Global organellar proteomics

Steven W. Taylor, Eoin Fahy and Soumitra S. Ghosh

MitoKor, 11494 Sorrento Valley Rd, San Diego, CA 92121, USA

Cataloging the proteomes of single-celled microorganisms, cells, biological fluids, tissue and whole organisms is being undertaken at a rapid pace as advances are made in protein and peptide separation, detection and identification. For metazoans, subcellular organelles represent attractive targets for global proteome analysis because they represent discrete functional units, their complexity in protein composition is reduced relative to whole cells and, when abundant cytoskeletal proteins are removed, lower abundance proteins specific to the organelle are revealed. Here, we review recent literature on the global analysis of subcellular organelles and briefly discuss how that information is being used to elucidate basic biological processes that range from cellular signaling pathways through protein–protein interactions to differential expression of proteins in response to external stimuli. We assess the relative merits of the different methods used and discuss issues and future directions in the field.

The basic problem of complexity poses a significant challenge in studies to unravel the protein complement of the genome, the proteome. Many of the estimated 30 000–40 000 genes in the human genome [1] are expected to provide multiple protein products that might arise as a result of alternate splicing and post-translational modification. One approach to the daunting prospect of cataloging entire proteomes has been to focus on protein subsets. Organelles, subcellular compartments in a dynamic intracellular membrane system, provide such subsets because they can be subfractionated [2]. Furthermore, the proteomes of organelles comprise a focused set of proteins that fulfills discrete but varied cellular functions. Here, we cover literature published since the last substantial review on the proteomes of major organelles [3] on both isolated organelles, major suborganellar structures and selected large multiprotein complexes that can be isolated in highly purified preparations (Table 1). The reports range in scope to include cataloguing studies that tested the ability of a method to identify as many unique proteins as possible, in particular known low-abundance proteins specific to a particular organelle. Other papers strive to localize newly discovered proteins to specific organelles, and in some cases, perform functional validation. Thus, as we attempt a comprehensive review of recent global organellar proteomic literature, we will refer to the context in which the work was performed and highlight the significance of each paper in terms of methodological and/or functional biological significance. Where applicable, papers dealing with bioinformatic predictions of the total number of proteins that could be targeted to a given organelle will also be discussed. Finally, we discuss whether a protein, once identified, can be ascribed with certainty to a particular organelle, whether it can have a dual localization or whether it might be a contaminant. Recent examples in the literature challenge the traditional notion of organelles as discrete entities, and reflect complex inter-relationships. The generalized structures of plant and animal cells and their respective organelles are illustrated in Fig. 1 [4].

Mitochondria

The mitochondrion provides 80–90% of all the ATP required by mammalian cells in respiring tissues. Mitochondria also signal cell death, modulate ionic balance and drive carbohydrate and fatty acid metabolism. Several lines of evidence point to the existence of at least 1500 distinct proteins in the human mitochondrion. Recent immunolocalization studies in yeast indicated the presence of almost 800 mitochondrial proteins, representing 13% the 6100 gene products in this organism [5]. Therefore, it would not be unreasonable to predict that at least 5% of human genes encode mitochondrial proteins. A detailed analysis of the mitochondrial proteome of *Arabidopsis thaliana* resolved up to 800 proteins on 2D gels [6]. This is almost certainly a gross underestimate because many hydrophobic proteins are not separated in the isoelectric focusing (IEF) dimension and also because low-abundance proteins are either not visible on these gels owing to limitations in sample loading, or are masked by

Table 1. Comprehensive global organellar proteomic investigations with well-curated lists of distinct proteins[a]

Structure	Proteins	Method	Ref.
ER: microsomes[b]	491	ICAT/MudPIT	[31]
Spliceosome	311	Aff/1D-PAGE/GPF LC/MS/MS	[50]
Nucleolus	271	1D-PAGE/ESI/MS/MS	[46]
Nuclear envelope	148	2D-PAGE/PMF/PSD	[42]
Peroxisomes	181	GPF LC/MS/MS	[37]
Mitochondrion	179	1D-PAGE/LC/MS/MS	[16]
Phagosome	140	LB/2D-PAGE/ESI/MS/MS	[40]
Golgi	81	1D-PAGE/Edman/PMF/ESI/MS/MS	[26]
Chloroplast	81	2D-PAGE/Edman/PMF/ESI/MS/MS	[24]
Lysosomes	27	2D-PAGE/PMF/ESI/MS/MS	[39]
Exosomes	21	1D-PAGE/PMF/ESI/MS/MS	[41]

[a]Abbreviations: Aff, affinity capture; ER, endoplasmic reticulum; ESI, electrospray ionization (static nanospray); GPF, gas phase fractionation; ICAT, isotope coded affinity tagging; LB, latex bead encapsulation; LC/MS/MS, liquid chromatography-tandem mass spectrometry (dynamic nanospray); MudPIT, multidimensional protein identification technology; PAGE, polyacrylamide gel electrophoresis; PMF, peptide mass fingerprinting; PSD, post source decay.
[b]Mixed with other organelles.

