Review TRENDS in Genetics Vol.20 No.11 November 2004

Mitochondriomics or what makes us breathe

Andreas S. Reichert and Walter Neupert

Adolf-Butenandt-Institut für Physiologische Chemie, Ludwig-Maximilians-Universität München, Butenandtstr. 5, 81377 München, Germany

Mitochondria perform several fundamental cellular processes in higher eukaryotes including oxidative phosphorylation, Fe/S cluster formation and apoptosis. Dysfunction of the organelle is associated with a wide range of human diseases. To gain a better understanding of mitochondrial function, several recent proteomic, genetic, transcriptomic and bioinformatic approaches have set out to determine the complete set of mitochondrially located proteins in yeast, plants and mammals. Here, we review these studies and discuss the advances and limitations of individual strategies. Integration of various approaches proves to be a successful and useful way to identify the mitochondrial proteome with high sensitivity and specificity. The most comprehensive dataset is available for *Saccharomyces cerevisiae*, giving an estimated number of 700 different proteins located in mitochondria.

Mitochondria perform a large number of reactions in eukaryotic cells. Probably the most impressive of these is oxidative phosphorylation, in which five major multisubunit complexes cooperate to transduce the energy of nutrient-derived substrates into the energy stored in ATP. Mitochondria provide the majority of this ATP to the rest of the cell. Other important reactions occurring in mitochondria include the Krebs cycle, heme biosynthesis, β-oxidation of fatty acids and metabolism of certain amino acids. A more recently discovered pathway is the formation and export of Fe/S clusters (Figure 1).

All the proteins catalyzing these reactions occur in an organelle that is bounded by two membranes and thereby separated, although not functionally isolated, from the rest of the cell. Interaction with the cytosol and with other organelles requires a large number of carriers, transporters and channels in the two membranes. An example is the ADP–ATP carrier, which in a human individual transports 50–60 kg of ATP per day across the inner membrane of its mitochondria.

The majority of mitochondrial proteins are nuclear encoded, synthesized in the cytosol, and are post-translationally imported into mitochondria; the mitochondrial genome itself encodes only a few proteins (e.g. eight in *Saccharomyces cerevisiae*, 13 in *Homo sapiens*, 25 in *Arabidopsis thaliana* and 64 in *Reclinomonas americana*; for more information, see http://megasun.bch.umontreal.ca/gobase/gobase.html). A considerable number of mitochondrial proteins are involved in the processes of import and assembly of these nuclear encoded proteins. A group of chaperones monitors the correct *de novo* folding of mitochondrial proteins and participates in decisions about refolding or degradation. Proteolytic removal of misfolded proteins and regulated turnover of proteins is mediated by a series of proteases associated with the various subcompartments of mitochondria. Some proteins are involved in the fusion and fission of these organelles, which in many cases form highly dynamic interconnected networks [1]. These latter proteins are often indispensable for the proper inheritance of mitochondrial DNA, and mutations in some are associated with human diseases (Box 1).

In addition to these often-complex functions, mitochondria have a key role in apoptosis. Release from mitochondria of cytochrome *c*, apoptosis-inducing factor and several other proteins promotes programmed cell death [2]. Furthermore, ageing is strongly correlated with impaired mitochondrial function; for example, the accumulation of mitochondrial DNA (mtDNA) mutations was proposed to lead to premature ageing in mice expressing an error-prone mitochondrial DNA polymerase γ [3].

To gain better insight into the functional networks of mitochondrial proteins, the complete set must be known. Several studies in recent years have addressed this aspect from different angles. Here, we present an overview of these studies and discuss the strengths and weaknesses of the various approaches. A major emphasis will be on mitochondria of *S. cerevisiae* (baker's yeast), from which most of the large-scale datasets were obtained. This organism has proven an excellent model system for mitochondriology. This is not surprising in view of the conservation of the fundamental functions of this organelle between many organisms. Recent advances in determining mammalian and plant mitochondrial proteomes are also discussed. Furthermore, we emphasize the importance of the integration of bioinformatic, genetic and biochemical approaches, as well as the subsequent detailed functional analysis of several mitochondrial molecular machines.

Proteomics

The most direct approach to obtain the complete mitochondrial proteome is to isolate highly purified

Corresponding author: Andreas S. Reichert (Andreas.Reichert@bio.med.uni-muenchen.de).

