Systematic Identification of Mitochondrial Proteins by LC–MS/MS

Delphine Pflieger, Jean-Pierre Le Caer, Claire Lemaire, Bruno Alain Bernard, Geneviève Dujardin, Jean Rossier

Systematic Identification of Mitochondrial Proteins by LC−MS/MS

Delphine Pflieger,*,† Jean-Pierre Le Caer,† Claire Lemaire,‡ Bruno Alain Bernard,# Geneviève Dujardin,‡ and Jean Rossier†

† Laboratoire de Neurobiologie et Diversite´ Cellulaire, CNRS, UMR 7637, Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris, 10 rue Vauquelin, 75231 Paris Cedex 05, France,

‡ Groupe Biologie du Cheveu, L'Oréal, Centre de recherche Charles Zviak, 90 rue du Général Roguet, 92110 Clichy, France,

#Centre de Génétique Moléculaire du CNRS, Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

Abstract

In eukaryotic cells, the mitochondrion is the key organelle for cellular respiration. Mitochondrial proteome analysis is difficult to perform by the classical proteomic approach involving two-dimensional gel electrophoresis (2DE), because this organelle contains a large number of membrane-associated and highly alkaline proteins usually requiring specific treatments to be successfully analyzed. Here, an alternative approach was evaluated and led to the rapid and sensitive identification of ~35% of the yeast mitochondrial proteins. It consists of an SDS-PAGE gel electrophoresis of the total mitochondrial protein in combination with the LC-MS/MS analysis of the digestion products of gel slices. The use of only 40 µg of mitochondrial protein enabled the identification of 179 different gene products divided into similar proportions of membrane and soluble proteins. The distribution of the identified proteins in terms of pI and hydrophobicity revealed that the present analytical strategy is largely unbiased. The identification of 28 proteins of previously unknown subcellular localization demonstrated the ability of SDSPAGE-LC-MS/MS to rapidly supplement the knowledge of the mitochondrial proteome.

Two-dimensional gel electrophoresis (2DE) was successfully used for the characterization of the proteome of various organisms. Nevertheless, at least in the case of complex eukaryotic cells, the number of proteins expressed largely exceeds the resolving power of 2DE. An elegant strategy to overcome this limitation consists of taking advantage of the subcellular compartmentalization of the cell and focusing the analysis on an organelle. This approach was adopted by a large number of researchers.[1] Nevertheless, 2DE suffers from other restrictions: proteins of low abundance are unlikely to be identified;[2] highly hydrophobic proteins are difficult to solubilize and maintain in solution during the first separation step;[3] and finally, proteins of extremely low and high molecular weights and very basic proteins are often badly focused on the gel when isoelectric conditions are not optimized. [4] To be rid of the limitations inherent to the isoelectrofocusing (IEF) step, alternative separation techniques have been developed. The most widely used approach combines one or several chromatographic separations with tandem mass spectrometry (LC^n−MS/MS). Briefly, after digestion of proteins in mixtures, the generated peptides are separated by liquid chromatography. The chromatographic system is coupled on-line with a tandem mass spectrometer that fragments automatically the eluted peptides. Both the separation and identification steps are thus performed in a single analysis. Depending on the complexity of the initial sample, LC−MS/MS analysis can be preceded by a separation step providing less complicated protein mixtures: SDS-PAGE, LC, or capillary isoelectric focusing (CIEF) are commonly used techniques.

Until now, few studies based on LC−MS/MS analysis have taken advantage of the subcellular organization of eukaryotic cells.[5-8] In the present work, a specific organelle, yeast mitochondria, was analyzed by the combination of SDS-PAGE and LC−MS/MS. *Saccharomyces cerevisiae* is an

attractive model organism, because its genome is completely sequenced; hence, no bias is encountered between known and unknown proteins. Moreover, mitochondria are an interesting validation model because they contain many membrane-bound and very basic proteins that are quite difficult to analyze by 2DE. Here, we first discuss the analysis conditions to identify a significant number of the proteins present in a total mitochondrial extract. In addition, the identified proteins are classified according to their isoelectric point (pI); molecular weight (MW); codon bias, which may be regarded as a rough indicator of their cellular abundance;(9) and hydrophobicity.

Material and Methods

Materials. Chemicals employed for gel electrophoresis were from Bio-Rad (Bio-Rad S. A., France). Dithiothreitol (DTT) and iodoacetamide were purchased from Sigma. Trypsin sequencing grade was obtained from Roche (Mannheim, Germany). The HPLC grade solvents used were acetonitrile and formic acid (FA).

Isolation of Mitochondria from *S. cerevisiae*. *S. cerevisiae* cells of strain CW04 (genotype *α ade2−1 ura3−1 his3−11,15 trp1−1 leu2−3,112 can1−100*) were grown in lactate medium (0.75% yeast extract, 0.75% bactopeptone, 0.5% lactic acid, 0.02 mg/mL adenine; pH adjusted to 4.5 with KOH 3 M). Cells were harvested in mid-exponential phase, rinsed with 10 mM EDTA, and resuspended in 1.2 M sorbitol buffer (1.2 M sorbitol, 50 mM Tris/HCl, pH 7.5, 10 mM EDTA, 0.3% 2-mercaptoethanol) at 3 mL/g wet mass cells. The cell wall was digested enzymatically with Zymolyase-100T (Seikagaku Kogyo Co.) at 1 mg/g of cells at 37 °C for 30 min. The spheroplasts were harvested by centrifugation at 1800g and were resuspended in ice-cold 0.7 M sorbitol buffer (0.7 M sorbitol, 50 mM Tris/HCl, pH 7.5, 0.2 mM EDTA). Cell fragments were subsequently removed by centrifugation at 2500g. Mitochondria were pelleted by centrifugation at 20000g and then washed three times in the same buffer except that protease inhibitors were added (complete cocktail from Roche Boehringer Mannheim, ref 1697498). The mitochondrial preparation was estimated to be pure at 90% according to electronic microscopic analysis, because it contained a few cell debris whose identity could not be established. The lysis of 5 × 10^6 cells provided ~40 μg of mitochondrial protein.