Corresponding author: Steven W. Taylor (taylors@mitokor.com).

http://tibtec.trends.com 0167-7799/03/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved. PII: S0167-7799(02)00037-9

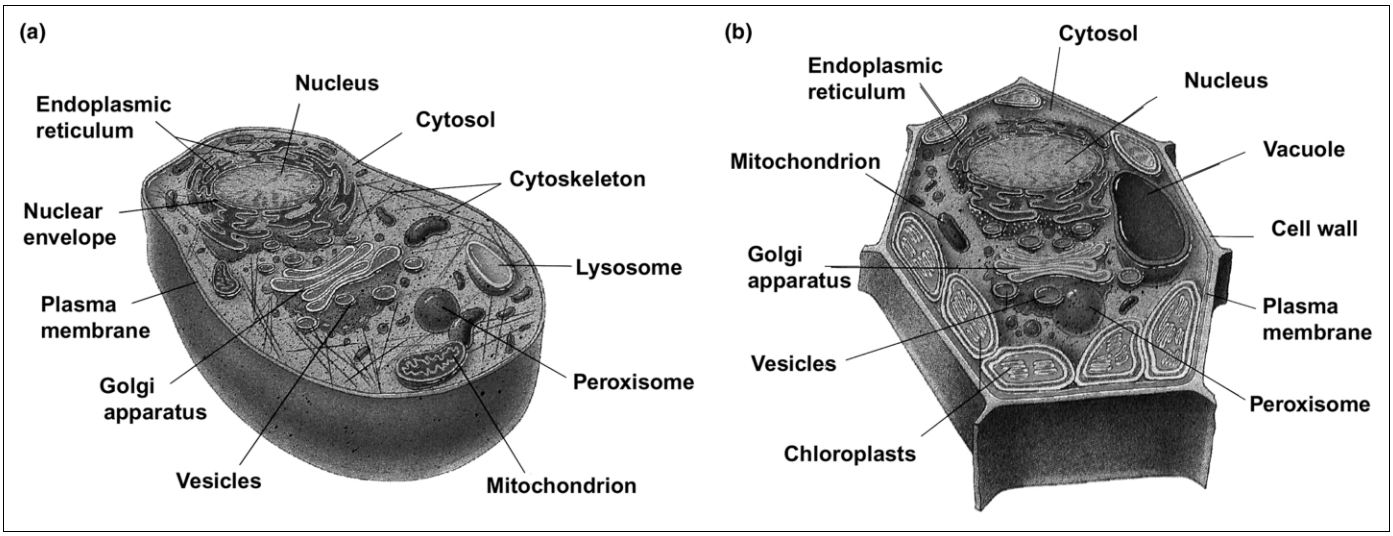

Fig. 1. Eukaryotic cells. (a) Composite animal cell and (b) composite plant cell. Reproduced, with permission, from [4]. © Pearson Education, Inc.

high-abundance proteins. The ability of programs (e.g. PSORT, TargetP and MitoProt) to predict the number of mitochondrial proteins from various eukaryotes has been investigated. However, these *in silico*-based methods have, as yet, failed to attain sufficient specificity and sensitivity to produce confident estimates, in part because numerous mitochondrial membrane proteins do not contain the characteristic N-terminal sequences found in most matrix proteins. However, it will now be possible to incorporate the many newly identified proteins that have been assigned to subcellular compartments into much more comprehensive training sets, which in turn, should lead to a considerable improvement in targeting prediction algorithms.

Several global surveys of mitochondrial proteins have appeared in the literature since the last review on mitochondrial proteomics [7]. *Arabidopsis* mitochondrial proteins have been subfractionated into soluble, total membrane and integral membrane proteins [8]. Following resolution by 2D-polyacrylamide gel electrophoresis (PAGE), 91 proteins were identified from 170 spots by peptide mass fingerprinting (PMF) with the use of matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. The entire dataset of protein sequences was queried through TargetP, PSORT and MitoProt. The three programs predicted, in common, a subset of 32 of these proteins as being targeted to the mitochondrion. The shared feature of the known mitochondrial proteins that these programs did not recognize was the absence of cleavable N-terminal extensions, which act as import signals. The most comprehensive 2D-PAGE report using PMF identified 192 gene products from rat liver mitochondrial proteins, separated by six 2D-PAGE gels with immobilized pH gradient (IPG) strips spanning three different pH ranges [9]. Unfortunately, the 192 identified proteins were not well curated and included redundant identifications and keratin contaminants. Curation of the dataset is an issue that should be closely scrutinized when assessing claims of a large dataset, because mass spectral searching algorithms might match the same set of peptides to two or more different proteins, especially when matches are based on homology.