Available online 11 September 2004

 0168-9525/$ - see front matter doi:10.1016/j.tig.2004.08.012

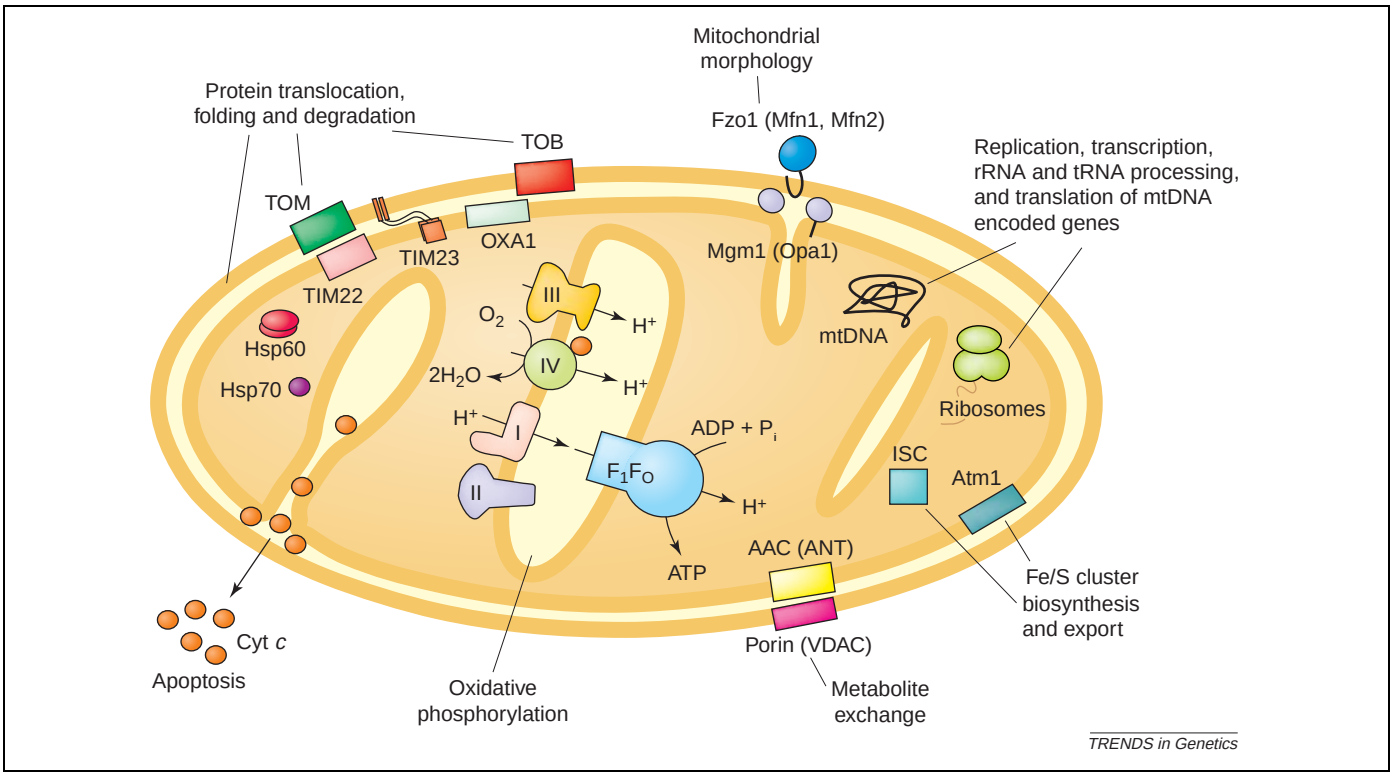

Figure 1. Overview of cellular processes related to mitochondria. Schematic cross-section of a mitochondrium showing a selection of proteins or protein complexes involved in: protein translocation [translocase complex of the outer membrane (TOM), topogenesis of mitochondrial outer membrane β-barrel proteins (TOB) TIM22, TIM23 and OXA1, the three translocase complexes of the inner membrane], protein folding (Hsp60, Hsp70) and degradation; mitochondrial morphology (Fzo1, Mfn1, Mfn2, Mgm1 and Opa1); replication, transcription, rRNA and tRNA processing, and translation (ribosomes) of genes encoded by mitochondrial DNA (mtDNA); Fe/S cluster biosynthesis [iron-sulfur cluster assembly complex (ISC)] and export (Atm1); metabolite exchange [ADP–ATP carrier (AAC) or adenine nucleotide exchanger (ANT), porin or voltage-dependent anion exchanger (VDAC)]; oxidative phosphorylation (complex I, II, III, IV, and the F_1F_O-ATP synthase); and apoptosis (cytochrome *c*). The yeast proteins are shown with their human orthologs in parenthesis. ATP synthesis in the mitochondrial matrix using the proton gradient across the inner membrane is indicated. Basic metabolic processes such as the Krebs cycle, the metabolism of certain amino acids and lipids, and the heme biosynthesis are omitted. Invaginations of the inner membrane, the cristae, are shown as they often appear in electron micrographs of sections of mitochondria. That part of the cristae where the invagination starts, the so-called cristae junction, is only visible when it is in the same plane as the section plane. If this is not the case, it leads to an isolated appearance of the cristae membrane although continuity over the whole inner membrane is maintained.

mitochondria and determine the subset of mitochondrial proteins by mass spectrometry. Several groups have purified yeast, mouse and human mitochondria by classical differential centrifugation, gradient ultracentrifugation or free-flow electrophoresis (FFE) [4–9]. Similar studies on the mitochondrial proteome of plants have been undertaken [10–15]. In general, mitochondria were solubilized and the proteins were separated in various ways. Isoelectric focusing and subsequent SDS–polyacrylamide gel electrophoresis (2D PAGE) is a method that has major advantages as it offers the possibility of, for instance, comparing a large set of proteins under the influence of different growth conditions or in response to drug administration. This allows for fast identification of the proteins whose levels are affected.

Ohlmeier and colleagues [5] compared the proteome of baker's yeast under fermentative versus respiratory growth. The overall differences were rather minor, as only 18 of the 252 identified proteins changed significantly in their abundance. However, in this and other studies using 2D PAGE, a major drawback of this method became obvious. The set of proteins identified is largely biased towards abundant and easy-to-solubilize proteins [16].

One possible way to overcome this problem was exemplified in a thorough and extensive study by Sickmann *et al.* [8]. Highly purified mitochondria were prepared from baker's yeast and different proteomic approaches were applied. In addition to 2D PAGE, the approaches included digestion with different proteases and subsequent multidimensional liquid chromatography (MDLC) or electrospray ionization mass spectrometry (ESI-MS), and one dimensional SDS-polyacrylamide gel electrophoresis (1D SDS–PAGE) with subsequent nano-scale liquid chromatography tandem mass spectrometry (n-LC-MS/MS). In total, >20 million spectra were recorded, yielding a list of 749 non-redundant mitochondrial proteins. In this approach, virtually saturation was achieved as, in those spectra analyzed at the end, only very few new proteins were detected compared with those spectra analyzed in the beginning [8]. About 92% of already known mitochondrial proteins contained in the MITOP database [17] were covered by these predicted mitochondrial proteins. Because the latter reference itself is biased towards more-abundant and soluble proteins and certain functional classes such as respiratory chain complexes, this sensitivity of 92% might overestimate the true coverage of the mitochondrial proteome by this study (Box 2). Therefore, we chose to use what in our opinion appears as the best currently available reference set, the one provided by the MITOP2 database. It represents an up-to-date list of proteins that were unambiguously localized to mitochondria in experiments

