions but not in the PM fraction, TL/C/CM/PM means in all the four fractions. B presents sensitivity to a wide variety of protein classes. The bars indicate the percentage of proteins identified in this study from a variety of protein classes. C indicates the functional classification of 372 of the 564 identified proteins according to GO annotation terms.
PM fraction, 93 (41.0%) proteins are within pI 8 intervals. On the contrary, among the proteins identified in the subcellular fractionation enriched alkaline proteins (pI 8/H11022 C fraction distributed in the pI 5 in different fractions. Most of the proteins only identified in the largest part of the TL-only proteins, while no protein with pI 8 only and PM-only proteins, while PM has more proteins in pI 7.3% and 8.2% components with pI 10 was detected in C-only proteins. On the other hand, there 14 proteins with GRAVY value have the theoretical pI –2.04 –0.072. Sixty-five (11.5%) proteins have positive values. Additionally, more proteins with GRAVY value –0.25 have been identified in the PM (55.1%) or CM (52.4%) fractions than in the C (46.8%) or TL (43.4%) fractions (Fig. 2C). For the theoretical TM domains predicted by TMHMM, 70 (12.4%) proteins of the total 564 proteins have one or more predicted TM domain (Fig. 1B), of which 11 proteins have three or more TM domains (Table III). In particular, eight of the 11 proteins with three or more TM domains are all annotated as IMPs, which mainly identified in the CM or PM fractions. Accordingly, more proteins with one or more TM domain were identified in the CM (17.2%) and PM (11.8%) fractions than in the TL (9.4%) and C (4.0%) fractions (Fig. 2D).

Subcellular Location Annotation

Swiss-Prot Annotation—About 62.4% (352/564) of the 564 proteins obtained subcellular annotation results in the Swiss-Prot database. Eighty-two (23.3%) were located as cytoplasmic, 78 (22.2%) as mitochondrial, 44 (12.5%) as ribosomal, 16 (4.5%) as nuclear, and 57 (16.2%) as membrane proteins (including organellar membrane proteins) (Fig. 3A; Table IV).

In 240 TL fraction proteins with Swiss-Prot annotation, 57 (23.8%) were annotated as cytoplasmic, 51 (21.3%) as mitochondrial, 43 (17.9%) as ribosomal, and 12 (5.0%) as nuclear. One hundred twenty-nine C fraction proteins have subcellular location annotation, including 66 (51.2%) cytoplasmic proteins, 17 (13.2%) mitochondrial proteins, and 3 (2.3%) nuclear proteins. For the PM fraction, 150 proteins have subcellular location annotation, including 14 (9.3%) cytoplasmic proteins, 74 (49.3%) mitochondrial proteins, and 4 (2.7%) nuclear proteins. As expected, cytoplasmic proteins are mainly identified in the C and TL fractions, and 94.9% (74/78) of the mitochondrial proteins appear in the PM fraction. Many proteins annotated as endoplasmic reticulum (ER), peroxisomal, Golgi, lysosomal, and nuclear were also included in the CM fraction but apparently decreased in the PM fraction. Most of the ribosomal proteins locate on the ER, which was separated from the cytoplasmic fraction in this work by density gradient centrifugation. Therefore, few ribosomal proteins were detected in the cytoplasmic fraction, and almost all the ribosomal proteins (43/44) were identified only in the TL fraction (Tables IV and V).

PSORT Prediction—In this work, PSORT was used to predict the subcellular location of the total 564 proteins. Prediction results indicated that the total of 564 proteins includes 54.6% cytoplasmic, 12.9% mitochondrial, and 8.7% extracellular proteins. However, the C fraction includes more cytoplasmic proteins (65.3%) than the TL (576.1%), CM (48.7%), and PM (48.5%) fractions, while the PM fraction includes