Another approach favored by several laboratories has been initial separation of intact mitochondrial protein complexes. Resolution of mitochondrial proteins using two different 2D-PAGE systems led to the identification of 52 proteins in *Arabidopsis* mitochondria by PMF, direct protein sequencing and immunoblotting [10]. 2D blue-native (BN) electrophoresis was optimized to give a high-throughput and reproducible system for mitochondrial proteomics [11]. BN-/Tricine-SDS PAGE was used in another study to separate human heart mitochondrial proteins, which were identified using PMF and liquid chromatography-tandem mass spectrometry (LC/MS/MS), giving good coverage of the five complexes of the oxidative phosphorylation machinery [12]. The three electrophoretic systems (BN/IEF/SDS PAGE) were combined to produce a novel '3D proteome map' [6]. An analogous strategy for separating intact protein complexes with the use of sucrose gradients followed by 2D-PAGE [13] was subsequently adapted into a high-throughput MALDI PMF [14] study of highly purified heart mitochondrial proteins, but using 1D-PAGE. A list of 82 distinct mitochondrial proteins was compiled from ten sucrose fractions and from the *N*-dodecyl-β-D-maltoside-insoluble pellet. Approximately half of these proteins had previously escaped detection on 2D gels because they were small, basic or highly hydrophobic. Interestingly, both these studies demonstrated the value of separating proteins by the size of their multiprotein complexes, and found that the 30-kDa protein, prohibitin, participated in a high-molecular weight complex of similar size to complex I [12,14]; this is significant because the function of prohibitin had been poorly defined until evidence arose that it acts a molecular chaperone for respiratory complex assembly in a 1-MDa complex called PHB [15].

Pflieger *et al.* [16] adopted a LC/MS/MS approach and identified 179 gene products from yeast mitochondria that had been directly subjected to SDS PAGE. The physiochemical properties of these proteins were found to span wide ranges of pI, molecular weight and hydrophobicity. The broad coverage of proteins afforded by this approach circumvents some of the limitations of 2D-PAGE [16].

To date, this study is the most extensive report for a well-curated dataset of mitochondrial proteins.

Chloroplasts

With the full genome of *Arabidopsis* completed in 2000 [17], the proteome of chloroplasts has never been so accessible. Using bioinformatics approaches, Abdallah *et al.* [18] predicted that the *Arabidopsis* chloroplast proteome consists of 1900–2500 proteins, including 87 proteins encoded by chloroplast DNA. The rest are nuclear-encoded and transported to the chloroplast. This estimate was the result of a homology-based comparison of the chloroplast proteome of *Arabidopsis*, the photosynthetic cyanobacterium *Synechocystis* and yeast as a model of the ancient mitochondriate host. Several criteria were considered for the prediction of chloroplast proteins that included: (1) the presence of N-terminal presequences called transit peptides that target nuclear-encoded proteins to the chloroplast; (2) the level homology to *Synechocystis*; or (3) lack of homology to yeast. Interestingly, on the basis of homology 35% of all chloroplast proteins were predicted to be of cyanobacterial origin. In a major review of chloroplast proteomics, the status of the analysis of chloroplast compartments was discussed [19]. Major developments have since come from studies focusing on thylakoid membranes (the site of the light-dependent reactions that produce NADPH and ATP) and the lumen (the aqueous space within the thylakoid membrane).

Hippler *et al.* [20] isolated thylakoid membranes from *Chlamydomonas reinhardtii*, a eukaryotic green algae. Following organic solvent extraction of the membranes, they used a 2D gel-based approach to resolve and identify several light-harvesting complex proteins, some of which were very hydrophobic, in 30 spots from 2D gels using a combination of immunoblotting and MS. Tandem MS analysis identified a putative transit sequence peptide of a light-harvesting complex II protein. Differential processing and/or post-translational modification of several light-harvesting complex I proteins was suggested by the patterns of immunoreactivity on 2D gels. Kashino *et al.* [21] identified 31 distinct proteins in a purified photosystem II preparation from *Synechocystis* with a combination of SDS PAGE, PMF and Edman sequencing. By contrast, Gomez *et al.* [22] probed photosystem II-enriched thylakoid membrane subdomains (grana) from pea and spinach by using LC/MS to measure the mass of reversed phase HPLC-separated proteins. The resultant 'intact mass tags' were used to identify proteins based on a comparison with genomic data in which protein molecular weights were corrected for post-translational modification. Around 90 mass tags were used to provisionally identify 40 gene products. This method is useful for systems where the number of gene products and post-translational modifications is limited.

Lumenal proteins have several well-characterized functions that include water splitting, electron transport and the violaxanthin cycle. Two groups have combined mass spectrometry-based proteomic and bioinformatics approaches to predict the number of lumenal proteins. One group used PMF, post source decay (PSD) and Edman sequencing to identify 36 proteins from 2D gels [23]. Because many lumenal proteins contain a C-terminal twin arginine motif, in addition to N-terminal chloroplast transit peptide, TargetP and SignalP algorithms were used in combination with MS data to predict 80 lumenal proteins. The other group used PMF and electrospray ionization (ESI) MS/MS on 2D gels of thylakoid lumen to identify 81 proteins, 30 of which were lumenal and 32 of which were peripheral thylakoid proteins [24]. Using the TargetP and SignalP algorithms, combined with a transmembrane predictor and the MS data, they estimated a lumenal proteome of 71 proteins. Because the overlap between these two studies of the lumenal proteome consists of only 24 proteins, it will be important to compare the actual sequences to evaluate the congruency of the two datasets. Both groups found similarities between their data for *Arabidopsis* and those for other higher plants, suggesting that *Arabidopsis* is a good model organism for higher plants.