SDS-PAGE Separation and Protein Digestion. After addition of sample buffer (62.5 mM Tris/HCl pH 6.8, 2% SDS, 25% glycerol, 0.01% Bromophenol Blue), mitochondria were lysed by boiling at 100 °C for 5 min. Electrophoretic separation of proteins was carried out on a 0.75-mm-thick 8 × 10-cm gel prepared at 12.5% acrylamide. A 40-μg portion of total protein in sample buffer was loaded into a 7-mm well of the gel and separated at 40 mA. A total of 5.86 μg of standards (LMW electrophoresis calibration kit, Pharmacia Biotech) migrated in a neighboring lane. After coloration with colloidal Coomassie blue (Biosafe, Bio-Rad), the gel was regularly cut into 27 slices of equal size (around 2 mm) while avoiding as much as possible grouping intense and pale bands in the same slice. The relative protein amount in the slices was estimated on the digital image of the gel using the software Kodak Digital Science 1D. Gel slices were reduced and alkylated using DTT and iodoacetamide, respectively, and subjected to digestion with trypsin following the protocol of Shevchenko et al.[10] Extracted peptides were dried and stored at −23 °C until use. For LC−MS/MS analyses, peptides were resolubilized in buffer A (H$_2$O/acetonitrile/FA, 96/4/0.1, v:v). Typically one-sixth of the digestion product of a gel slice was used per LC−MS/MS analysis.

LC−MS/MS Analyses. Nanoscale capillary liquid chromatography-tandem mass spectrometric (LC−MS/MS) analyses of the digested proteins were performed using an UltiMate capillary LC system (LC Packings, Amsterdam) coupled to a hybrid quadrupole orthogonal acceleration time-of-flight tandem mass spectrometer (Q-TOF2, Micromass, Manchester, U.K.). The LC−MS union was made with a PicoTip (New Objective, Woburn, MA) fitted on a ZSPRAY (Micromass) interface. Chromatographic separations were conducted on a reversed-phase (RP) capillary column (Pepmap C18, 75-μm i.d., 15-cm length, LC Packings) with a 200 nL/min flow. The gradient profile used consisted of a linear gradient from 100% A to 45% B (H$_2$O/acetonitrile/FA, 10/90/0.085, v:v) in 50 min, followed by a linear gradient to 100% B in 10 min. Mass data acquisitions were piloted by Masslynx software (Micromass) using automatic switching between MS and MS/MS modes. Peptides eluted from the chromatographic column were detected for 1 s;

when their signal reached a user-defined threshold (4 counts/s), they could be selected for fragmentation. An MS/MS scan (1 s) was then performed on up to the 8 most intense peptide ions detected. MS/MS scans of each selected ion were summed until the total fragmentation time attributed to one selected precursor had been reached (see Results and Discussion) or the signal in an MS/MS scan had fallen to 0. Acquisitions were performed with the dynamic exclusion of m/z ratios of already fragmented ions (exclusion of a ±0.5 Da mass window around the m/z ratio of previously selected precursors). Fragmentation was performed using argon as the collision gas and with a collision energy profile optimized for various mass ranges of precursor ions.

Protein Identification. Mass data collected during a LC−MS/MS analysis were processed and converted into a .PKL file to be submitted to the search software Mascot (http://www.matrixscience.com/). Protein identifications were obtained by comparison of experimental data to the NCBInr database. No taxonomic restriction was used to reveal the presence of both proteins of interest and mammalian contaminants. Searches were done with a tolerance on mass measurement of 0.15 Da in MS mode and 0.25 Da in MS/MS mode. Nevertheless, the observed mass difference between a precursor ion and the corresponding to theoretical peptide assigned by Mascot was usually lower than 0.05 Da (20−50 ppm).

The identifications provided automatically by Mascot sometimes needed manual validation. A protein identification was regarded as reliable in automatic mode when (i) the identification was based on the fragmentation of more than 3 peptides and (ii) at least one of these peptides was associated with an MS/MS spectrum whose score calculated by Mascot was superior to 50. MS/MS spectra with scores above this value could usually be interpreted in a sequence of 5 amino acids or more (S/N > 5 for fragments leading to the sequence information). When these two conditions were not simultaneously fulfilled, the interpretation of the MS/MS spectra leading to the identification of a protein was checked according to the following criteria: (i) peaks of major intensities in the spectrum must correspond to fragments of the theoretical peptide; (ii) a series of consecutive y-type fragments must be readable starting from higher masses, typically leading to a sequence information of 5 amino acids.