| Unique Peptides | Totala | TLb | Cb | CMb | PMb | TL/C/CM/PMb | TL-onlyc | C-onlyc | C-onlyd | CM-onlyd | PM-onlyd |
|-----------------|-------|-----|----|-----|-----|-------------|----------|--------|--------|----------|----------|
| 1               | 265 (47.0%) | 164 (46.9%) | 107 (48.2%) | 145 (54.3%) | 104 (45.8%) | 9.75 (21.2%) | 67 (72.0%) | 45 (84.9%) | 46 (83.6%) | 46 (75.4%) |
| 2               | 100 (17.7%) | 67 (19.1%) | 28 (12.6%) | 41 (15.4%) | 44 (19.4%) | 5.25 (11.4%) | 20 (21.5%) | 4 (9.4%) | 7 (12.7%) | 11 (18.0%) |
| 3               | 49 (8.7%) | 33 (9.4%) | 28 (12.6%) | 24 (9.0%) | 14 (6.2%) | 5.25 (11.4%) | 4 (4.3%) | 2 (3.8%) | 2 (3.6%) | 2 (3.3%) |
| >3              | 150 (26.6%) | 86 (24.6%) | 59 (26.6%) | 57 (21.4%) | 65 (28.6%) | 25.75 (56.0%) | 2 (2.2%) | 1 (1.9%) | 0 (0.0%) | 2 (3.3%) |
| Total           | 564   | 350 | 222 | 267 | 227 | 46           | 93       | 53     | 55     | 61       |

\( ^a \) A total of 564 proteins identified in the four TL, C, CM, and PM fractions. The peptides attributed to the identification of certain protein in different fractions were combined to calculate the unique peptides.

\( ^b \) Proteins identified in certain fraction.

\( ^c \) Proteins identified only in the TL, C, CM, or PM fractions.

\( ^d \) Proteins identified in all four TL, C, CM, and PM fractions. The protein numbers or percentages according to different unique peptides were calculated as the average of the numbers or percentages for the four different fractions.
FIG. 2. Distribution of the proteins identified from different fractions in relation to their theoretical molecular mass (A), pI (B), GRAVY (C) values, and the number of predicted helices (PredHels) (D). TL-only, C-only, CM-only, or PM-only represents the proteins identified only in the TL, C, CM, or PM fractions; TL/C/CM/PM represents the proteins identified in all the four fractions. The bars indicate the percentage of proteins in certain fractions.
more mitochondrial proteins (25.6%) than the TL (14.3%), C (9.5%), and CM (14.6%) fractions (Fig. 3B).

**TargetP Prediction**—The total of 564 proteins includes 20.7% mitochondrial, 17.6% secreted, and 61.7% other proteins predicted by TargetP (Fig. 3C). In accordance with the Swiss-Prot subcellular annotation and PSORT prediction results, the TargetP prediction results indicated that the PM fraction included twice as many mitochondrial proteins (40.5%) as the TL fraction (19.1%) and the CM fraction (22.5%) and three times as many as the C fraction (12.2%). In addition, the CM fraction has more secreted proteins (29.6%) than other fractions (PM 18.1%, TL 15.4%, C 9.0%), which might resulted from contamination of the Golgi complex and ER that are involved in the synthesis, maturation, and traffic of the secreted proteins.

**Evaluation of the Bioinformatics Tools Using the 352 Proteins with Swiss-Prot Annotation as a Test Dataset**

It has been reported that TargetP has a sensitivity of 0.89 and a specificity of 0.67 while PSORT has a sensitivity of 0.81 and a specificity of 0.60 for mitochondrial protein prediction in a non-plant test dataset (31). However, many problems are involved in the prediction (27), and it is questionable whether the efficiency still holds when applied to proteome data (28). To evaluate the separation and purification effect and prediction efficiency, the 352 proteins with subcellular location annotation in the Swiss-Prot database are used as a test dataset for analyzing the sensitivity and specificity of subcellular fractionation and those predictors for separating or predicting cytoplasmic, mitochondrial, membrane, and ribosomal proteins. The results are shown in Table V. In this simple actual test dataset, TargetP has a sensitivity of 0.71 and a specificity of 0.66 for mitochondrial protein prediction while PSORT has a sensitivity of 0.45 and a specificity of 0.66 for mitochondrial protein prediction and a sensitivity of 0.71 and a specificity of 0.31 for cytoplasmic protein prediction. The values implied that the two tools had been overestimated. But surprisingly, the combination usage of TargetP and PSORT has a high specificity up to 0.86 for mitochondrial protein prediction. The separation and purification effect is again shown when compared to the sensitivity and specificity of a different fraction for examining cytoplasmic, mitochondrial, membrane, or ribosomal proteins (Table V). It is excitingly to find the PM fraction has a high sensitivity up to 0.95 for examining the mitochondrial proteins with specificity of 0.49, which indicates that the PM fraction has a lowered complexity and provides more specific mitochondrial proteins.