Golgi complex

The Golgi complex has a crucial role in the post-translational modification of newly synthesized proteins and of complex carbohydrates through precisely programmed oligosaccharide-processing pathways. In addition, the Golgi complex is crucial for accurately sorting the transit of the mature glycoproteins and membrane lipids to their select destinations within the cell and for cellular export. It is estimated that there are 1000 proteins that comprise the Golgi proteome, of which <200 are known [25].

A proteomic analysis of proteins from a purified rat hepatic Golgi fraction that was partitioned into Triton X114 has been reported [26]. The proteins were separated using 1D-SDS PAGE and subjected to a combination of Edman sequencing, PMF and MS/MS analyses. A total of 81 proteins was identified, of which 49 were characterized as integral membrane proteins on the basis of one or more predicted transmembrane domains. The identified protein set was predominated by Golgi-resident enzymes and membrane trafficking proteins, such as Rabs, KDEL receptors, p24 family members and soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs). In addition, a previously unidentified 34-kDa protein, GPP34, was identified, which is conserved from yeast to humans. The intra-Golgi localization of GPP34 and selected trafficking proteins was confirmed by analytical fractionation and gold immunolabeling. Protein contaminants from the endoplasmic reticulum (ER) and the mitochondria were also identified and represented by 24 proteins, which tallied with an estimate of $\sim$84% purity by morphometry of the starting Golgi fraction.

Highly specialized post-Golgi carrier membranes mediate the polarized trafficking of rhodopsin to rod outer segments of retinal rod photoreceptor cells. These carrier membranes bud from the trans-Golgi network and deliver their rhodopsin cargo by fusing with the specialized domain of plasma membrane of rod inner segment. A subcellular fractionation procedure of frog retinal cells obtained an enriched fraction of rhodopsin-bearing post-Golgi carrier membranes [27], and several cytoskeletal proteins were identified by MS/MS after 2D-PAGE.

During late pregnancy, rat mammary epithelial cells

maintain basal secretion levels of milk components that within 1–4 days after parturition, shift to maximal secretion. Organellar compartments involved in milk secretion in these epithelial cells undergo rapid proliferation and expansion during the transition from late pregnancy to lactation. A comparative study of basal and maximal secretion states of rat mammary epithelial cells illuminated the proteomic changes that occur in the Golgi complex to accommodate the increased secretory demands [28]. Thirty upregulated proteins, belonging to the regulatory, trafficking and signal-transduction protein classes, were identified in the maximal secretory state. The observed changes reflect a high degree of regulation during the functional transition of the organelle, and it would be pertinent to ascertain if the altered protein expression occurs in a coordinated or temporal fashion.

ER

ER has a crucial cellular role in the biosynthesis and transport of proteins and lipids and in the metabolism of xenobiotics and endogenous compounds. The hepatic ER system uses the action of cytochrome P450 isozymes, NADPH and NADH reductases and transferases to carry out oxidation, glycosation and sulfation of drug substrates for eventual renal excretion of hydrophilic metabolites. Proteins from liver microsomes isolated from phenobarbital-treated and control rats were analyzed by 1D- or 2D-gel electrophoresis, which was then followed by PMF of in-gel digests [29]. This study highlighted the superiority of the 1D-gel approach for the identification of membrane-bound proteins. In addition, phenobarbital was shown to induce the expression of a large number of proteins, exemplified by the detection of cytochromes P450 2B1 and 2B2 and several stress-related ER proteins. A 2D-PAGE-based proteomic analysis of the ER from developing and germinating seed of castor found that sample complexity was reduced by separation of ER proteins into lumenal, peripheral membrane and integral membrane subfractions [30]. Protein identification by PMF and ESI/MS/MS was limited by the small genomic database for the plant.

Although 2D-gel electrophoresis is widely used in proteomic studies, the method suffers from a significant drawback in the separation of hydrophobic, membrane-associated proteins. In a major advance, Han *et al.* [31] used a non-gel-based isotope-coded affinity tagging (ICAT) approach [32] to map and quantify microsomal fractions, which had been isolated from naive and phorbol ester-differentiated human myeloid leukemia (HL-60) cells. It should be noted that in addition to the ER, microsomal fractions consisted of all membrane-bound vesicles separated from the soluble cytosol, such as plasma membranes, Golgi, mitochondria and lysosomes. They identified 491 proteins, many of which were transmembrane or membrane associated. The data indicated changes in signal transduction proteins and enzymes that might be directly or indirectly related to phorbol ester treatment.

The secretion of lipids by mammary epithelial cells is a highly ordered process. Following synthesis in the ER, the lipids are released, packaged into cytoplasmic lipid droplets (CLD) and vectorially transported to the apical plasma membrane for fusion and secretion as milk fat globules. A 2D-PAGE study of CLDs and the encasing milk fat globule membrane (MFGM) component of the globule has provided evidence for a mechanistic role of ER in this secretory pathway [33]. The key findings are that mammary CLDs have a different protein composition from liver CLDs and that they contain components in their membrane-like structures that are of ER origin. The CLDs and MFGM share a subset of ER proteins that suggests a role for these proteins in the secretory process.