Box 1. Mitochondrial dysfunction in human diseases

Mutations in mitochondrial DNA (mtDNA) affecting essential subunits of the respiratory chain or mitochondrial tRNA genes are associated with a large number of mitochondrial diseases such as mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes (MELAS), myclonic epilepsy and ragged red fiber disease (MERRF), cardiomyopathy, Leber's hereditary opticus neuropathy (LHON), Leigh's syndrome and many others [40]. In addition, mtDNA alterations have been observed in early tumors of the liver, prostate, bladder, and head and neck, as well as in astrocytomas, adenocarcinomas and primary lung tumors and in Barrett's oesophagus [41]. For further reading, see Ref. [40] or http://www.mitomap.org (Table 2).

Mutations in nuclear genes encoding mitochondrial proteins can also be the cause of human diseases. For example,

(i) Friedreich's ataxia is caused by mutations in frataxin, a protein involved in the biogenesis of iron-sulfur protein [42,43].

(ii) Deafness dystonia peptide 1 (DDP1), a protein involved in the import of proteins into mitochondria, was found to be associated with Mohr–Tranebjaerg syndrome [44–47].

(iii) Opticus atrophy type I is inherited in an autosomal-dominant fashion in which the gene coding for the mitochondrial dynamin-like protein Opa1 is affected [48,49]. Opa1 is involved in maintaining mitochondrial morphology and its yeast ortholog, Mgm1, was shown to be essential for mitochondrial fusion [50,51].

(iv) Recently, another component involved in mitochondrial fusion, Mitofusin-2, was shown to be mutated in families with Charcot-Marie-Tooth neuropathy type 2A [52].

In addition to these well-documented diseases, several other common human diseases are accompanied by mitochondrial defects. Impairment of glucose and fatty acid metabolism due to mitochondrial dysfunction are important factors in diabetes and obesity [53]. Moreover, a major contributing factor to Parkinson's and Alzheimer's disease is reduced cellular ATP levels caused by mitochondrial dysfunction [54,55], although how mitochondrial dysfunction is connected to these diseases is far from understood.

Box 2. Terms and definitions

Sensitivity

The sensitivity of an approach to determine the mitochondrial proteome is a measure of the probability to detect a mitochondrial protein by this approach but does not say whether any candidate protein is truly a mitochondrial protein. It is a measure of the coverage of an approach as it gives the percentage of known mitochondrial proteins that were identified by this approach. The absolute value of the sensitivity depends on the reference set used and should therefore only be taken as a rough estimate. However, when the same reference set is always used, it is possible to compare different approaches with each other (Table 1).

Specificity

The specificity of an approach to determine the mitochondrial proteome is a measure of the accuracy of a given approach. It can be estimated from the percentage of the predicted mitochondrial proteins of this approach that are present in the reference set. This value is also dependent on the quality of the reference set and only gives a rough estimate of the accuracy of an approach. For a given reference set, different approaches can be compared with each other.

Reference set

The reference set for *Saccharomyces cerevisiae* as given by the curators of the MITOP2 database is based on single gene studies only and does not contain any information from any of the approaches compared in Table 1 [7]. In this way, no bias towards either of the discussed high-throughput approaches is generated, allowing a better comparison between them. Owing to the incompleteness of the reference set itself, the calculated specificities and sensitivities are best estimates.

Sensitivity versus specificity

An approach that is only able to predict a limited number of proteins will therefore always have a low sensitivity (given that the reference set is much larger). However, when many of these few predicted proteins are present in the reference set (meaning that there are presumably few false positives), the specificity is high. Conversely, an approach that predicts almost all known proteins in a cell to be localized in the mitochondria will have a high sensitivity, ut most likely also has a low specificity owing to the large number of false positives.

with individual proteins (http://ihg.gsf.de/mitop2; [7,17]). This reference set contains 477 proteins of which 76% were identified by Sickmann and colleagues [8] (Table 1). This is the highest sensitivity (Box 2) so far reported for a study to evaluate the complete mitochondrial proteome of a single organism (Table 1). By using a combination of proteomic approaches, Sickmann and colleagues [8] minimized the biases and problems associated with the various individual methods. For example, 2D PAGE analysis tends to underrepresent hydrophobic proteins such as peripheral or integral membrane proteins [16]. This was, at least partially, overcome by the combination of 1D SDS–PAGE and n-LC-MS/MS. Another problem often encountered is that certain proteins (e.g. small proteins or proteins with an unusual amino acid composition) do not yield peptide fragments of suitable size due to lack of properly spaced cleavage sites. Therefore, Sickmann *et al.* used four different proteases. In summary, these various approaches finally led to the high specificity (48.5%) of this study (Box 2).

In a recent extensive proteomic study of mitochondria from *S. cerevisiae*, cells were grown on fermentable or non-fermentable carbon sources and mitochondria were purified either by classical gradient centrifugation or by free-flow electrophoresis (FFE) [7]. Subsequently, matrix and membrane-associated fractions were generated and analyzed separately by LC-MS/MS or liquid chromatography fourier transform-ion cyclotron resonance mass spectrometry (LC-FTICR-MS). Thereby, 28 datasets were generated, which led to the identification of 546 non-redundant proteins. Surprisingly, this study revealed numerous (209) mitochondrial candidates that where not discovered in the study by Sickmann *et al.* [8] discussed above. Conversely, 412 proteins predicted by the latter study were not detected here. The authors suggest that many, but not all, of those proteins detected by only one of the two approaches are non-mitochondrial contaminants (for further reading, we recommend Ref. [7]). This might be explained by the differences in growth conditions and in purification methods. Nevertheless, 337 proteins were detected by both proteomic studies, representing 62% of all proteins detected by Prokisch *et al.* [7]. When integrating both studies, the sensitivity decreases to 49.9% and the specificity increases to 63.1% (Box 2; Table 1).