Only 54% membrane proteins have one or more TM helices predicted by TMHMM. But eight proteins with three or more TM helices are all annotated as IMPs, which mainly identified in CM or PM fractions. So, when with three or more TM helices predicted by TMHMM, a protein would be intensively indicated as an IMP (Table III). Typically, 99% of cytoplasmic proteins have no TM domain predicted by TMHMM, and 89% of cytoplasmic proteins have a GRAVY value $\leq 0$.

**Functional Annotation**

The identified proteins were functionally categorized based on universal GO annotation terms (45). Three hundred seventy-two (66.0%) of the identified proteins were mapped to at least one annotation term within the GO molecular function category, including 19 (5.1%) electron transfer flavoproteins, 54 (14.5%) proteins with electron transfer activity, 48 (12.9%) proteins with oxygen binding activity, 40 (10.8%) ribosome constituent proteins, 62 (16.7%) proteins with enzyme activity, 28 (7.5%) proteins with metal ion binding activity, 14 (3.8%) proteins with transcription factor activity, and 19 (5.1%) proteins with translation factor activity (Fig. 1C). The PM fraction enriched electron transfer flavoproteins (10.0%), proteins with electron transfer activity (22.0%), and proteins with oxygen binding activity (18.7%), while the C fraction enriched more metabolic enzymes (Fig. 4). In each fraction-only protein, the functional classification presents more specificity. In the TL-only fraction, ribosomal proteins are the major
part (42.5%). As before mentioned, ribosomal proteins locate in the ER, which was contained in the TL fraction but removed in the C fraction. On the other hand, several ribosomal proteins can be found in the CM-only fraction and none found in the PM-only fraction. The CM fraction was from primary purification by differential centrifugation, in which may contain the ER component and thus are mixed with ribosomal proteins. After density gradient centrifugation, the ER component was removed more completely, resulting in none of the ribosomal constituents that were detected in the PM-only proteins. In the C-only proteins, enzymes occupy the greatest number. After differential centrifugation, the CM-only fraction

---

**Fig. 3.** Subcellular location distribution patterns of different fractions according to Swiss-Prot annotations (A), PSORT prediction results (B), and TargetP prediction results (C). TL-only, C-only, CM-only, or PM-only represents the proteins identified only in the TL, C, CM, or PM fractions; TL/C/CM/PM represents the proteins identified in all the four fractions. The bars indicate the percentage of proteins in certain fractions.
still includes many enzyme proteins, whose coverage (24.1%) is even more than that of the whole CM fraction (16.3%) and C-only fraction (22.6%). The CM fraction may be contaminated with ER and lysosome (as shown in Table IV), which contain many enzymes. After density gradient centrifugation, the enzyme proteins decrease obviously, while electron transfer flavoproteins, proteins with electron transfer activity and proteins with oxygen binding activity become the major parts of the PM-only proteins. These observations indicate that the 2D-LC-MS/MS provide a fast and direct strategy to evaluate the results of subcellular fractionation and assess the purity of each faction.

**DISCUSSION**

In our study, we first combined the differential centrifugation, nycodenz density gradient centrifugation, and 2D-LC-MS/MS. Only one run of 2D-LC-MS/MS analysis of the trypsin-digested peptide mixtures from 300 μg of protein (which is far lower than the usual 1–2 mg of protein loaded in preparative 2D-PAGE gels) was performed for the four fractions (TL, C, CM and PM). With the strict and widely accepted SEQUEST criterion and only the rat protein database, a total of 564 unique rat proteins were identified in the four fractions, of which 350 proteins were characterized in the TL fraction, 222 proteins in the C fraction, and 227 proteins in the PM fraction (Table I). Although the number of proteins identified from all or each rat liver fraction in our study are far from the theoretical protein number of rat liver, which may have suffered from a limited protein database for rat compared with those of human and mouse, we have established the largest proteome database for rat liver and its cytosol and mitochondrial fractions at the present time. Compared with 2D-PAGE-MS or 1D-PAGE-LC-MS/MS, 2D-LC-MS/MS showed improved automation and high-throughput (24, 26, 50).