Peroxisomes and lysosomes

Peroxisomes and lysosomes – eukaryotic organelles encased in a single membrane – carry out oxidative metabolism reactions and degradation of macromolecules, respectively, to provide small metabolites that can be neutralized or excreted by the cell. In the yeast genome, there are ~300 proteins with peroxisomal targeting signal (PTS) motifs [34]. In the greening cotyledon leaves of *Arabidopsis*, 29 peroxisomal proteins were identified by 2D-PAGE and PMF, including five belonging to the glycolate pathway and four involved in scavenging of hydrogen peroxide [35]. Twenty novel proteins, including kinases, phosphatases and five proteins with PTS motifs, have also been identified. In addition, Schäfer *et al.* [36] used a 1D-gel electrophoretic approach, coupled with PMF and LC/MS/MS, to identify 45 proteins (22 of which are peroxisomal) from isolated yeast peroxisomal membranes, and discovered a phosphorylation site in the long-chain fatty acid CoA ligase 2. The significance these kinases, phosphatases and the phosphorylation site is unknown, but might reveal signaling pathways available to the peroxisome that have not been well defined.

Another effort to characterize proteins on yeast peroxisomes used a gel-free strategy, subjecting tryptic peptides to LC/MS/MS directly [37]. First, data-dependent mass-spectral acquisition parameters were varied for identical injections of a whole-cell lysate of yeast in which peroxisome proliferation had been stimulated by treatment with oleic acid. Different peptides were selectively fragmented by gas-phase fractionation (GPF) according to their mass:charge ratio (*m/z*) or their relative ion intensity (RI). Triplicate injections over three *m/z* ranges of the whole-cell lysate tended to identify more proteins, whereas triplicate injections according to three RI criteria tended to yield lower abundance proteins because low-intensity peptide ions were being selectively fragmented. When GPF in 13 overlapping 100 *m/z* ranges was applied to Nycodenz gradient-purified peroxisomes, 181 proteins were identified, including 38 of 41 experimentally characterized or predicted peroxisomal proteins. This study provided the most comprehensive coverage of the peroxisome proteome to date, illustrated by the detection of low-abundance proteins such as Pex5p.

Mammalian soluble lysosomal proteins are distinguished by an *N*-linked modification with mannose-6-phosphate, which can be exploited by an affinity-based approach to isolate and identify these proteins from human monocytic U937 cells [38]. A study describes the progress towards the construction of a 2D reference map of the soluble lysosomal proteome [39]. Mass spectrometry revealed 22 proteins, including 16 well-known lysosomal

hydrolases, in addition to newly identified members of the lysosome [a novel mannosidase homolog, cystatin F and CREG (cellular repressor of E1A-stimulated genes)]. A comparison of 2D-PAGE difference maps for U937 control cells, cells undergoing phorbol ester-induced differentiation into macrophage-like cells, and human breast cancer MCF7 cells revealed differences in the limited proteolytic cleavage or maturation of a discrete number of hydrolases. Five additional lysosomal hydrolases were identified from MCF7 samples after 2D-PAGE.

Phagosomes and exosomes

Macrophages participate in tissue remodeling: they restrict the spread of intracellular pathogens and clear apoptotic cells by engulfing particulate material and forming phagosomes. The first comprehensive analysis of the phagosome proteome has revealed key insights into the function of this organelle [40]. 2D- and 1D-PAGE, in combination with PMF and MS/MS, were used to deconvolute the protein complement in latex bead-associated mouse phagosomes. More than 140 proteins were identified that included several new proteins and a series of proteins not known to be associated with this organelle (e.g. galectin-3, Alix, TRAIL, 14-3-3, rab14, flotillin-1, prohibitin and stomatin). Sensitivity to pronase proteolysis revealed the suborganellar localization of several proteins. Modulation of phagosome composition during biogenesis, particularly in hydrolases, was apparent from 2D-gel maps of phagosomes isolated at different time points.

Various types of blood cells and dendritic cells secrete a population of homogenous membrane vesicles (exosomes) into the extracellular environment, which have been suggested to play a role in humoral immune responses. A proteomic study that used SDS PAGE coupled with PMF and MS/MS identified 21 proteins in dendritic exosomes, including several transmembrane proteins already known to be present in this organelle [41]. The proteins that were newly identified came predominantly from the cytosol, where they are known to be involved in intracellular membrane transport, signaling, cytoskeletal structure and apoptosis. Importantly, exosomes appeared to be biochemically and morphologically distinct from the vesicles formed by membrane blebbing of apoptotic dendritic cells.

The nucleus

The nucleus can be subdivided into several components. The nuclear envelope is a bimembrane structure perforated by protein-rich nuclear pores and is continuous with the ER. The envelope is lined by a filamentous protein structure called the nuclear lamina that, in addition to helping maintain organellar structure, interacts with nuclear pore complexes and genetic material. Furthermore, this nuclear proteomic structure is important for importing protein from and exporting RNA to the cytoplasm. Differential detergent-solubilization of the nuclear envelope into three subfractions, 2D-PAGE, and MALDI PMF and PSD fragmentation yielded the identity of 148 different proteins [42]. Of these, 19 were previously uncharacterized, two of which were cloned and their recombinant constructs immunolocalized to the inner nuclear membrane.