Initial studies on the proteome of plant mitochondria from *A. thaliana* based on 2D PAGE and mass spectrometry led to the identification of nearly 100 non-redundant proteins [14,15]. The same limitations for low abundance and hydrophobic proteins as discussed above could be resolved in a recent study by Heazlewood *et al.* [11].

Table 1. Quantitative comparison of various approaches to determine the mitochondrial proteome in yeast, humans and mice[a]

Species	Approach	Predicted number of mitochondrial proteins[b]	Sensitivity (%)[b]	Specificity (%)[b]
Saccharomyces cerevisiae				
Proteomics	Highly purified mitochondria analyzed by 2D PAGE, 1D SDS–PAGE and n-LC-MS/MS [8]	749	76.1	48.5
	Highly purified (by density gradient and free-flow electrophoresis) mitochondria analyzed by LC-MS/MS and FTICR-MS [7]	546	53.7	46.9
	Proteins found in both studies [7,8]	377	49.9	63.1
	Highly purified mitochondria analyzed by SDS–PAGE and LC-MS/MS [6]	177	28.3	76.3
	2D PAGE of highly purified mitochondria (fermentative versus non-fermentative growth) [5]	252	30.4	57.5
Genetics	Deletion phenotype screen for respiratory-deficient strains (*petites*) [22]	381	39.2	49.1
	Deletion phenotype screen for slow growth on non-fermentable carbon source [23]	466	47.6	48.7
	Immunofluorescence analysis of epitope-tagged proteins [20]	364	46.5	61.0
	Fluorescence analysis of C-terminally GFP-tagged proteins [19]	527	69.2	62.6
Bioinformatics	*Rickettsia* homolog (MITOP2 database [7,38])	931	42.8	21.9
	Prediction of N-terminal targeting signal with MITOPROT score >0.8 [30]	789	54.7	33.1
	Prediction of N-terminal targeting signal with MITOPROT score >0.9 [30]	574	46.3	38.5
	Prediction of N-terminal targeting signal by PSORT [29]	981	52.8	25.7
	Prediction of N-terminal targeting signal by Predotar [31]	397	43.4	52.1
	Prediction of subcellular localization by Bayesian algorithm [39]	500	43.4	41.4
Transcriptomics	Synexpression after Hap4 overexpression [28]	514	45.1	41.8
	Synexpression after diauxic shift [25]	416	15.9	18.3
Integration of proteomics, transcriptomics, genetics and bioinformatics	Integrative analysis of 22 datasets and a MITOP2 score >95 [7]	557	81.1	69.5
	High-confidence protein–protein interaction dataset [35] combined with known mitochondrial proteins [7,38]	154	17.4	53.9
	Medium-confidence protein–protein interaction dataset [35] combined with known mitochondrial proteins [7,38]	815	49.7	29.1
Homo sapiens* and *Mus musculus				
Proteomics	Highly purified mitochondria from human heart analyzed by 1D SDS-PAGE and LC-MS/MS [9]	544	32.9	39.7
	Highly purified mitochondria from mouse heart analyzed by LC-MS/MS [4]	181	32.5	71.9
	Highly purified mitochondria from mouse brain analyzed by LC-MS/MS [4]	190	31.5	67.9
	Highly purified mitochondria from mouse liver analyzed by LC-MS/MS [4]	279	42.8	62.7
	Highly purified mitochondria from mouse kidney analyzed by LC-MS/MS [4]	265	37.9	58.5
	Highly purified mitochondria from mouse of all four tissues (heart, brain, liver and kidney) analyzed by LC-MS/MS [4]	409	53.1	53.1

[a]Abbreviations: LC-FTICR-MS, liquid chromatography fourier transform-ion cyclotron resonance mass spectrometry; n-LC-MS/MS, nanoscale liquid chromatography tandem mass spectrometry; 1D SDS–PAGE, one dimensional SDS-polyacrylamide gel electrophoresis.
[b]Sensitivity is the fraction of the reference set covered by the predicted mitochondrial dataset; specificity is the fraction of predicted mitochondrial proteins confirmed by the reference set (Box 2). Predicted mitochondrial datasets of the studies in *Saccharomyces cerevisiae* and *Homo sapiens,* and the mitochondrial reference sets containing 477 and 656 proteins, respectively, were taken from the MITOP2 database (http://ihg.gsf.de/mitop2; [38]). For analysis of the mouse proteomic study, we used the raw data from Supplementary Table S1 in Ref. [4].

Highly purified mitochondria from *A. thaliana* using two successive Percoll gradients where analyzed by LC-MS/MS. Here, 416 putative mitochondrial proteins were detected. A considerable amount of proteins involved in cellular signaling such as a Ser/Thr kinase, a mitogen-activated protein kinase, a 2A protein phosphatase regulatory subunit, a GTP-binding protein and several putative Ca^{2+}-binding proteins were detected. Interestingly, several candidates for mitochondrially localized proteins in plants with unknown function seem to be absent from humans, yeast and also from *Rickettsia prowazekii*, which is the closest living relative to the bacterial endosymbiont [11]. This points to the unique functions of some mitochondrial proteins in plants.

The human mitochondrial proteome was extensively studied using highly purified heart mitochondria [9]. Mitochondria were solubilized with a non-ionic detergent and the resulting protein complexes were separated by sucrose gradient centrifugation and subsequent 1D SDS–PAGE. Mass spectrometric analysis revealed >540 putative mitochondrial proteins (Table 1). Similar to plants, orthologous kinases were also detected that are absent in yeast mitochondria, which is consistent with the presence of advanced regulatory circuits in multicellular organisms.

