Organelle proteomics: looking at less to see more

Sylvain Brunet[1], Pierre Thibault[2], Etienne Gagnon[1], Paul Kearney[2], John J.M. Bergeron[2,3] and Michel Desjardins[1,2]

[1]Département de pathologie et biologie cellulaire, Université de Montréal, C.P. 6128, Succ. Centre ville, Montreal, Canada H3C 3J7
[2]Caprion Pharmaceuticals, Montreal, Canada H4S 2C8
[3]Department of Anatomy and Cell Biology, McGill University, Montreal, Canada H3A 2B2

The recent finding that the human genome comprises between 21 000 and 39 000 genes, a number much lower than expected, has in no way simplified the complexity associated with the understanding of how cells perform their functions. Elucidation of the molecular mechanisms underlying cell functions will require a global knowledge of the expressed proteins, including splice variant products, their post-translational modifications, their subcellular localizations and their assembly into molecular machines as deduced from protein–protein interactions, at any given time during the life of the cell or under any cellular conditions. Current and expected advances in mass spectrometry and bioinformatics might help the realization of these goals in a shorter time than is currently predicted.

Although promising, the information gathered in the last few years from large-scale proteomics programs indicates that analysis of whole cells or tissues might be too complex for the available technology and not suited for the study of proteins that are present in low copy numbers [1,2]. There are currently several limitations to the analysis of complex samples. First, although the total number of gene products present in any given cell is estimated to be around 10 000, the effective number is likely to be several fold higher owing to splice variants and the wide variety of possible post-translational modifications [2]. Second, when one considers the large variation in the level of expression of various proteins in a cell or tissue (which can range from 3–4 to over 10 orders of magnitudes), the complexity becomes such that proteins in low abundance are inevitably masked by those in high abundance [3]. Since most of the regulatory proteins such as kinases, GTPases and certain membrane receptors are present in low copy numbers, important layers of information are missing from studies of whole cells or tissue. In several cases, the low abundance proteins become even more elusive because of the limited amount of material available for analysis (e.g. from medical biopsies). Although genomic approaches have taken advantage of amplification methods such as the polymerase chain reaction, proteomics is still limited by the inability to amplify the protein content of any biological sample.

Nevertheless, biologists can use various approaches to concentrate and enrich the low abundance proteins of interest. For example, the total protein content of cells can be simplified by fractionation methods based on affinity procedures to isolate groups of proteins displaying similar features (lectin-based isolation of glycoproteins, the use of specific antibodies to isolate phosphoproteins, etc.). This procedure has the advantage of simplifying the complexity of crude cell or tissue extracts, thereby maximizing the probability of detecting low abundance proteins in the mass spectrometer; even so, this technique does not provide information on the location of proteins in the cell, organ or tissue under analysis. The organization of the eukaryotic cell into subcompartments of specialized function, the organelles, provides a unique opportunity to link proteomics data with functional units. It is thus not surprising that, for cell biologists, identifying the proteins or 'players' present in organelles has always been considered an important step towards understanding the molecular mechanisms governing their functions. Here, we discuss how proteomics analysis of organelles is helping to unravel novel concepts in cell biology, leading to a better understanding of complex cellular processes in health and disease.

Cell biologists isolate organelles

The first observation of cells under the electron microscope in the late 1940s and early 1950s indicated a level of organization that was much more elaborate than had previously been assumed [4]. A new world was revealed, where compartmentalization was achieved by intracellular membranes. In 1974, the Nobel Prize in Physiology or Medicine was awarded to Albert Claude, George Palade and Christian de Duve for their discoveries concerning the structural and functional organization of the cell. These founding fathers of modern cell biology were among the first to subfractionate the eukaryotic cell into functional compartments, leading to the first molecular characterization of mitochondria, lysosomes, peroxisomes and the entire secretory apparatus (for an historical review see [5]).

Refinement of the methods and media used for subcellular fractionation has provided the means to analyze the composition of isolated organelles and develop cell-free assays to study and reconstitute complex cellular processes (for a recent review, see [6]). However, a major

Corresponding author: Michel Desjardins (michel.desjardins@umontreal.ca).

 0962-8924/$ - see front matter doi:10.1016/j.tcb.2003.10.006

limitation is the difficulty in assessing the degree of purity of these fractions. At best, a level of enrichment can be assigned by measuring the level of expected proteins of these fractions compared with the starting materials and the decrease, or lack of, markers from other compartments. Using these criteria, various levels of enrichment have been claimed for various organelles in the literature. Recently, the improved ability to generate highly enriched organelle preparations has provided what might turn out to be a key approach to compensate for the lack of amplification methods to study protein samples. The advantage of using an organelle approach to analyze the cell proteome and characterize its low abundance proteins is illustrated in Figure 1.