More recently, 135 proteins were catalogued that interacted with nucleoporins from the supermolecular assembly of the nuclear pore protein complex in yeast [43]. Thirteen specific nucleoporins with FG peptide repeats (FG Nups) as immobilized protein 'bait', were used to affinity capture interacting partners, which were then separated by 1D-SDS–PAGE and analyzed by PMF and LC/MS/MS. The captured proteins consisted mainly of other nucleoporins, in addition to karyopherins, proteins that facilitate nucleocytoplasmic transport. Interestingly, the study found that the presence of the GTP-bound form of the GTPase, Gsp1p, modulates nucleoporin–karyopherin–cargo interactions. Comprehensive data maps of protein interactions in the nuclear pore complex were assembled to mimic the effect of the steep Gsp1p–GTP gradient across the complex. This gradient is thought to drive nucleocytoplasmic transport, with Gsp1p–GTP binding to importins facilitating dissociation of cargo from importins, and Gsp1p–GTP binding to exportins enhancing association of cargo with exportins [44]. Analysis of nuclear pore complexes has recently been extended to a mammalian system [45]. Of the proteins identified from highly purified nuclear pore complexes in rat nuclei, 29 were classified as nucleoporins and 18 as proteins associated with the complexes.

The nucleolus is a dense mass in the nucleus, which is a major site of ribosomal RNA synthesis and ribosome assembly. This sub-organelle was purified from human HeLa cell cultures [46] and subjected to 1D- and 2D-PAGE, followed by PMF and MS/MS. From a total of 271 proteins, 20 were only found with the 2D approach. The bulk of the proteins identified came from the analysis of two different 1D gels. The major classes of functionally classified proteins included the nucleotide and nucleic acid-binding proteins (24%), ribosomal proteins (14%), and RNA-modifying enzymes and related proteins (11%). However, a large proportion (>30%) was novel and of unknown localization (although a subset of 18 proteins from this group was cloned as fusion products with yellow fluorescent protein and 15 of the chimeric constructs localized to the nucleolus).

The response of the nucleolus proteome to structure-induced changes by the transcription inhibitor actinomycin D was monitored with 1D-PAGE, followed by MS to identify 11 differentially expressed proteins. Treatment with actinomycin D was found to increase the localization of all these proteins to the nucleolus (tracked by either antibodies or fluorescent cloned fusion products), supporting the idea that proteins are rapidly trafficked between different nuclear compartments [47].

The nuclear matrix is a proteinaceous scaffold involved in the spatial organization of the interphase nucleus. During apoptosis, nuclear matrix proteins (e.g. lamins) are degraded by caspases, followed by a process called apoptotic chromatin condensation (ACC). Changes in the 2D-gel maps of isolated nuclear matrix proteins in cultured cells were monitored during ACC by PMF, western blotting and Edman sequencing, highlighting 26 proteins that decreased in relative abundance and 28 proteins that

accumulated in the nuclear matrix [48]. It was proposed that 14 unaffected proteins were constituents of an elementary nuclear scaffold linked to chromatin by the proteins that were differentially expressed.

Two independent surveys of the spliceosome were recently undertaken [49,50]. In one study, maltose-binding affinity chromatography was used to isolate splicesomes in a highly purified and functional form. From the complex purified by 1D-PAGE, ~145 distinct spliceosomal proteins were identified by LC/MS/MS of in-gel digests of 10 gel slices [49]. In the other study, biotinylated, radioactively labeled RNA was used to assemble a mixture of spliceosomal complexes that was subsequently subjected to gel filtration and affinity selection on streptavidin beads [50]. Proteins were then identified from in-gel trypsin digests from 1D-SDS PAGE by LC/MS/MS over several *m/z* ranges, and 311 distinct proteins were subsequently identified. To date, this subset of nuclear proteins is the largest dataset generated for any purified subcellular organelle in a single study. The dataset included all known essential spliceosome-associated factors, in addition to 96 novel proteins (55 of which contain domains indicative of functional roles in splicing and RNA processing).

Contamination, dual localization or interorganellar contact sites?

In spite of the common view that organelles are discrete entities carrying out independent cellular functions, there are complex mechanisms of intracellular communication and contact sites between the organelles. This complexity makes it hard to evaluate the biological significance of proteins that are usually associated with one organelle but are detected in the proteome of another organelle. Although these proteins could be artifacts of subcellular fractionation procedures (resulting in contamination from keratin, proteins from other organelles, the plasma membrane and cytoplasm), they might also be biologically significant.

Schäfer *et al.* [36] identified many mitochondrial proteins in their peroxisomal membrane preparation, but could not distinguish whether this was because of cross contamination or, as for carnitine acetyl transferase, because they were located in both organelles. Similarly, lysosomal cathepsin B and some peroxisomal proteins were earlier found amongst proteins released from mouse liver mitochondria undergoing the membrane-permeability transition [51].