Advances and perspectives in proteomic analyses of mitochondria

The number of false positives in proteomic studies obviously depends on the purity of isolated mitochondria. One recent technical development to improve this is the use of FFE to purify mitochondria [7,18]. Although solubilization and detection procedures will certainly be improved in future studies, any proteomic approach is likely to remain biased against proteins of low abundance (compare Figure 2a in Ref. [7]) and also proteins that are difficult to extract from biological samples. Interestingly, there was no significant bias against the identification of membrane-associated proteins in the study by Prokisch and colleagues [7]. This was probably due to specific

technological advances in purification (i.e. FFE) and mass spectrometry (LC-MS/MS and LC-FTICR-MS). However, in another study, bias against the detection of membrane-associated proteins was observed [4]. The authors of this latter study attributed this to the lower protein abundance on average of membrane proteins compared with soluble proteins and not to the limited sensitivity of mass spectrometry for hydrophobic proteins. Nevertheless, hydrophobic proteins still appear to be an issue for approaches based on 2D PAGE [16]. Furthermore, many proteins are expressed in a conditional manner (for instance under stress) or at specific stages of growth or development of an organism. Therefore, it will be difficult to identify such potentially interesting proteins by a standard proteomic approach.

Genetics, bioinformatic and transcriptomic approaches complete the picture

Genetics

Two recent studies report on tagging a large number of yeast genes and localizing the expressed fusion proteins by fluorescence microscopy [19,20]. Tagging was carried out in a systematic manner by fusion of the green fluorescent protein (GFP) to the C-terminus [19], by random transposon-mediated mutagenesis, or by cloning of PCR-amplified open reading frames (ORFs) into an overexpression/tagging vector [20]. In this way, many essential and non-essential proteins could be assigned not only to mitochondria but also to other cellular compartments. As particularly shown in the study by Huh and colleagues [19], this gives high sensitivity and specificity when tested against the MITOP2 reference set (Box 2; Table 1). In a further tagging approach, the N-terminal half of a split GFP was randomly fused to sequences from a cDNA library and coexpressed with a matrix-targeted reporter containing the C-terminal half of GFP [21]. When the expressed fusion protein contained targeting information for the mitochondrial matrix, a full-length GFP was generated by protein splicing. Automated cell sorting enriched fluorescent clones and subsequent identification of expressed cDNA fusion proteins became possible. It will be interesting to see what novel mitochondrial proteins will be identified by this approach. In its present form, it is limited to the detection of proteins exposing the C-terminus to the matrix of mitochondria.

All tagging approaches require that expression levels are sufficiently high and that the tags do not interfere with correct targeting to the respective location. Furthermore, changing the authentic amino acid sequence of a protein (by insertion of a tag), as well as overexpression, can lead to mislocalization [8]. Additional independent experimental evidence for a reliable localization of a protein is desirable; for instance, the similarity of the phenotype of a strain in which the candidate gene was deleted to the phenotype of a strain deficient in a known mitochondrial protein. Respiratory deficiency and slow growth on non-fermentable carbon sources are two such phenotypes exploited in large-scale studies [22,23]. These methods are restricted to non-essential genes and clearly do not prove that a protein is a mitochondrial protein as, for example, nuclear transcription factors or cytosolic regulators also affect mitochondrial function. Synthetic lethality caused by deletion of a candidate gene together with deletion of a gene encoding a known mitochondrial protein also indicates a functional and, therefore, often a physical interaction. Although a large-scale study on synthetic lethality in baker's yeast was performed [24], the results to date have not been analyzed for predicting mitochondrial proteins.

Transcriptomics

Another approach to detect mitochondrial proteins is to study the expression level of mRNAs under various growth conditions. Similar responses in transcript levels of different genes allow their clustering into groups. These groups are more likely to represent functionally and physically interacting protein networks than transcripts that are differentially regulated. Several synexpression studies were reported, including genes involved in mitochondrial functions [25–28]. Two of those are described in Table 1. These studies can predict the subcellular localization of proteins, albeit only with low-to-medium specificity.

Bioinformatics

Another possibility to identify proteins located in mitochondria is to use *in silico* predictions. Some predictions rely on the fact that a large number of proteins targeted to the matrix or the inner membrane contain a positively charged, amphipathic N-terminal signal sequence [29–31]. Several algorithms exist that make possible fast and initial testing for the presence of such a sequence (Tables 1 and 2). By contrast, proteins targeted to the endoplasmic reticulum or highly positively charged proteins such as ribosomal subunits also often give high scores, explaining the low-to-medium specificity of these approaches. Mitochondrial proteins containing targeting signals with other characteristics, in particular internal signals, are being missed. This group includes inner membrane proteins of the carrier family (a class of proteins containing six transmembrane helices and involved in the transport of small solutes such as ADP or ATP across the inner membrane), many intermembrane space proteins and presumably all outer membrane proteins. A more comprehensive bioinformatic approach has integrated information derived from genetic, *in silico* predictions and correlated mRNA expression datasets using a Bayesian probabilistic algorithm [32] (Tables 1 and 2). An advantage of this type of analysis lies in the possibility of easily including other types of information such as proteomic data.

Evolutionary considerations on the origin of the mitochondrial proteome

Mitochondria are widely believed to have originated from a single endosymbiotic event in which an amitochondriate pro-eukaryotic host took up an α-proteobacterial endosymbiont [33,34]. Although most of the mitochondrial genes have been transferred during evolution to the nucleus, the mitochondrial location and function of the proteins encoded by these genes often appear to be maintained. Therefore, a candidate protein