Our strategy appears to give an overall understanding of proteins in rat liver or their subcellular fractions, with little restrictions on the molecular mass, pI, hydrophobicity, and even membrane proteins. The proteins detected in 2D-PAGE gels are in general hydrophilic with negative GRAVY values (35, 49). For the 170 proteins identified by Fountoulakis et al. (35) from 2D-PAGE gels of rat liver, only 14 proteins had low positive GRAVY values (below 0.21). In our study, 65 (11.5%) hydrophobic proteins with GRAVY value >0 (up to 0.72) were identified. Without specific methods for enrichment or treatment of membrane proteins, we still identified 70 (12.4%) proteins from the total of 564 proteins having one or more predicted TM domain in which 11 proteins have three or more TM domains (Fig. 1B). In particular, eight of the 11 proteins with three or more TM domains are almost all known IMPs mainly identified in the CM or PM fraction (Table III). On the contrary, in the 170 proteins identified from rat liver or its cytosol sample through 2D-PAGE (35), only one protein with three TM domains has been identified. Furthermore, the 352 proteins with Swiss-Prot subcellular location annotation were used as an actual subcellular proteome dataset to evaluate the widely used bioinformatics tools for protein subcellular location prediction such as PSORT, TargetP, TMHMM, and GRAVY values. The sensitivity and specificity for each prediction tool might have been overevaluated previously according to our results.

Nycodenz is an iodinated density gradient media with os-
TABLE V
Evaluation of the efficiency of subcellular fractionation and PSORT, TargetP, and TMHMM prediction using the 352 proteins with Swiss-Prot annotation as a test dataset

| Test dataset | Subcellular fraction | Methods of prediction | GRAYV value | Molecular mass (kDa) | pI |
|--------------|----------------------|-----------------------|-------------|---------------------|----|
| Proteins with Swiss-Prot annotation | TL C CM PM | PSORT | TargetP | PSORT/TargetP | TMHMM (PredHel) | Cyto. | Mito. | Mito. | Mito. | 0 | ≥1 | ≥2 | ≥3 | ≤0 | >0 | <20 | 20–60 | 60–100 | >100 | <6 | 6–8 | 8–10 | >10 |
| Proteins in test dataset | 352 | 240 | 129 | 170 | 150 | 186 | 53 | 83 | 37 | 301 | 51 | 18 | 8 | 312 | 40 | 69 | 210 | 49 | 24 | 92 | 107 | 95 | 58 |
| Cytoplasmic | 82 | 57 | 66 | 32 | 14 | 58 | 5 | 5 | 1 | 81 | 1 | 0 | 0 | 73 | 9 | 14 | 57 | 7 | 4 | 29 | 35 | 18 | 0 |
| Sensitivity | 0.70 | 0.80 | 0.39 | 0.17 | 0.71 | 0.06 | 0.06 | 0.01 | 0.99 | 0.01 | 0.00 | 0.00 | 0.89 | 0.11 | 0.17 | 0.70 | 0.09 | 0.05 | 0.35 | 0.43 | 0.22 | 0.00 |
| Specificity | 0.24 | 0.51 | 0.19 | 0.09 | 0.31 | 0.09 | 0.06 | 0.03 | 0.27 | 0.02 | 0.00 | 0.00 | 0.23 | 0.23 | 0.20 | 0.27 | 0.14 | 0.17 | 0.32 | 0.33 | 0.19 | 0.00 |
| Mitochondrial | 78 | 51 | 17 | 45 | 74 | 38 | 35 | 55 | 32 | 68 | 10 | 5 | 1 | 61 | 17 | 16 | 53 | 7 | 2 | 9 | 21 | 37 | 11 |
| Sensitivity | 0.65 | 0.22 | 0.58 | 0.95 | 0.49 | 0.45 | 0.71 | 0.41 | 0.87 | 0.13 | 0.06 | 0.01 | 0.78 | 0.22 | 0.21 | 0.68 | 0.09 | 0.03 | 0.12 | 0.27 | 0.47 | 0.14 |
| Specificity | 0.21 | 0.13 | 0.26 | 0.49 | 0.20 | 0.66 | 0.66 | 0.86 | 0.23 | 0.20 | 0.28 | 0.13 | 0.20 | 0.18 | 0.23 | 0.25 | 0.14 | 0.08 | 0.10 | 0.20 | 0.39 | 0.19 |
| Membrane | 57 | 27 | 13 | 25 | 30 | / | / | / | / | 26 | 31 | 15 | 8 | 47 | 10 | 8 | 27 | 15 | 7 | 12 | 22 | 18 | 5 |
| Sensitivity | 0.47 | 0.23 | 0.44 | 0.53 | 0.09 | 0.40 | 0.26 | 0.14 | 0.82 | 0.18 | 0.14 | 0.47 | 0.26 | 0.12 | 0.21 | 0.39 | 0.32 | 0.09 |
| Specificity | 0.11 | 0.10 | 0.15 | 0.20 | 0.11 | 0.61 | 0.83 | 1.00 | 0.15 | 0.25 | 0.12 | 0.13 | 0.31 | 0.29 | 0.13 | 0.21 | 0.19 | 0.09 |
| Ribosomal | 44 | 43 | 1 | 7 | 2 | 6 | 0 | 44 | 0 | 0 | 0 | 43 | 1 | 24 | 20 | 0 | 0 | 3 | 0 | 2 | 39 |
| Sensitivity | 0.98 | 0.02 | 0.16 | 0.05 | 0.68 | 0.05 | 0.14 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.98 | 0.02 | 0.55 | 0.45 | 0.00 | 0.00 | 0.07 | 0.00 | 0.05 | 0.89 |
| Specificity | 0.18 | 0.01 | 0.04 | 0.01 | 0.16 | 0.04 | 0.07 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 | 0.14 | 0.03 | 0.35 | 0.10 | 0.00 | 0.00 | 0.03 | 0.00 | 0.02 | 0.67 |
| Total Proteins | 564 | 350 | 222 | 267 | 227 | 307 | 73 | 117 | 50 | 494 | 70 | 24 | 11 | 499 | 65 | 112 | 325 | 79 | 48 | 149 | 199 | 151 | 65 |