Garin *et al.* [40] discussed finding plasma membrane, mitochondrial, cytoplasmic and ER proteins in their phagosome preparation. However, unlike most subfractionation procedures that are based on organelle density, they separated phagosomes by floating them away from other organelles on latex beads. The presence of the ER proteins was subsequently found to have great biological significance when it was shown that the ER is directly recruited to the plasma membrane to form phagosomes [52], changing long-held views of phagosome formation.

Conclusion and outlook

Improved proteome coverage of highly purified preparations of organelles is being rapidly achieved through innovation in non-gel, 1D-, 2D- and 3D- protein and peptide separation systems. Recent research with large proteome-scale epitope tagging and immunolocalization [5], and improved subcellular fractionation procedures is providing direct evidence for the existence of proteins in one or more compartments within the cell. Future efforts will undoubtedly focus on incorporating such data to predict the multiple localization of gene products that might result from the existence of multiple sequence forms (splice-variants or otherwise) or because they have a more generalized functional role in the cell. As demonstrated above for the nucleoporins [43], co-precipitation with tagged target proteins can be used to generate comprehensive protein–protein-interaction data maps. Databases of such networks might provide researchers with the first insights into the function of many of these unannotated proteins. However, in spite of these advances, the bottleneck to further progress will probably be the complete functional elucidation of the thousands new proteins that are likely to be identified in the next few years [53]. The challenge of global organellar-based proteomics is to provide a functional context for these proteins by associating them with a distinct group of proteins in defined intracellular environments.

References

1 Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351
2 Storrie, B. and Madden, E.A. (1990) Isolation of subcellular organelles. *Methods Enzymol.* 182, 203–225
3 Jung, E. *et al.* (2000) Proteomics meets cell biology: the establishment of subcellular proteomes. *Electrophoresis* 21, 3369–3377
4 Horton, R.H. *et al.* (1996) *Principles of Biochemistry*, Prentice Hall
5 Kumar, A. *et al.* (2002) Subcellular localization of the yeast proteome. *Genes Dev.* 16, 707–719
6 Werhahn, W. and Braun, H.P. (2002) Biochemical dissection of the mitochondrial proteome from *Arabidopsis thaliana* by three-dimensional gel electrophoresis. *Electrophoresis* 23, 640–646
7 Lopez, M.F. and Melov, S. (2002) Applied proteomics: mitochondrial proteins and effect on function. *Circ. Res.* 90, 380–389
8 Millar, A.H. *et al.* (2001) Analysis of the *Arabidopsis* mitochondrial proteome. *Plant Physiol.* 127, 1711–1727
9 Fountoulakis, M. *et al.* (2002) The rat liver mitochondrial proteins. *Electrophoresis* 23, 311–328
10 Kruft, V. *et al.* (2001) Proteomic approach to identify novel mitochondrial proteins in *Arabidopsis*. *Plant Physiol.* 127, 1694–1710
11 Brookes, P.S. *et al.* (2002) High throughput two-dimensional blue-native electrophoresis: a tool for functional proteomics of mitochondria and signaling complexes. *Proteomics* 2, 969–977
12 Devreese, B. *et al.* (2002) Mass spectrometric identification of mitochondrial oxidative phosphorylation subunits separated by two-dimensional blue-native polyacrylamide gel electrophoresis. *Electrophoresis* 23, 2525–2533
13 Hanson, B.J. *et al.* (2001) A novel subfractionation approach for mitochondrial proteins: a three-dimensional mitochondrial proteome map. *Electrophoresis* 22, 950–959
14 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 Research* 1, 451–458
15 Nijtmans, L.G. *et al.* (2002) The mitochondrial PHB complex: roles in mitochondrial respiratory complex assembly, ageing and degenerative disease. *Cell. Mol. Life Sci.* 59, 143–155
16 Pflieger, D. *et al.* (2002) Systematic identification of mitochondrial proteins by LC-MS/MS. *Anal. Chem.* 74, 2400–2406
17 Initiative, T.A. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
18 Abdallah, F. *et al.* (2000) A prediction of the size and evolutionary