Table 2. Databases and useful tools related to mitochondrial proteomes

Databases	Description	Website
MITOP2 database [7,38]	Database containing a large number of datasets from various studies related to mitochondria in *Saccharomyces cerevisiae*, *Neurospora crassa* and *Homo sapiens*	http://ihg.gsf.de/mitop2
MitoProteome [9]	Database on the human mitochondrial proteome	http://www.mitoproteome.org
TRIPLES database [20]	Database for the localization of proteins in baker's yeast by immunofluorescence microscopy after epitope tagging	http://ygac.med.yale.edu/triples
Yeast deletion project and proteomics of mitochondria database (YDPM) [7,23]	Database containing information on large-scale proteomic and expression studies on baker's yeast	http://www-deletion.stanford.edu/YDPM/YDPM_index.html
Yeast GFP fusion localization database [19]	Database for the localization of proteins in baker's yeast by immunofluorescence microscopy of GFP fusion proteins	http://yeastgfp.ucsf.edu
GOBASE database [56]	Compilation of mitochondrial genomes	http://megasun.bch.umontreal.ca/gobase/gobase.html
MITOMAP database [57]	Database on human mitochondrial DNA sequences including known polymorphisms and mutations	http://www.mitomap.org
Bioinformatic prediction of mitochondrial proteins		
Mitoprot [30]	N-terminal targeting signal analysis	http://ihg.gsf.de/ihg/mitoprot.html
Predotar [31]	N-terminal targeting signal analysis	http://www.inra.fr/predotar/index.html
PSORT and other versions [29]	N-terminal targeting signal analysis for bacterial, plant and other eukaryotic sequences	http://www.psort.org
Bayesian [39]	Prediction of subcellular localization based on diverse range of 30 features	http://bioinfo.mbb.yale.edu/ genome/localize
Mitopred [58]	Prediction based on occurrence patterns of protein domains, amino acid composition and pI values	http://mitopred.sdsc.edu

that shows homology to a protein of the supposedly very close-living relative of the α-proteobacterial ancestor of mitochondria, *R. prowazekii*, is thought to be more likely to be targeted to mitochondria than a protein that has no homologs in this organism. This does not apply to all mitochondrial proteins because some are clearly encoded by nuclear genes that were already present in the ancestral eukaryotic host. In addition, some gene products of eubacterial origin have been recruited to cellular compartments other than mitochondria during evolution [33]. Therefore, this kind of analysis can, at best, only be taken as an indication for a mitochondrial location of the corresponding protein.

Integrative approaches

Protein–protein interaction networks

A new possibility of deciphering the proteome of mitochondria is to study protein–protein interactions. The rationale is that any interaction of an unidentified protein with a known mitochondrial protein reveals its mitochondrial location. In addition to the identity of a candidate protein, one can learn something about its physiological function, given that there is information about the interacting partner protein(s). Several large-scale studies were performed to determine networks of protein–protein interactions in yeast. They are based on high-throughput yeast two-hybrid assays, mass spectrometric identification of proteins co-purified upon isolation of protein complexes, correlation of mRNA expression, synthetic lethality of two genes, and *in silico* predictions through genome analysis. These approaches have been analyzed and reviewed by von Mering *et al.* [35]. When using affinity purification of protein complexes, only rather stable interactions are detected as well as interactions that are not disturbed by the affinity tag. Two-hybrid assays are able to find more-labile interactions but are particularly prone to false positives (see Ref. [35] for further discussion). When applied on their own, false positives are a common drawback of high-throughput methods. Therefore, the data derived from these various whole-genome studies were integrated and protein–protein interactions were grouped in high and medium confidence classes [35]. To assign novel proteins to mitochondria by non-proteomic procedures, the curators of the MITOP2 database (http://ihg.gsf.de/mitop2; [7,17]) used these confidence classes and looked for the presence of at least one known mitochondrial protein in these protein–protein interactions (Tables 1 and 2). In doing so, the mitochondrial location of 154 and 815 proteins, respectively, was predicted. The sensitivity and specificity (Box 2) are affected differently depending on the class. When the high-confidence protein–protein interaction class was analyzed, low sensitivity but high specificity was observed. Conversely, when using the medium-confidence protein–protein interaction class, a high sensitivity but a low specificity was observed (Table 1). This nicely exemplifies the trade-off between sensitivity and specificity.

Integration of transcriptomics and proteomics applied to animal mitochondria

A rather sophisticated, integrative type of analysis was reported recently by Mootha and colleagues [4]. It shows

in an exemplary manner how far one can get with a combination of tissue-based proteomics and a large-scale mRNA expression analysis. Proteins of highly purified mitochondria obtained from liver, brain, heart and kidney were separated by gel filtration and individual fractions were subjected to digestion and LC-MS/MS. In this way, 409 non-redundant proteins present in at least one tissue were obtained. By combining these 409 proteins with a reference set of known mitochondrial proteins and subtracting ten obvious contaminants, they obtained a 'mitochondria-associated' set of 591 proteins that were used for their subsequent analysis. Among the 399 (409 minus ten) proteins detected, 163 represented newly identified proteins.

We analyzed the data reported by the authors [4] and determined sensitivity and specificity for individual tissues. Out of the reference set of 409 mitochondrial proteins, 217 proteins (53.1% sensitivity) were identified in at least one tissue (Table 1). For heart, brain, liver and kidney, the sensitivity was 32.5%, 31.5%, 42.8% and 37.9%, respectively (Table 1). Furthermore, of the proteins that were identified and that were also in the reference set of known mitochondrial proteins, only ~40% were present in all four tissues [4]. This low percentage does not seem to be due to methodological limitations such as undersampling but rather can be attributed to, as the authors suggest, tissue-specific differences. This interpretation that now integrates transcriptomic and proteomic data is based on the following arguments. The authors tested whether a protein identified in one tissue behaves in a manner that is concordant or not concordant with its corresponding mRNA. For example, a protein that is identified in heart but not in liver mitochondria is considered to behave in a concordant manner when its mRNA expression level is higher in heart than in liver. Indeed, there was a strong concordance between protein identification and mRNA abundance. Furthermore, the authors used five tissue samples, two from liver and one each from heart, kidney and brain. Next they determined the probability that a protein detected in the first liver sample is detected in the second liver sample and in the samples from other tissues. For the second liver sample, this probability was 92%, but 79% on average for the other tissues. The authors propose a model in which mitochondria harbor a subset of ~50% ubiquitous and 50% tissue-specific proteins, of which half are shared between two given tissues. Two different tissues, therefore, should share ~75% of mitochondrial proteins that, at least for liver, is close to the observed 79%.