Visualization of the Subcellular Localization of the GFP-Tagged Ygr086cp Protein. We tagged Ygr086cp (38 kDa) at its C-terminus with the green fluorescent protein (GFP, 27 kDa). Proteins of the mitochondrial pellet and the supernatant were separated on a 12% SDS-polyacrylamide gel, and the accumulation of the tagged protein was analyzed with anti-GFP antibodies by Western blotting. Cytochrome $c1$ (Cyt1p), a subunit of the mitochondrial respiratory complex $bc1$, was used as a control for mitochondria purification. Detection of $c1$ heme was performed as described by Vargas et al.[11]

Results and Discussion

Analysis Parameters. The optimization of protein analysis by LC−MS/MS is predominantly based on the selection of the most informative peptides. We first examined which charge states are most represented among precursor ions contributing to protein identifications. This optimization was performed on the SDS-PAGE−LC−MS/MS analysis of a total protein extract of *S. cerevisiae,* which was prepared according to Yaffe et al.[12] Analyses were conducted with no restriction on the charge state of precursor ions, and with 4 s of total fragmentation time for each selected species. The analysis of the tryptic digests of 10 gel slices distributed on the whole molecular weight range of the gel led to the identification of 174 proteins based on the fragmentation of 435 peptides and showed that protein identifications mostly came from the fragmentation of doubly and triply charged species. The former counted for 80.8% ± 7.0 of the peptides and the latter for 12.9% ± 4.8. The contribution of singly charged ions was 5.8% ± 4.0. In automatic mode, they generally gave poor MS/MS spectra, generating little sequence information. Moreover, a singly charged precursor ion of low m/z ratio usually corresponds to too short a peptide to establish a protein identification. Finally, the contribution of ions carrying 4 charges or more was not significant (<1%). On the basis of these results, the subsequent analyses of mitochondrial proteins by LC−MS/MS were restricted to the selection of doubly and triply charged precursor ions.

Another parameter to be adjusted in the LC−MS/MS experiments is the fragmentation duration of a given precursor ion. This parameter must meet two contradictory requirements. MS/MS signal must be accumulated long enough to give a spectrum that can be interpreted in terms of an amino acid sequence. On the other hand, this time of fragmentation has to be reduced as much as possible to enable a maximum of peptides to be analyzed during the LC−MS/MS run. In the analysis of the mitochondrial protein extract, fragmentation duration was varied between 6 and 10 s, depending on the color intensity of the digested gel slice so as to favor the use of the 8 MS/MS acquisition channels. Digests of pale gel slices were indeed expected to contain either less numerous peptides or peptides in smaller amounts than intensely colored slices. These samples allowed MS/MS acquisition duration to increase to 8 s or even 10 s without the 8 acquisition channels being used. Figure 1 represents the gel of mitochondrial extract and indicates which slices were regarded as intense or pale with an estimation of the corresponding protein content.
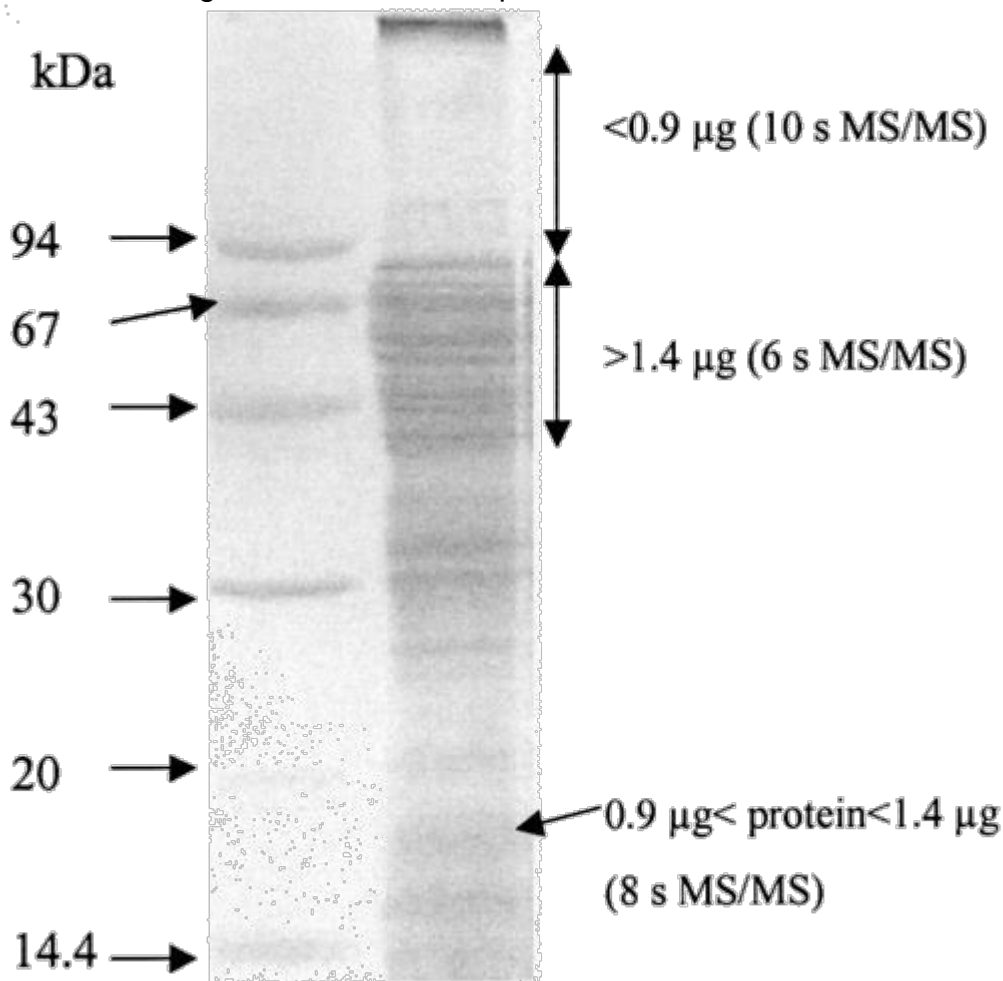

Figure 1 Image of the gel of mitochondrial protein extract used for LC−MS/MS analyses. Some regions of the gel are annotated, with an estimation of the protein quantity contained in 2 mm-large slices and with the MS/MS duration used for the analysis of their digest.