origin of the proteome of chloroplasts of *Arabidopsis*. *Trends Plant Sci.* 5, 141–142
19 Jan van Wijk, K. (2000) Proteomics of the chloroplast: experimentation and prediction. *Trends Plant Sci.* 5, 420–425
20 Hippler, M. *et al.* (2001) Towards functional proteomics of membrane protein complexes: analysis of thylakoid membranes from *Chlamydomonas reinhardtii*. *Plant J.* 28, 595–606
21 Kashino, Y. *et al.* (2002) Proteomic analysis of a highly active photosystem II preparation from the *Cyanobacterium synechocystis* sp. PCC 6803 reveals the presence of novel polypeptides. *Biochemistry* 41, 8004–8012
22 Gomez, S.M. *et al.* (2002) The chloroplast grana proteome defined by intact mass measurements from liquid chromatography mass spectrometry. *Mol. Cell Proteomics* 1, 46–59
23 Schubert, M. *et al.* (2002) Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. *J. Biol. Chem.* 277, 8354–8365
24 Peltier, J.B. *et al.* (2002) Central functions of the lumenal and peripheral thylakoid proteome of *Arabidopsis* determined by experimentation and genome-wide prediction. *Plant Cell* 14, 211–236
25 Taylor, R.S. *et al.* (2000) Proteomics of rat liver Golgi complex: minor proteins are identified through sequential fractionation. *Electrophoresis* 21, 3441–3459
26 Bell, A.W. *et al.* (2001) Proteomics characterization of abundant Golgi membrane proteins. *J. Biol. Chem.* 276, 5152–5165
27 Morel, V. *et al.* (2000) Towards the proteome of the rhodopsin-bearing post-Golgi compartment of retinal photoreceptor cells. *Electrophoresis* 21, 3460–3469
28 Wu, C.C. *et al.* (2000) Proteomic analysis of two functional states of the Golgi complex in mammary epithelial cells. *Traffic* 1, 769–782
29 Galeva, N. and Altermann, M. (2002) Comparison of one-dimensional and two-dimensional gel electrophoresis as a separation tool for proteomic analysis of rat liver microsomes: Cytochromes P450 and other membrane proteins. *Proteomics* 2, 713–722
30 Maltman, D.J. *et al.* (2002) Proteomic analysis of the endoplasmic reticulum from developing and germinating seed of castor (*Ricinus communis*). *Electrophoresis* 23, 626–639
31 Han, D.K. *et al.* (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* 19, 946–951
32 Gygi, S.P. *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999
33 Wu, C.C. *et al.* (2002) Proteomics reveals a link between endoplasmic reticulum and lipid secretory mechanisms in mammary epithelial cells. *Electrophoresis* 21, 3470–3482
34 Geraghty, M.T. *et al.* (1999) Detecting patterns of protein distribution and gene expression *in silico*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 2937–2942
35 Fukao, Y. *et al.* (2002) Proteomic analysis of leaf peroxisomal proteins in greening cotyledons of *Arabidopsis thaliana*. *Plant Cell Physiol.* 43, 689–696
36 Schäfer, H. *et al.* (2001) Identification of peroxisomal membrane proteins of *Saccharomyces cerevisiae* by mass spectrometry. *Electrophoresis* 22, 2955–2968
37 Yi, E.C. *et al.* (2002) Approaching complete peroxisome characterization by gas-phase fractionation. *Electrophoresis* 23, 3205–3216
38 Journet, A. *et al.* (2000) Towards a human repertoire of monocytic lysosomal proteins. *Electrophoresis* 21, 3411–3419
39 Journet, A. *et al.* (2002) Proteomic analysis of human lysosomes: application to monocytic and breast cancer cells. *Proteomics* 2, 1026–1040
40 Garin, J. *et al.* (2001) The phagosome proteome: insight into phagosome functions. *J. Cell Biol.* 152, 165–180
41 Théry, C. *et al.* (2001) Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles. *J. Immunol.* 166, 7309–7318
42 Dreger, M. *et al.* (2001) Nuclear envelope proteomics: novel integral membrane proteins of the inner nuclear membrane. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11943–11948
43 Allen, N.P. *et al.* (2001) Proteomic analysis of nucleoporin-interacting proteins. *J. Biol. Chem.* 276, 29268–29274
44 Nakielny, S. and Dreyfuss, G. (1999) Transport of proteins and RNAs in and out of the nucleus. *Cell* 99, 677–690
45 Cronshaw, J.M. *et al.* (2002) Proteomic analysis of the mammalian nuclear pore complex. *J. Cell Biol.* 158, 915–927
46 Andersen, J.S. *et al.* (2002) Directed proteomic analysis of the human nucleolus. *Curr. Biol.* 12, 1–11
47 Misteli, T. (2001) Protein dynamics: implications for nuclear architecture and gene expression. *Science* 291, 843–847
48 Gerner, C. *et al.* (2002) Proteome analysis of nuclear matrix proteins during apoptotic chromatin condensation. *Cell Death Differ.* 9, 671–681
49 Zhou, Z. *et al.* (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature* 419, 182–185
50 Rappsilber, J. *et al.* (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res.* 12, 1231–1245
51 Patterson, S.D. *et al.* (2000) Mass spectrometric identification of proteins released from mitochondria undergoing permeability transition. *Cell Death Differ.* 7, 137–144
52 Gagnon, E. *et al.* (2002) Endoplasmic reticulum-mediated phagocytosis is a mechanism of entry into macrophages. *Cell* 110, 119–131
53 Hubbard, M.J. (2002) Functional proteomics: The goalposts are moving. *Proteomics* 2, 1069–1078

Do you want to reproduce material from a *Trends* journal?

This publication and the individual contributions within it are protected by the copyright of Elsevier Science. Except as outlined in the terms and conditions (see p. ii), no part of any *Trends* journal can be reproduced, either in print or electronic form, without written permission from Elsevier Science. Please address any permission requests to:

Rights and Permissions,
Elsevier Science Ltd,
PO Box 800, Oxford, UK OX5 1DX.