In addition, the authors determined genes that are transcriptionally regulated in a similar manner as those genes encoding mitochondrial proteins identified by proteomic means. These co-regulated genes extended the set of 'mitochondria-associated' proteins to a so-called 'mitochondrial neighborhood' comprising 643 genes altogether. This work nicely demonstrates the power of integrative approaches in which proteomic and mRNA tissue-specific data complement each other.

The most exhaustive integration of proteomic, transcriptomic, bioinformatic and genetic data was reported recently in an analysis of the mitochondrial proteome of *S. cerevisiae* [7]. Altogether 22 approaches were integrated, including a novel proteomic study and many of those discussed in detail above. Overlaps between different approaches were analyzed. For each protein, a value termed the MITOP2 score was calculated – this represents the specificity of the best combination of methods (http://ihg.gsf.de/mitop2). The MITOP2 score has a range 0–100, indicating a low-to-high probability of a protein being located in mitochondria. The power of this type of integration is impressively exemplified when the MITOP2 score is used as a criterion to predict mitochondrial proteins. The authors calculated the specificity and sensitivity of the different protein sets when a MITOP2 score >90, 93, 95, 96 or 97 was used. This was superior in terms of high sensitivity, still maintaining a high specificity when compared with any single approach applied so far (compare with Figure 4 in [7]). This is also evident from our comparison in Table 1, where using a MITOP2 score >95 yields a specificity of 69.5% and a sensitivity of 81.1%.

Outlook

The goal of determining the complete set of mitochondrial proteins is getting closer. But how far are we now? On the basis of the most exhaustive study on the mitochondrial proteome of *S. cerevisae* so far, ~700 proteins are predicted to be located in this organelle [7]. This is a conservative estimate that yielded a relatively low number of mitochondrial proteins as compared with earlier estimates of 800 [20], and compared with corresponding estimates for *A. thaliana* (2800) or humans (1500–4200 [36,37]). Mammalian mitochondria are involved in several processes that do not occur in yeast mitochondria (e.g. nonshivering thermogenesis, steroid biosynthesis and apoptosis). Tissue-specific and developmentally regulated proteins or isoforms are likely to increase further the number of mitochondrial proteins in higher organisms. Therefore, although not formally proven, it seems likely that the set of proteins present in the mitochondria of mammals and plants is larger than that of fungi.

However, many low-abundance and/or non-extractable proteins, particularly in mammalian cells, might have resisted identification so far. Therefore, at present, a reliable estimate of the total number of mitochondrial proteins in a given organism is not possible. Further technological improvements are needed that can be applied in integrative types of analysis. Tools have to be developed that allow identification of those 'hard-to-get' proteins. A future challenge will be the assignment of all mitochondrial proteins to their respective submitochondrial location such as the outer membrane, the inner membrane, the matrix and the intermembrane space. Finally, it should be emphasized that the detailed functional analysis of each individual protein is indispensable to achieve a complete understanding of 'what makes us breathe'.

References

1 Shaw, J.M. and Nunnari, J. (2002) Mitochondrial dynamics and division in budding yeast. *Trends Cell Biol.* 12, 178–184