Finally, it was important to evaluate whether the SDS-PAGE−LC−MS/MS association gave sufficient separation of the large number of proteins to be analyzed in order to provide a significant overview of the complex sample studied.[13,14] Indeed, the number of digested proteins that can be successfully identified in one LC−MS/MS run is governed by the maximum number of peptides that can be characterized per time unit by the mass spectrometer. With the LC gradient used here, most of the tryptic peptides were eluted in a time window of 30 min. With a 6-s fragmentation time attributed to each selected precursor, the maximum number of peptides the mass spectrometer can analyze in MS/MS mode is ~300. The systematic identification of proteins present in a mixture, thus, requires the analysis of samples of limited complexity. In the present work, the first step of sample fractionation consisted in focusing the analysis on a single organelle. Since yeast proteins predicted to be mitochondrial (http://www.proteome.com/databases/YPD) represent near 8% of all theoretical proteins, mitochondria constituted a protein mixture ~10 times less

complicated than the whole cell. This type of fractionation had the further advantage of enabling the assignation of unknown proteins to that organelle. Each SDS-PAGE gel slice represented a protein mixture further simplified on average by the number of slices cut, in our case, 27.

The repeatability of protein identifications provided by the SDS-PAGE−LC−MS/MS was tested by performing three replicate analyses on the same digest. The protein identifications obtained from the analysis of gel slices distributed on the whole molecular weight range of the gel showed that the majority of the proteins could be identified in all three experiments. Yet, some proteins could only be characterized twice or even once. Proteins that were detected only once were most often identified on the basis of one unique peptide. These were likely to be proteins of low abundance whose identification was disfavored, since the automatic LC−MS/MS acquisition process selects the most intense ions.[15] More generally, the variability in protein identification results can be explained by a variability in precursor selection that is due to a variation in peptide retention times from run to run. During an LC−MS/MS analysis, one given ion may thus be detected in MS mode and then selected for fragmentation, or on the contrary, it may happen to elute within a time period during which the mass spectrometer works in MS/MS mode, performing no MS scan to select new precursors. In the case of highly complex protein mixtures, this factor then contributes to the variability of protein identifications.

This limitation of the SDS-PAGE−LC−MS/MS technique can be overcome by several means. The number of fragmented ions can be increased by improving the separation of peptides. This can be obtained by extending the LC gradient profile. Nevertheless, a slower gradient tends to broaden the elution of a given peptide and often reduces the S/N ratio in its MS/MS spectrum (data not shown). This may be critical in the case of peptides giving weak signals. Another approach consists of conducting two analyses of the same sample while restricting the precursor selection to two different mass ranges. This restriction typically reduces by a factor of 2 the number of candidates for fragmentation in each analysis. Such a strategy was employed by Spahr et al. for the analysis of tryptic peptide mixtures.[16] The authors emphasized the fact that complementary analyses performed on three to four different mass ranges brought more information than a single one, yet without precisely quantifying the gain in peptide and protein characterization.

Here, this second strategy was systematically studied on the digestion products of 5 gel slices of the mitochondrial extract. LC−MS/MS analyses were performed either by selecting the precursors on the whole mass range (*m/z*) [400, 1300], or by splitting the selections into the overlapping intervals, [400, 720] and [680, 1300]. Indeed, in our previous experiments on a total yeast extract, the analysis of 10 gel slices (415 doubly and triply charged peptides contributing to protein identifications) had indicated that tryptic peptides were similarly distributed into these two mass ranges. In addition, the quality of MS/MS spectra is also a critical parameter. As estimated from the analysis of four gel slices of the mitochondrial extract that produced 683 MS/MS spectra, 31.7% ± 4.6 of the MS/MS spectra could be automatically interpreted by Mascot software, whereas 47.3% ± 9.8 were of poor quality (too few or too weak fragment ions to be able to read as little as a three-residue sequence). These proportions are quite comparable to those observed by Simpson et al., who analyzed 400 μg of proteins from an enriched membrane preparation by a combination of SDS-PAGE and LC−MS/MS.[15] The authors studied in detail the MS/MS spectra acquired during the analysis of one digested gel slice and stated in particular that 40% of the spectra were automatically identified, whereas 41% were unusable. In an attempt to improve the quality of MS/MS spectra of ions giving weak signals, we then increased the MS/MS analysis time when working on [400, 720] and [680, 1300]. For dark blue gel slices, MS/MS fragmentation lasted 6 s when working on [400, 1300], and 10 s on [400, 720] or [680, 1300]; for pale gel slices, MS/MS durations were 10 and 15 s, respectively. The 15-s duration was unlikely to be effective for all ions: when more than five or six coeluted peptides were fragmented, some of them could not be detected long enough to benefit from the 15 s of MS/MS analysis. But this duration could be effective for peptides eluting at the beginning or the end of the gradient, when candidates for fragmentation were less numerous.