a PreHel, predicted helices.
b Cyto., cytoplasmic.
c Mito., mitochondrial.
d Percentage of the proteins annotated as cytoplasmic, mitochondrial, membrane, or ribosomal proteins that were identified in different fractions or predicted by different bioinformatics tools, respectively.
e Percentage of the proteins identified in different fractions or predicted by different bioinformatics that were annotated as cytoplasmic, mitochondrial, membrane, or ribosomal proteins, respectively.
f Includes membrane proteins annotated as organelle membrane proteins.
molarity and viscosity lower than those of sucrose gradients (51). It can be used in density gradient centrifugation with a high-speed mode rather than an ultracentrifuge mode (42, 51). In this report, crude mitochondria by differential centrifugation and purified mitochondria by nycodenz density gradient centrifugation were obtained. Many proteins annotated as cytoplasmic, ER, peroxisomal, Golgi, lysosomal, and nuclear proteins were included in the CM fraction but apparently decreased in the PM fraction. A total of 41.9% (93/222) of proteins identified in C fractions were also identified in the CM fraction, but for the PM fraction, the value decreased to 24.8% (55/222). Excitingly, for all the 78 proteins annotated as mitochondrial (see Tables IV and V and Fig. 3), 74 (94.8%) of them had been identified in the PM fraction; moreover, 21 (26.9%) of them had only been identified in the PM fraction. Pflieger et al. (52) had validated that the mitochondria contain a large number of membrane-associated and highly alkaline proteins in yeast mitochondrial proteome research. In this work, nearly half (48.4%) of the 227 proteins identified in the PM fraction distribute in pI > 8 and even in pI > 10, and 27 (11.9%) of them have one or more TM domain. Mitochondria proteins were dramatically enriched by the purification process. Thus purified subcellular fraction would be in favor of identification of more organelle-specific proteins.