2 Jiang, X. and Wang, X. (2004) Cytochrome *c*-mediated apoptosis. *Annu. Rev. Biochem.* 73, 87–106
3 Trifunovic, A. *et al*. (2004) Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* 429, 417–423
4 Mootha, V.K. *et al*. (2003) Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* 115, 629–640
5 Ohlmeier, S. *et al*. (2004) The yeast mitochondrial proteome, a study of fermentative and respiratory growth. *J. Biol. Chem.* 279, 3956–3979
6 Pflieger, D. *et al*. (2002) Systematic identification of mitochondrial proteins by LC-MS/MS. *Anal. Chem.* 74, 2400–2406
7 Prokisch, H. *et al*. (2004) Integrative analysis of the mitochondrial proteome in yeast. *PloS. Biol.* 2, 795–804
8 Sickmann, A. *et al*. (2003) The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13207–13212
9 Taylor, S.W. *et al*. (2003) Characterization of the human heart mitochondrial proteome. *Nat. Biotechnol.* 21, 281–286
10 Millar, A.H. and Heazlewood, J.L. (2003) Genomic and proteomic analysis of mitochondrial carrier proteins in *Arabidopsis*. *Plant Physiol.* 131, 443–453
11 Heazlewood, J.L. *et al*. (2004) Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell* 16, 241–256
12 Heazlewood, J.L. *et al*. (2003) Towards an analysis of the rice mitochondrial proteome. *Plant Physiol.* 132, 230–242
13 Bardel, J. *et al*. (2002) A survey of the plant mitochondrial proteome in relation to development. *Proteomics* 2, 880–898
14 Kruft, V. *et al*. (2001) Proteomic approach to identify novel mitochondrial proteins in *Arabidopsis*. *Plant Physiol.* 127, 1694–1710
15 Millar, A.H. *et al*. (2001) Analysis of the *Arabidopsis* mitochondrial proteome. *Plant Physiol.* 127, 1711–1727
16 Taylor, S.W. *et al*. (2002) An alternative strategy to determine the mitochondrial proteome using sucrose gradient fractionation and 1D PAGE on highly purified human heart mitochondria. *J. Proteome Res.* 1, 451–458
17 Scharfe, C. *et al*. (2000) MITOP, the mitochondrial proteome database: 2000 update. *Nucleic Acids Res.* 28, 155–158
18 Zischka, H. *et al*. (2003) Improved proteome analysis of *Saccharomyces cerevisiae* mitochondria by free-flow electrophoresis. *Proteomics* 3, 906–916
19 Huh, W.K. *et al*. (2003) Global analysis of protein localization in budding yeast. *Nature* 425, 686–691
20 Kumar, A. *et al*. (2002) Subcellular localization of the yeast proteome. *Genes Dev.* 16, 707–719
21 Ozawa, T. *et al*. (2003) A genetic approach to identifying mitochondrial proteins. *Nat. Biotechnol.* 21, 287–293
22 Dimmer, K.S. *et al*. (2002) Genetic basis of mitochondrial function and morphology in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 13, 847–853
23 Steinmetz, L.M. *et al*. (2002) Systematic screen for human disease genes in yeast. *Nat. Genet.* 31, 400–404
24 Tong, A.H. *et al*. (2004) Global mapping of the yeast genetic interaction network. *Science* 303, 808–813
25 DeRisi, J.L. *et al*. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686
26 Causton, H.C. *et al*. (2001) Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* 12, 323–337
27 Hughes, T.R. *et al*. (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126
28 Lascaris, R. *et al*. (2003) Hap4p overexpression in glucose-grown *Saccharomyces cerevisiae* induces cells to enter a novel metabolic state. *Genome Biol.* 4, R3
29 Nakai, K. and Horton, P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 24, 34–36
30 Claros, M.G. and Vincens, P. (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* 241, 779–786
31 Small, I. *et al*. (2004) Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 4, 1581–1590
32 Drawid, A. *et al*. (2000) Genome-wide analysis relating expression level with protein subcellular localization. *Trends Genet.* 16, 426–430
33 Andersson, S.G. *et al*. (2003) On the origin of mitochondria: a genomics perspective. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 165–177
34 Burger, G. *et al*. (2003) Mitochondrial genomes: anything goes. *Trends Genet.* 19, 709–716
35 von Mering, C. *et al*. (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* 417, 399–403
36 Taylor, S.W. *et al*. (2003) Global organellar proteomics. *Trends Biotechnol.* 21, 82–88
37 Richly, E. *et al*. (2003) Evolutionary diversification of mitochondrial proteomes: implications for human disease. *Trends Genet.* 19, 356–362
38 Andreoli, C. *et al*. (2004) MitoP2, an integrated database on mitochondrial proteins in yeast and man. *Nucleic Acids Res.* 1, 32
39 Drawid, A. and Gerstein, M. (2000) A Bayesian system integrating expression data with sequence patterns for localizing proteins: comprehensive application to the yeast genome. *J. Mol. Biol.* 301, 1059–1075
40 Wallace, D.C. (1999) Mitochondrial diseases in man and mouse. *Science* 283, 1482–1488
41 Verma, M. *et al*. (2003) Proteomic analysis of cancer-cell mitochondria. *Nat. Rev. Cancer* 3, 789–795
42 Koutnikova, H. *et al*. (1997) Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin. *Nat. Genet.* 16, 345–351
43 Wilson, R.B. and Roof, D.M. (1997) Respiratory deficiency due to loss of mitochondrial DNA in yeast lacking the frataxin homologue. *Nat. Genet.* 16, 352–357
44 Jin, H. *et al*. (1996) A novel X-linked gene, DDP, shows mutations in families with deafness (DFN-1), dystonia, mental deficiency and blindness. *Nat. Genet.* 14, 177–180
45 Roesch, K. *et al*. (2002) Human deafness dystonia syndrome is caused by a defect in assembly of the DDP1/TIMM8a–TIMM13 complex. *Hum. Mol. Genet.* 11, 477–486
46 Rothbauer, U. *et al*. (2001) Role of the deafness dystonia peptide 1 (DDP1) in import of human Tim23 into the inner membrane of mitochondria. *J. Biol. Chem.* 276, 37327–37334
47 Bauer, M.F. *et al*. (1999) The mitochondrial TIM22 preprotein translocase is highly conserved throughout the eukaryotic kingdom. *FEBS Lett.* 464, 41–47
48 Alexander, C. *et al*. (2000) OPA1, encoding a dynamin-related GTPase, is mutated in autosomal dominant optic atrophy linked to chromosome 3q28. *Nat. Genet.* 26, 211–215
49 Delettre, C. *et al*. (2000) Nuclear gene OPA1, encoding a mitochondrial dynamin-related protein, is mutated in dominant optic atrophy. *Nat. Genet.* 26, 207–210
50 Wong, E.D. *et al*. (2003) The intramitochondrial dynamin-related GTPase, Mgm1p, is a component of a protein complex that mediates mitochondrial fusion. *J. Cell Biol.* 160, 303–311
51 Sesaki, H. *et al*. (2003) Mgm1p, a dynamin-related GTPase, is essential for fusion of the mitochondrial outer membrane. *Mol. Biol. Cell* 14, 2342–2356
52 Zuchner, S. *et al*. (2004) Mutations in the mitochondrial GTPase mitofusin 2 cause Charcot-Marie-Tooth neuropathy type 2A. *Nat. Genet.* 36, 449–451
53 Gerbitz, K.D. *et al*. (1996) Mitochondria and diabetes. Genetic, biochemical, and clinical implications of the cellular energy circuit. *Diabetes* 45, 113–126
54 Orth, M. and Schapira, A.H. (2002) Mitochondrial involvement in Parkinson's disease. *Neurochem. Int.* 40, 533–541
55 Hashimoto, M. *et al*. (2003) Role of protein aggregation in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's diseases. *Neuromolecular Med.* 4, 21–36
56 O'Brien, E.A. *et al*. (2003) GOBASE – a database of mitochondrial and chloroplast information. *Nucleic Acids Res.* 31, 176–178
57 Kogelnik, A.M. *et al*. (1998) MITOMAP: a human mitochondrial genome database – 1998 update. *Nucleic Acids Res.* 26, 112–115
58 Guda, C. *et al*. (2004) MITOPRED: a web server for the prediction of mitochondrial proteins. *Nucleic Acids Res.* 32, W372–W374