Mass data coming from the two analyses performed on [400, 720] and [680, 1300] were associated into a single .PKL file and submitted to the Mascot software. Table 1 shows a

comparison of the results of protein identifications obtained while working on the whole mass range of precursors or by associating the data collected on [400, 720] and [680, 1300]. Summing the mass data from the analyses performed on these two intervals had clear advantages. In general, the overall number of identified proteins was increased. The number of peptides contributing to each identification was always higher than that observed in the single analysis on [400, 1300]. Last, the quality of MS/MS spectra, evaluated by the cumulative scores of the identified proteins, was usually improved (data not shown). When the number of identified proteins did not increase, at least the reliability of the identifications was improved in terms of peptides and cumulative scores. To characterize even more proteins present in the mixtures, it would be conceivable to conduct the LC−MS/MS analysis on 3 different mass ranges or more. Another possibility would consist of adding a chromatographic separation before the RPLC−MS/MS analysis, for example with an SCX or a size exclusion column.[5,17,18] This scheme would thus provide a 3D separation of the mitochondrial proteins.

Table 1. Results of Protein Identification Obtained from the LC−MS/MS Analyses of the Digestion Products of Five Gel Slices[a]

gel (av molecular mass of proteins identified in gel slices, kDa)	no. identified proteins		no. peptides contributing to protein identifications	
	[400, 1300]	[400, 720] + [680, 1300]	[400, 1300]	[400, 720] + [680, 1300]
G1 (23)	16	16	58	69
G2 (27)	16	18	37	45
G3 (40)	28	29	73	84
G4 (60)	15	23	63	97
G5 (63)	18	24	64	79

[a] Designated G1 to G5. The analyses were performed either by selecting precursor ions on the whole mass range [400, 1300] Th or by restricting the selection of precursor ions on [400, 720] Th and [680, 1300] Th. Mass data collected during the analyses carried out on [400, 720] and [680, 1300] were associated into a single .PKL file for protein identification by Mascot software.

Identified Mitochondrial Proteins. The LC−MS/MS analyses of the 27 digested gel slices led to the identification of 179 different gene products out of about 500 expected yeast mitochondrial proteins. On the basis of the information available in the YPD web site (Yeast Protein Database, http://www.proteome.com/databases/YPD/), the identified proteins were characterized by their pI, MW, codon bias, and hydrophobicity. In addition, the biological relevance of the identified proteins was discussed.

Physicochemical Characteristics of the Identified Proteins. Tabulated values of pI and MW correspond to predictions for the mature form of the proteins as numerous mitochondrial proteins present an N-terminal presequence that is cleaved after import within the organelle. Possible posttranslational modifications that could significantly modify the pI and MW of a protein are not taken into account. Nevertheless, these tabulated values help to estimate which kinds of polypeptides can be successfully analyzed by LC−MS/MS.

Proteins distributed over the entire pI range [3.8−11.8] were identified. In particular, 28 proteins with pI above 10 and 75 proteins with pI above 9 were seen, which represents nearly one-half of the analyzed proteins. The identification of so many alkaline proteins is not conceivable with

standard 2DE techniques. Indeed, a few proteomic studies of mitochondria were carried out in the past few years.[19-21] Rabilloud et al. identified 58 proteins from human placenta mitochondria using 2DE gels ranging from pH 4−8; however, basic proteins were ignored, as the authors acknowledged.[19] Hanson et al. fractionated the proteins of human mitochondria by sucrose gradient centrifugation before running a 2DE separation on the pH range 3−10.[20] This method provides a better insight into the whole population of mitochondrial proteins and information about protein complexes within the organelle, but the number of gels to be run is multiplied by the number of protein fractions.

Another challenge concerns the MW of the analyzable proteins. Mitochondria contain a significant proportion of low MW proteins. YPD enables an evaluation with results that ~6% of yeast mitochondrial proteins are under 10 kDa. These proteins are often difficult to resolve by standard 2DE methods[20] and usually cannot be identified with confidence by mass fingerprinting alone. The small number of tryptic peptides they generate may also disfavor their identification by LC−MS/MS when analyzed in mixture with high MW proteins. This was emphasized by Washburn et al., who used two-dimensional chromatography (SCX and RP) to analyze the digestion product of three fractions of a total yeast protein extract.[17] This bias between small and large proteins may be largely prevented here, thanks to the separation step by SDS-PAGE. Our LC−MS/MS approach enabled the detection of 10 proteins with MW lower than 10 kDa out of 28 recorded in YPD. In addition, it was observed that 3 of the 16 recorded mitochondrial proteins were identified above 100 kDa.

Among the 179 identified proteins, 30 possess a codon bias below 0.1, suggesting a low level of expression. This value of codon bias can be regarded as the limit under which proteins are rarely identifiable by 2DE separation without previous enrichment.[22] It is interesting to note that the identifications have been made while starting from only 40 µg of mitochondrial protein, which demonstrates the excellent sensitivity of our LC−MS/MS system. Starting from 50 mg of total yeast protein extract separated by SDS-PAGE, Gygi et al. analyzed the digestion product of a gel strip by two-dimensional chromatography and identified ~60 proteins with a codon bias below 0.1 out of a total of 193 characterized proteins.[23] In that work, proteins with codon bias below 0.1 thus represented nearly one-third of all of the identifications. Here, with a 1000 times lower quantity of initial sample, this class of proteins represents nearly one-sixth of all of the identifications. A bias yet remains in favor of more abundant proteins, since 44% of the expected mitochondrial proteins possess codon bias values under 0.1.