Characterization of the protein content of organelles began long before the introduction of the neologism 'proteomics' and the use of the mass spectrometer. The advance of high-throughput methods of analysis has simply changed the scale at which this effort can be undertaken (Figure 2, Box 1). Accordingly, major research programs to analyze the proteome of each of the cell organelles are currently underway in various laboratories.

The cell map

Like maps of the New World, cell maps [7] are valuable tools for further discovery. Beyond the generation of simple lists of proteins, organelle proteomics studies are leading to major discoveries in cell biology and are contributing to reshape our understanding of the cell and its organelles (Figure 3). It is beyond the scope of this article to report and comment on all the studies performed to analyze and identify proteins from cellular organelles. Selected examples have been chosen that have led to the uncovering of paradigms which themselves have generated unexpected hypotheses regarding cellular functions.

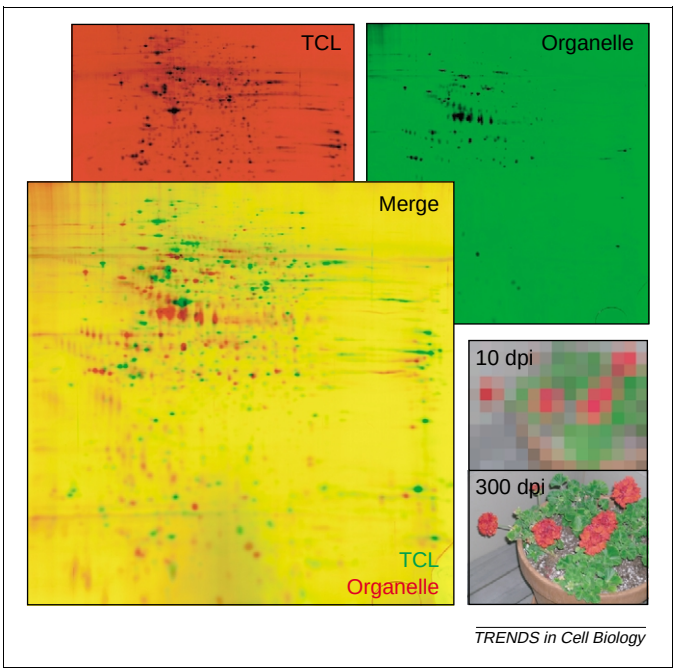

Figure 1. Organelle proteomics brings a new resolution to cellular processes. When identical amounts of proteins from a total macrophage lysate or purified subcellular phagosome preparation are loaded onto separate 2-D gels, several hundred spots are identified in each gel. In principle, all the proteins present in the phagosome preparation (organelle) should be present in the total cell lysate (TCL) from which the phagosomes came. However, superimposition of the two gels (merge) reveals that only 20 or so of the phagosome protein spots are detected in the total cell lysate, of which actin is the most prominent, indicating that more than 90% of the phagosome proteins would have been undetected by analysis of the total cell lysate alone.

The nucleus

The long-standing model of the nucleus as a compartment in which the nuclear components are randomly distributed is being revisited as more and more studies indicate the existence of a higher level of structural organization [8]. Recent proteomics analysis has contributed to that view. Among the subcompartments of the nucleus, the nuclear envelope is one of the least characterized. Analysis of purified nuclear envelope by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) led to the identification of 148 proteins, of which 19 were 'new' proteins (predicted gene products) [9]. Of these new proteins, two integral membrane proteins (LUMA protein and one related to Unc-84A of *Caenorhabditis elegans*) were identified. Overexpression of these proteins in COS-7 cells and immunofluorescence staining confirmed their localization in the inner nuclear membrane where they are likely to interact with the nuclear lamina. The nuclear pore complex, a substructure of the nuclear envelope of prime importance in the control of the trafficking of molecules between the nucleus and the cytoplasm, has also been thoroughly analyzed by proteomics. A total of 174 proteins were identified from yeast nuclear pore preparations, of which 29 were localized to the nuclear pore by immunoelectron microscopy, enabling a fine molecular characterization of this structure. This in turn led to a radical new hypothesis on the mechanism of transport of cargo across the nuclear pore in which a stochastic mechanism selects for cargo transport into and out of the nucleus [10].

In the nucleus, the transcription of genes into mRNA involves the removal of introns by a splicing process performed by the multiprotein/RNA complex named spliceosome [11]. High-throughput proteomics led to the identification of 292 proteins associated to spliceosomes assembled *in vitro* [12,13]. Several new proteins were identified, including 20 related to transcription, suggesting a direct physical connection between this process and splicing.