Certain subcellular fractions enriched specific functional proteins. About 40~50% of the proteins identified in the PM fraction are predicted or annotated as mitochondrial proteins. Moreover, the PM fraction enriched electron transfer flavoproteins (10.0%), proteins with electron transfer activity (22.0%), and proteins with oxygen binding activity (18.7%) (Fig. 4). Many proteins annotated as cytoplasmic, ER, peroxisomal, Golgi, lysosomal, and nuclear proteins are included in the CM fraction but apparently decreased in the PM fraction. Table VI shows the identified proteins involved in the electron transport and oxidative phosphorylation occurring in mitochondria and their distribution in certain fractions. A total of 25 components involved in the electron transport and oxidative phosphorylation were obtained in this work. All of them could be found in purified mitochondria, while only three components (cytochrome c, α-ETF, and H+-transporting two-sector ATPase β chain) were detected in the C fraction. It is concluded that protein purification and fractionation are efficient at enriching specific proteins located in certain subcellular organelle. On the other hand, it is noted that the distribution difference between TL and CM fractions are not obvious. There are 17 and 16 proteins involved in the electron transport and oxidative phosphorylation found in TL and CM fractions, respectively. These results show that differential centrifugation separated mitochondria from other subcellular components but that CM still maintained large complexity. In addition, the proteins in electron transport and the oxidative phosphorylation pathway are relatively high-abundant proteins, and some

Fig. 4. Functional classification of 372 out of the 564 identified rat liver proteins according to GO annotation terms and their distributions in different fractions. TL-only, C-only, CM-only, or PM-only represents the proteins identified only in the TL, C, CM, or PM fractions; TL/C/CM/PM represents the proteins identified in all the four fractions. The bars indicate the percentage of proteins in certain fractions.
Mitochondria inner membrane
Mitochondria
Mitochondria inner membrane
Mitochondria inner membrane
Mitochondria inner membrane
Mitochondria inner membrane
Mitochondria inner membrane
Mitochondria inner membrane
Mitochondria inner membrane
Integral membrane
Mitochondria inner membrane
Mitochondria inner membrane
Mitochondria inner membrane
Mitochondria inner membrane
Mitochondria inner membrane
Mitochondria inner membrane

of the components were easily found in TL even with no purification of mitochondria. However, all of the 25 proteins were found in PM, which showed the efficient decrease of complexity of sample and enabled more mitochondria-specific proteins to be identified with the nycodenz density gradient centrifugation. Moreover, there are two IMPs involved in electron transfer and oxidative phosphorylation (ATP synthase E chain and cytochrome c oxidase subunit II) detected in this work, and both of them were only found in PM. The nycodenz density gradient centrifugation increases the capability of identification of membrane proteins that are usually low-abundant proteins. However, there are many proteins in the PM fraction possessing subcellular location other than mitochondrial, and about 4% of the TM proteins were detected in the C fraction. These phenomena could result from

i) contamination during the preparation, ii) the fact that other fraction or organelles are in close contact with certain fraction, or iii) cross-localization in different fractions (50).

In summary, we provided a strategy for subcellular proteomics research: identification of proteins from subcellular fractions using 2D-LC-MS/MS followed by bioinformatics annotation, which was proved as a high-throughput, sensitive, and effective analytical approach for subcellular proteomics research. It could be further used in the analysis of the membrane and soluble proteins of certain subcellular fractions, or other suborganelle proteomics research. Because this method is rapid and sensitive, it may be used to check the purity of subcellular fractions as an alternative to detection of marker proteins with specific antibody. Moreover, isotope-coded affinity tag technology has been successfully used in
quantitative and differential analysis of complex protein mixtures (53–57). Combination of isotope-coded affinity tag with subcellular fractionation may promote the quantitative and differential study between different subcellular fractions under different physiology or pathology states.

* This work was supported by National High-Techology Project (2001AA333031, 2002BA711A11) and Basic Research Foundation (2001CB210501). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

To whom correspondence should be addressed: Research Center for Proteome Analysis, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 YueYang Road, Shanghai 200031, China. Tel.: 86-21-54920170; Fax: 86-21-54920171; E-mail: zr@silbs.ac.cn.

REFERENCES

1. Rabilloud, T., Kieffer, S., Procaccio, V., Louwagie, M., Courchesne, P. L., Patterson, S. D., Martinez, P., Garin, J., and Lunardi, J. (1998) Two-dimensional electrophoresis of human placental mitochondria and protein identification by mass spectrometry: Toward a human mitochondrial proteome. Electrophoresis 19, 1006–1014.

2. Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajah, P., Steeeman, J., Lamond, A., and Mann, M. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. Nature Genet. 20, 46–50.

3. Taylor, R. S., Wu, C. C., Hays, L., Eng, J. K., Yate, J. R., and Howell, K. E. (2000) Proteomics of rat liver Golgi complex: Minor proteins are identified through sequential fractionation. Electrophoresis 21, 3441–3459.

4. Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T., and Matunis, J. I. (1998) Analyzing glycoproteins separated by two-dimensional gel electrophoresis with immobilized pH gradients. Electrophoresis 18, 307–316.

5. Chevallet, M., Santoni, V., Poinas, A., Rouquie, D., Fuchs, A., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J., and Rabilloud T. (1998) New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. Electrophoresis 19, 1901–1909.

6. Giddings, J. C. (1987) Concepts and comparisons in multidimensional separation. J. High Resolut. Chromatogr. Commun. 10, 319–323.

7. Opit, G. J., Jorgenson, J. W., and Anderegg, R. J. (1997) Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. Anal. Chem. 69, 2283–2291.

8. Tong, K., Link, A., Eng, J. K., and Yates, J. R., III (1999) Identification of proteins in complexes by solid-phase microextraction/multistep elution/capillary electrophoresis/tandem mass spectrometry. Anal. Chem. 71, 2270–2278.

9. Wolters D. A., Washburn M. P., and Yates, J. R., III (2001) An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. 73, 5683–5690.

10. Wagner, K., Miliotis, T., Marqu-Varga, G., Bischoff, R., and Unger, K. (2002) An automated on-line multidimensional HPLC system for protein and peptide mapping with integrated sample preparation. Anal. Chem. 74, 809–820.

11. Link, A. J., Eng, J., Schietz, D. M., Carmack, E., Mize, G. J., Morris, D. R., and Garvik, B. M., and Yates, J. R. III (1999) Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. 17, 676–682.

12. 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.

13. Vihinen, M. (2001) Bioinformatics in proteomics. Biomed. Engineering 18, 241–248.

14. Wu, C. H., Huang, H., Yeh, L. S., and Barker, W. C. (2003) Protein family classification and function annotation. Comput. Biol. Chem. 27, 37–47.

15. Nakai, K. (2000) Protein sorting signals and prediction of subcellular localization. Adv. Protein Chem. 54, 277–344.

16. Fen, Z. P. (2002) An overview on predicting the subcellular location of a protein. In Silico Biol. 2, 291–303.

17. Nakai, K. (2001) Review: Prediction of in vivo fates of proteins in the era of genomics and proteomics. J. Struct. Biol. 134, 103–116.

18. Emanuelsson, O., and Heijne, G. V. (2001) Prediction of organelar targeting signals. Biochim. Biophys. Acta 1541, 114–119.

19. Nakai, K., and Horton, P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci. 24, 34–35.

20. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300, 1005–1016.

21. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

22. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305, 567–580.

23. Thome-Kromer, B., Bonk, I., Klett, M., Nebrich, G., Taufmann, M., Bryant, S., Wacker, U., and Kopke, A. (2003) Toward the identification of liver proteins in complexes by solid-phase microextraction/multistep elution/mass spectrometry. J. High Resolut. Chromatogr. Commun. 26, 1862.

24. Fountoulakis, M., and Suter, L. (2002) Proteomic analysis of the rat liver. J. Chromatogr. B. 782, 197–218.

25. Lopez, F. M., Kristal, B. S., Chernokalskaya, E., Lazarev, A., Shestopalov, A. I., Bogdanova, A., and Robinson, M. (2000) High-throughput profiling of the mitochondrial proteome using affinity fractionation and automatic peptide mapping with integrated sample preparation. Electrophoresis 21, 3427–3440.

26. Fountoulakis, M., Berndt, P., Langen, H., and Suter, L. (2002) The rat liver mitochondrial proteins. Electrophoresis 23, 311–328.

27. Taylor, R. S., Flia, I., Jones, S. M., Huber, L. A., and Howell, K. E. (1997) Two-dimensional mapping of the endogenous proteins of the rat hepatocyte Golgi complex cleaved of proteins in transit. Electrophoresis 18, 2601–2612.

28. Taylor, R. S., Wu, C. C., Hays, L. G., Eng, J. K., Yate, J. R., and Howell, K. E. (2000) Proteomics of rat liver Golgi complex: Minor proteins are identified through sequential fractionation. Electrophoresis 21, 3441–3459.

29. Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T., and Matunis, M. J. (2002) Proteomics analysis of the mammalian nuclear pore com-
plex. J Cell Biol. 158, 915–927
41. Cahill, A., Baio, D. L., and Cunningham, C. C. (1995) Isolation and charac-
terization of rat liver mitochondrial ribosomes. Anal. Biochem. 232, 47–55
42. Ayako, O. M., and Fridovich, I. (2002) Subcellular distribution of superoxide
dismutases (SOD) in rat liver. J. Biol. Chem. 276, 38388–38393
43. Eng, J. K., McCormack, A. L., and Yates, J. R., Ill (1994) An approach to
correlate tandem mass-spectral data of peptides with amino-acid-se-
quencies in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989
44. Yates, J. R., Ill, Carmack, E., Hays, L., Link, A. J., and Eng, J. K. (1999)
Automated protein identification using micrololumn liquid chromatogra-
phy-tandem mass spectrometry. Methods Mol. Biol. 112, 553–569
45. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M.,
Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill,
D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson,
J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene Ontol-
gy: Tool for the unification of biology. The Gene Ontology Consortium.
Nat. Genet. 25, 25–29
46. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrett, D., Bateman,
A., Binns, D., Biswas, M., Bradley, P., Bork, P., Boucher, P., Copley, R. R.,
Courecelle, E., Das, U., Durbin, R., Falquet, L., Fleischmann, W., Griffiths-
Jones, S., Haft, D., Harle, N., Hulo, N., Kahn, D., Kanapin, A., Krestyana-
inova, M., Lopez, R., Letunic, I., Lonsdade, D., Silventoinen, V., Orchard,
S. E., Pagni, M., Peyruc, D., Ponting, C. P., Selegut, J. D., Servant, F.,
Sigrist, C. J. A., Vaughan, R., and Zdobnov, E. M. (2003) The InterPro
Database, 2003 brings increased coverage and new features. Nucleic
Acids Res. 31, 315–318
47. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., and Gygi, S. P. (2003)
Evaluation of multidimensional chromatography coupled with tandem
mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the
yeast proteome. J. Proteome Res. 2, 43–50
48. Mawuenyega, K. G., Kaji, H., Yamuchi, Y., Shinkawa, T., Saito, H., Taoka,
M., Takahashi, N., and Isebe, T. (2003) Large-scale identification of
Caenorhabditis elegans proteins by multidimensional liquid chromatog-
raphy-tandem mass spectrometry. J. Proteome Res. 2, 23–35
49. Fountoulakis, M., and Takacs, B. (2001) Effect of strong detergents and
chaotropes on the detection of proteins in two-dimensional gels. Elec-
rophoresis 22, 1593–1602
50. Taylor, S. W., Warnock, D. E., Glenn, G. M., Zhang, B., Fahy, E., Gaucher,
S. P., Capaldi, R. A., Gibson, B. W., and Ghosh, S. S. (2002) An alter-
native strategy to determine the mitochondrial protein using sucrose
gradient fractionation and 1D PAGE on highly purified human heart
mitochondria. J. Proteome Res. 1, 451–458
51. Graham, J. M., Ford, T., and Rickwood, D. (1990) Isolation of the major
subcellular organelles from mouse liver using nycodenz gradients with-
out the use of ultracentrifuge. Anal. Biochem. 187, 315–323
52. Pfieger, D., Le Caer, J.-P., Lemaire, C., Bernard, B. A., Dujardin, G., and
Rossier, J. (2002) Systematic identification of mitochondrial proteins by
LC-MS/MS. Anal. Chem. 74, 2400–2406
53. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold,
R. (1999) Quantitative analysis of complex protein mixtures using iso-
tope-coded affinity tags. Nat. Biotechnol. 17, 994–999
54. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling
of differentiation-induced microsomal proteins using isotope-coded af-
finity tags and mass spectrometry. Nat. Biotechnol. 19, 946–951
55. Zhou, H., Ranish, J. A., Watts, J. D., and Aebersold, R. (2002) Quantitative
proteome analysis by solid-phase isotope tagging and mass spectrom-
metry. Nat. Biotechnol. 20, 512–515
56. Ranish, J. A., Yi, E. C., Leslie, D. M., Purvine, S. O., Goodlett, D. R., Eng, J.,
and Aebersold, R. (2003) The study of macromolecular complexes by
quantitative proteomics. Nat. Genet. 33, 349–355
57. 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