Highly hydrophobic proteins constitute another class of polypeptides whose analysis by 2DE is not straightforward. Indeed, cumbersome sample-preparation steps such as organic solvent extraction,[24] solubilization of membrane proteins with detergents,[25] or prefractionation by isoelectric separation in a multicompartment electrolyzer[26] have been tested. In the present work, 57% of the identified mitochondrial proteins are expected to be soluble and 43% to be membrane-bound. According to YPD, as many as 23% of the identified proteins possess transmembrane segments. A more detailed analysis was performed on identified proteins belonging to the respiratory chain. The numbers of transmembrane domains were either determined from the structure of the yeast *bc*1 complex[27] and the beef cytochrome *c* oxidase[28] or were estimated by sequence analysis with the algorithm of Kyte and Doolittle using an 11-residue window.[29] A total of 33 structural subunits of respiratory chain complexes were identified, among which 17 are soluble and 16 are membrane-bound, exhibiting from one to three transmembrane domains. The large majority of the peptides enabling the identification of these transmembrane proteins are localized within soluble portions of the proteins. Nevertheless, a few analyzed peptides corresponded to significant parts of transmembrane domains.

The proteins identified by LC−MS/MS thus appear to give an overview of all types of proteins expressed in mitochondria, with little restrictions on the pI, MW, and hydrophobicity.

Biological Relevance of the Protein Identifications. Among the 179 proteins identified by LC−MS/MS, 132 are already recorded as mitochondrial in YPD, 19 as belonging to subcellular

compartments other than mitochondria, and 28 are of unknown cellular localization. The information of subcellular localization was either experimentally demonstrated or predicted by sequence analysis and comparison. Our preliminary analysis of 60 μg of a total yeast protein extract by SDS-PAGE−LC−MS/MS (37 gel slices) had led to the identification of 408 different gene products, among which 54 were known as mitochondrial. The analysis of the purified organelle thus enabled us to increase the number of mitochondrial proteins characterized by a factor of 2.5. Table 2 indicates the number of identified proteins that are involved in each mitochondrial function. About one-half of the identified proteins take part in three major mitochondrial activities: carbohydrate metabolism, protein synthesis, and respiratory functions.

Table 2. Distribution of the Identified Proteins in the Different Mitochondrial Functions

cellular function	no. identified proteins
amino acid, nitrogen metabolism	11
carbohydrate metabolism	22
carriers, transporters	11
fatty acid metabolism	1
miscellaneous	1
nucleic acid metabolism	3
organellar morphology, inheritance	2
protein sorting	8
protein translation, stability	20
respiratory chain/ oxidative phosphorylation	39
stress response	6

Regarding the 19 proteins possessing subcellular localization other than mitochondria, their presence in the mitochondrial extracts could a priori result from (i) a contamination during the preparation, (ii) the fact that other cellular fractions or organelles are in close contact with mitochondria, or (iii) a true dual localization. Two proteins are histone proteins that are highly expressed and have been identified by only one or two peptides; hence, their presence could be due to a slight contamination of the mitochondrial fraction by nuclei. Nine are subunits of cytoplasmic ribosomes, and it has been shown that a fraction of these ribosomes are associated with mitochondria.[30] Two proteins experimentally located in the endoplasmic reticulum and five in

the cytoplasm could interact with mitochondria. Finally, Ybr026cp, which was experimentally proved to be nuclear, is essential for mitochondrial function and could exhibit a dual localization, as this was already reported for Rna14p.[31]

Our LC−MS/MS analyses also identified 28 proteins of unknown cellular localization (Table 3). The first five proteins are recorded as possibly involved in mitochondrial function, whereas the others are only defined as "proteins of unknown function". Table 3 also presents a prediction of cellular localization obtained with the PsortII software (http://psort/nibb.ac.jp/). This software searches for the presence of signal sequences involved in protein sorting to calculate the probability for a protein to be localized in the various cellular compartments. According to this prediction program, 10 proteins out of 28 are expected to be located within mitochondria (probability > 50%). For the others there are no clear-cut predictions. The cellular localization of 6 of these 28 proteins was experimentally tested with GFP- or HA-tagged proteins as part of the EUROFAN project (http://mips.gsf.de/proj/yeast). These 6 proteins were then conclusively proved to be mitochondrial. Figure 2 presents an example of these localization experiments for Ygr086cp and an MS/MS spectrum that enabled the identification of this protein in our LC−MS/MS experiments. This experimental evidence also demonstrates the limits of the PsortII program, since a protein (Yol027cp) with a probability to be located in mitochondria of 17.4% was proved to be mitochondrial. In conclusion, the LC−MS/MS data are in good agreement with the tagging experiments and may, therefore, constitute a powerful tool in the assignment of proteins to the mitochondrial compartment.
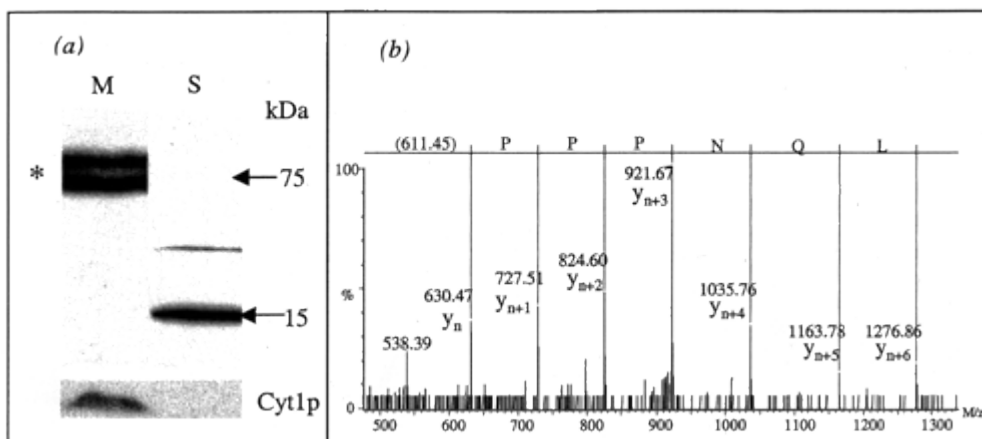