The nucleolus is defined as the site of rRNA synthesis and assembly of the ribosomal subunits. Proteomics analysis led to the identification of 271 gene products from a HeLa cell preparation of nucleoli, of which more then 30% were previously uncharacterized or unknown [14], and expression of fluorescently tagged proteins confirmed nucleolar localization for at least 18 of them. Of interest, results suggested that a dynamic reorganization of the nucleolus proteome could occur in response to the metabolic state of the cell. Moreover, one of the new proteins identified, paraspeckle protein 1 (PSP1), was shown to shuttle between the nucleolus and a novel location within the nucleus, named paraspeckle [15], suggesting its involvement in a complex spatiotemporal regulation of transcription within the nucleus [16].

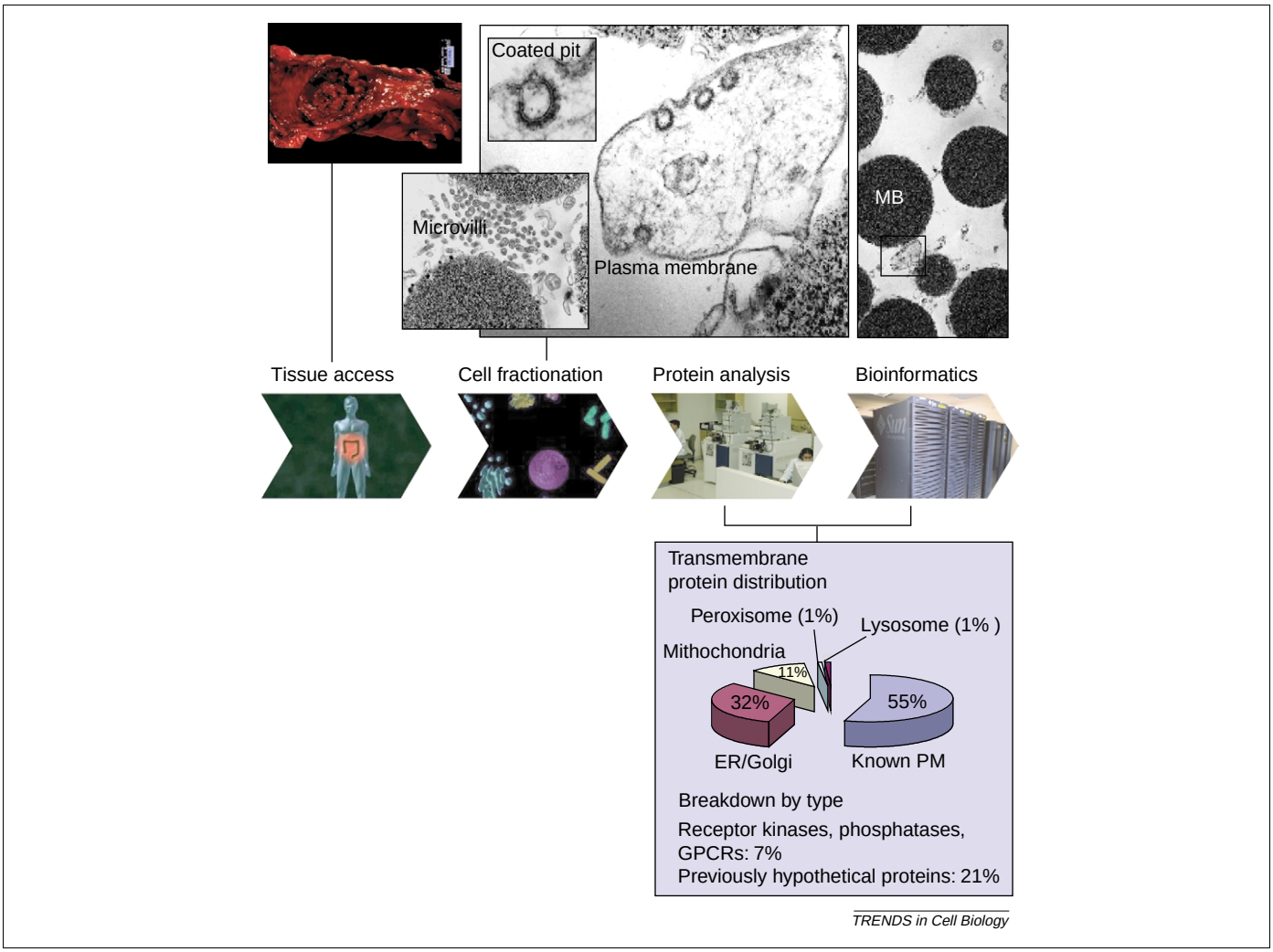

Figure 2. High-throughput proteomics platform. Proteomics analyses actively challenge and reshape several concepts in cell biology and diseases. A key aspect for this success is the integration of several steps in a seamless process enabling the identification of key proteins. Although the current technology might not be suitable for the meaningful analysis of complex samples such as cells or tissues, study of the cell subcompartments, the organelles, provides a unique way to access and identify low abundance proteins in a biological sample. This example illustrates the high-throughput platform used by Caprion Pharmaceuticals for the analysis of plasma membrane samples purified by immunoisolation from human tumor biopsies. The immunoisolation procedure on magnetic beads (MB) allows for the isolation of highly enriched fractions displaying the typical morphology of the plasma membrane with microvilli and coated pits. A single analysis in the mass spectrometer enables the identification of thousands of peptides, leading to the formal identification of several hundred proteins. Thorough bioinformatics analyses are required to annotate and help make sense of terabytes of data generated in a typical high-throughput analysis. ER, endoplasmic reticulum; GPCRs, G-protein-coupled receptor; PM, plasma membrane.

The endocytic apparatus: clathrin-coated vesicles and lysosomes

The endocytic pathway has been extensively studied. Surprisingly, only a few studies have taken advantage of proteomic approaches to identify proteins associated to the various organelles of this pathway, despite established protocols for their isolation. Nevertheless, information has been gathered using organelle proteomics, allowing the identification of new proteins and the refinement of our understanding of the molecular mechanisms involved in endocytosis. Endocytosis through clathrin-coated vesicles is a major pathway by which proteins and membrane components are internalized. Mass spectrometric analysis of purified clathrin-coated vesicles led to the identification of around 250 proteins, of which 10 were novel proteins [17]. Further characterization of these novel proteins allowed for the identification of a novel gene product displaying an epsin N-terminal homology (ENTH) domain, a membrane-interacting module found in several endocytic proteins [18]. This protein, named enthoprotin, was found to interact with the clathrin adaptor protein AP1 and, through its C-terminal, to the clathrin heavy chain, to stimulate clathrin assembly [17].

Lysosomes, considered for many years as the end compartment of the endocytic pathway, were among the first organelles to be described by cell biologists [19]. Their electron-microscopic appearance as dense membrane-bound vacuoles staining for acid phosphatase, used for their identification, rapidly drew attention to that compartment. The properties of lysosomes were first discussed at length at the first international symposium held by the Ciba Foundation in 1963 [5]. By virtue of their centrifugal properties, lysosomes were isolated and found to be associated with a variety of acid hydrolases. Several more hydrolases were found in lysosomes, raising their number to ~50 [19,20], but it took many years to begin identifying some of the membrane proteins associated with these organelles. The first proteomics analysis of lysosomes focused on the identification of new hydrolases, isolated on the basis of their affinity for the mannose 6-phosphate

Box 1. Mass spectrometry-based proteomics for system biology

The late 1980s witnessed an unparalleled expansion of mass spectrometry (MS) into protein chemistry, and the technological advances that led to this contribution played a key role in the numerous large-scale proteomics efforts that are currently under way. The discovery and development of matrix-assisted laser desorption ionization (MALDI) [64,65] and electrospray MS [66] (recently recognized by the 2002 Nobel Prize in Chemistry) have enabled the identification and quantitation of trace levels of proteins from complex cellular extracts, thereby providing indispensable structural tools to biochemists and cell biologists.

Three MS-based proteomics approaches have emerged to face the overwhelming complexity of the protein make-up of the cell and the extensive diversity of protein variants, isoforms and modifications (glycosylation, sulfation, phosphorylation, acylation, etc.). All three approaches involve gel electrophoresis and/or liquid chromatography before MS analysis (for a review, see [67]). 2-D gels have been the hallmark of proteomics research for several decades [68] to compare protein expression across different cell extracts, whereas identification of protein spots digested within the gel is obtained using MALDI MS and mass fingerprinting. The application of this technique must be balanced with its current limitations with respect to dynamic range, the under-representation of membrane proteins and proteins of high and low molecular weight or extreme isoelectric point [69]. A second approach involves protein separation by sodium dodecyl sulfate–polyacrylamide gel electrophoresis, excision of gel slices, tryptic digestion and analysis of the resulting peptides using capillary liquid chromatography–electrospray MS. Gel-based protein identification places the lower boundary of identification at ~5–50 femtomoles. A third approach avoid gels altogether and involves in-solution protein digestion followed by 2-D liquid chromatography (strong cation-exchange and reverse-phase C_{18} chromatography) coupled to electrospray MS [70].