Figure 2 Demonstration of the presence in mitochondria of Ygr086cp, a protein of unknown cellular localization in YPD. (a) Visualization of the subcellular localization of the GFP-tagged Ygr086cp protein. Note that in mitochondria (M), the tagged protein is present and migrates as a doublet (*) with an apparent molecular weight of ~75 kDa. In the supernatant (S), polypeptides of lower molecular weight were detected, indicating that the tagged protein would be subject to degradation before its importation within mitochondria. Cytochrome *c*1 (Cyt1p), a subunit of the mitochondrial respiratory complex *bc*1, was used as a control for mitochondria purification. (b) Experimental MS/MS spectrum contributing to the identification of Ygr086cp obtained during the SDS-PAGE−LC−MS/MS analysis of the mitochondrial protein extract.

Table 3. Identified Proteins with No Information of Cellular Localization in YPD[a]

proteins	PsortII[b], %	tagging[c]
Possibly Mitochondrial		

Ykr016wp	**78.3**	**mito**
Ynl177cp	**52.2**	
Ylr164wp	**43.5**	
Yor356wp	**21.7**	
Yml030wp	**11.1**	
Unknown Function		
Ykr065cp	**87**	
Yhl021cp	**82.6**	**mito**
Yer182wp	**69.6**	**mito**
Yor215cp	**69.6**	
Ygr086cp	**65.2**	**mito**
Ylr201cp	**65.2**	
Yer080wp	**65.2**	
Grx5p	**60.9**	
Yil157cp	**39.1**	
Yer048w-ap	**39.1**	
Ynl100wp	**26.1**	
Yfr011cp	**26.1**	
Ymr110cp	**26.1**	

Ylr077wp	26.1	
Ypl004cp	21.7	mito
Msc1p	17.4	
Pdh1p	17.4	
Pst2p	17.4	
Yol027cp	17.4	mito
Yer004wp	17.4	
Yhr083wp	17.4	
Ybr230cp	13	
Yfl030wp	4.3	

[a] Proteins are of unknown function or possibly involved in mitochondrial function. [b] From the results of subcellular localization prediction by PsortII, only the probability percentage of a mitochondrial localization is indicated. [c] The localization experimentally obtained by tagging experiments is given for the six proteins tested in the EUROFAN project.

Conclusion

The combination of SDS-PAGE and LC−MS/MS was evaluated as an analytical approach to characterize the protein content of a total yeast mitochondrial extract. This work enabled the identification of 179 proteins, among which 28 proteins were previously of unknown localization. To identify a greater complement of the mitochondrial proteins, a larger and gradient gel could be used to obtain a better resolution of proteins in the gel slices. We also consider implementing in the analytical process an SCX separation of the tryptic peptides or an affinity purification based on ICAT, as successfully used by Wolters et al.[14] and Han et al.[13] for the analysis of complex cellular extracts. Moreover, so as to localize more precisely the proteins inside mitochondria, a further analysis would consist of the separate analysis of the membrane and soluble mitochondrial protein fractions. Preliminary experiments have been conducted on both protein fractions. The LC−MS/MS analysis of digested gel slices enabled the confirmation of effective separation of membrane-bound and soluble proteins and the identification of proteins of unknown localization. Different solvent conditions have yet to be evaluated in order to optimize the solubilization of the membrane proteins.

Finally, the small amount of material used in this study should soon allow the characterization of the mitochondrial proteome of patients with mitochondrial pathologies. Indeed, mitochondria are

required for many biological processes, and a great variety of genetic diseases are due to mitochondrial dysfunctions. It would, thus, be very important to be able to determine the effect of these pathologies on the mitochondrial proteome.

The complete data will be available at http://mitochondria.cgm.cnrs-gif.fr/

Acknowledgment

This article references 31 other publications.

1. **1**

Jung, E.; Heller, M.; Sanchez, J.-C.; Hochstrasser, D. F. *Electrophoresis* **2000**, *21*, 3369−3377.

[Crossref], [PubMed], [CAS], Google Scholar open URL

---

2. **2**

Wilkins, M. R.; Gasteiger, E.; Sanchez, J.-C.; Bairoch, A.; Hochstrasser, D. F. *Electrophoresis* **1998**, *19*, 1501−1505.

[Crossref], [PubMed], [CAS], Google Scholar open URL

---

3. **3**

Santoni, V.; Molloy, M.; Rabilloud, T. *Electrophoresis* **2000**, *21*, 1054−1070.

[Crossref], [PubMed], [CAS], Google Scholar open URL

---

4. **4**

Görg, A.; Obermaier, C.; Boguth, G.; Weiss, W. *Electrophoresis* **1999**, *20*, 712−717.

[Crossref], [PubMed], [CAS], Google Scholar open URL

---

5. **5**

Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III. *Nat. Biotechnol.***1999**, *17*, 676−682.

[Crossref], [PubMed], [CAS], Google Scholar open URL

6. **6**

Koc, E. C.; Burkhart, W.; Blackburn, K.; Moyer, M. B.; Schlatzer, D. M.; Moseley, A.; Spremulli, L. L. *J. Biol. Chem.***2001**, *276*, 43958−43969.

[Crossref], [PubMed], [CAS], Google Scholar open URL

7. **7**

Schäfer, H.; Nau, K.; Sickmann, A.; Erdmann, R.; Meyer, H. E. *Electrophoresis***2001**, *22*, 2955−2968.

[Crossref], [PubMed], [CAS], Google Scholar open URL

8. **8**

Verma, R.; Chen, S.; Feldman, R.; Schieltz, D.; Yates, J.; Dohmen, J.; Deshaies, R. *J. Mol. Biol. Cell.***2000**, *11*, 3425−3439.

[PubMed], [CAS], Google Scholar open URL

9. **9**

Bennetzen, J. L.; Hall, B. D. *J. Biol. Chem.***1982**, *257*, 3026−3031.

[PubMed], [CAS], Google Scholar open URL

10. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.***1996**, *68*, 850−858.

[ACS Full Text ◆], [CAS], Google Scholar open URL

11. **11**

Vargas, C.; McEwan, A. G.; Downie, J. A. *Anal. Biochem.***1993**, *209*, 323−326.

[Crossref], [PubMed], [CAS], Google Scholar open URL

12.**12**

Yaffe, M. *Methods Enzymol.***1991**, *194*, 627−643.

[Crossref], [PubMed], [CAS], Google Scholar open URL

13.**13**

Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. *Nature***2001**, *19*, 946−951.

[Crossref], [CAS], Google Scholar open URL

14.**14**

Wolters, D. A.; Washburn, M. P.; Yates, J. R., III. *Anal. Chem.***2001**, *73*, 5683−5690.

[ACS Full Text ◆], [CAS], Google Scholar open URL

15.**15**

Simpson, R. J.; Connolly, L. M.; Eddes, J. S.; Pereira, J. J.; Moritz, R. L.; Reid, G. E. *Electrophoresis***2000**, *21*, 1707−1732.

[Crossref], [PubMed], [CAS], Google Scholar open URL

16.**16**

Spahr, C. S.; Susin, S. A.; Bures, E. J.; Robinson, J. H.; Davis, M. T.; McGinley, M. D.; Kroemer, G.; Patterson, S. D. *Electrophoresis***2000**, *21*, 1635−1650.

[Crossref], [PubMed], [CAS], Google Scholar open URL

17.**17**

Washburn, M. P.; Wolters, D.; Yates, J. R., III. *Nat. Biotechnol.***2001**, *19*, 242−247.

[Crossref], [PubMed], [CAS], Google Scholar open URL

18. **18**

Opiteck, J. G.; Jorgenson, J. W. *Anal. Chem.***1997**, *69*, 2283−2291.

[ACS Full Text ⬥], [CAS], Google Scholar open URL

19. **19**

Rabilloud, T.; Kieffer, S.; Procaccio, V.; Louwagie, M.; Courchesne, P. L.; Patterson, S. D.; Martinez, P.; Garin, J.; Lunardi, J. *Electrophoresis***1998**, *19*, 1006−1014.

[Crossref], [PubMed], [CAS], Google Scholar open URL

20. Hanson, B. J.; Schulenberg, B.; Patton, W. F.; Capaldi, R. A. *Electrophoresis***2001**, *22*, 950−959.

[Crossref], [PubMed], [CAS], Google Scholar open URL

21. **21**

Lopez, M. F.; Kristal, B. S.; Chernokalskaya, E.; Lazarev, A.; Shestopalov, A. I.; Bogdanova, A.; Robinson, M. *Electrophoresis***2000**, *21*, 3427−3440.

[Crossref], [PubMed], [CAS], Google Scholar open URL

22. **22**

Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. *Mol. Cell Biol.***1999**, *19*, 1720−1730.

[Crossref], [PubMed], [CAS], Google Scholar open URL

23. **23**

Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. *Proc. Natl. Acad. Sci. U.S.A.***2000**, *97*, 9390−9395.

[Crossref], [PubMed], [CAS], Google Scholar open URL

24.**24**

Ferro, M.; Seigneurin-Berny, D.; Rolland, N.; Chapel, A.; Salvi, D.; Garin, J.; Joyard, J. *Electrophoresis***2000**, *21*, 3517−3526.

[Crossref], [PubMed], [CAS], Google Scholar open URL

25.**25**

Santoni, V.; Rabilloud, T.; Doumas, P.; Rouquié, D.; Mansion, M.; Kieffer, S.; Garin, J.; Rossignol, M. *Electrophoresis***1999**, *20*, 705−711.

[Crossref], [PubMed], [CAS], Google Scholar open URL

26.**26**

Herbert, B. R.; Harry, J. L.; Packer, N. H.; Gooley, A. A.; Pedersen, S. K.; Williams, K. L. *TRENDS Biotechnol.***2001**, 19, S3−9.

[CAS], Google Scholar open URL

27.**27**

Hunte, C.; Koepke, J.; Lange, C.; Rossmanith, T.; Michel, H. *Struct. Fold. Des.***2000**, *8*, 669−684.

[Crossref], [PubMed], [CAS], Google Scholar open URL

28.**28**

Tsukihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itoh, K.; Nakashima, R.; Yaono, R.; Yoshikawa, S. *Science***1996**, *272*, 1136−1144.

[Crossref], [PubMed], [CAS], Google Scholar open URL

29.**29**

Kyte, J.; Doolittle, R. F. *J. Mol. Biol.***1982**, *157*, 105−132.

[Crossref], [PubMed], [CAS], Google Scholar open URL

30. Corral-Debrinsky, M.; Blugeon, C.; Jacq, C. *Mol. Cell Biol.***2000**, *20*, 7881−7892.

[Crossref], [PubMed], Google Scholar open URL

31. **31**
Rouillard, J. M.; Brendolise, C.; Lacroute, F. *Mol. Gen. Genet.***2000**, *26*, 1103−1112.

[Crossref], Google Scholar open URL