Peptide ions are sequenced through collisional activation of MS-selected tryptic peptide ions (detected in the survey scan), a process that gives rise to cleavage of the amide bonds and formation of discrete sequence ions that retain the charge at the C-terminus (y-type) or N-terminus (b-type) of the peptide backbone (Figure I**a**). Comprehensive protein identification is typically made using product ion spectra obtained through iterative exclusion of previously acquired ions (Figure I**b**). Such an approach can be both sample- and time-consuming given that gel bands typically contain several tens of proteins with more than 2000 peptide ions, including native tryptic peptides, unspecific cleavages and multiple peptide protonation (same peptide resulting in different *m*/*z* values). A more sample-efficient approach involves the identification of differentially abundant proteins from related cell extracts using an inclusion list of selected precursor ions, as shown in (b) for normal and tumor colon samples.

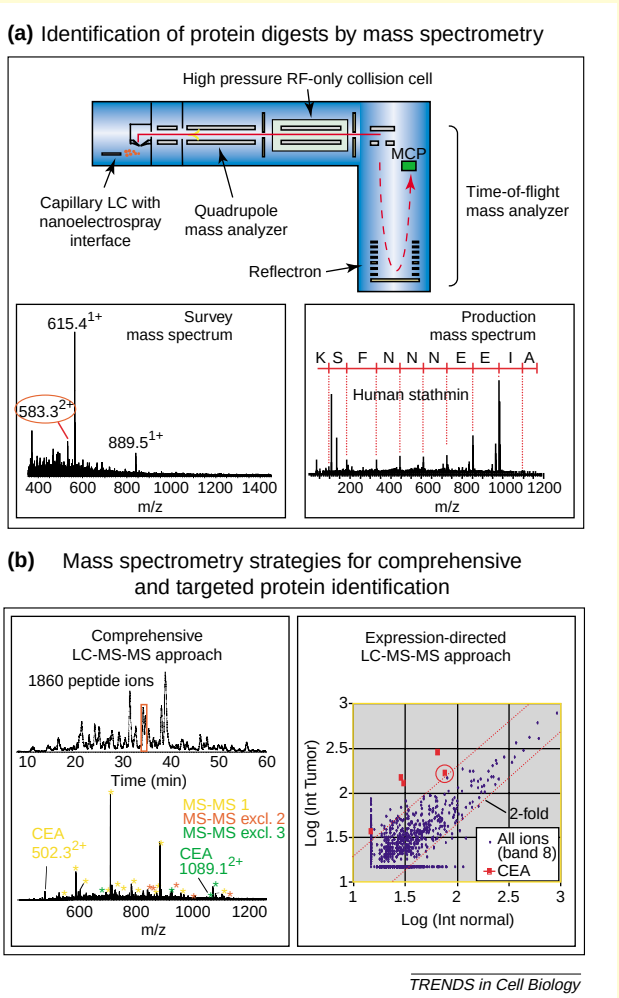

Figure I.

receptor involved in their exit from the Golgi apparatus and their targeting to endocytic compartments [20,21]. This approach led to the identification of cystatin F, a cysteine proteinase inhibitor [22] that could be involved in the control of cysteine peptidase activity known to be essential for antigen presentation in various blood cell lineages [23], and CREG (cellular repressor of E1A-stimulated genes), a secreted protein that might participate in a signaling cascade important for the differentiation of pluripotent stem cells [24].

Using a 2-D gel-based approach, Huber's group compared the protein profiles of early and late endosomes and found that several protein spots were selectively enriched in late endosomes [25]. One of the highly enriched proteins turned out to be a novel 14-kDa protein (p14) associated to the cytoplasmic side of late endosomes and lysosomes. p14 was shown to function as an adaptor protein essential for the assembly on endosomes of the MP1–MAPK scaffold complex crucial for signal transduction [26]. These findings suggested that, in addition to the direct modulation of the enzymatic activity of kinases, the localization of signaling units might exert a regulatory function to signal transduction. Proteomics analysis of purified lysosomes also led to the identification of protein of the γ-secretase complex [27], providing an interesting new twist to the search for the biogenesis of Alzheimer's disease.

The exosome

Although exosomes were first observed, while following the fate of internalized transferrin receptors, as small vesicles present within multivesicular bodies bearing the receptor at their external surface [28], the term 'exosome' was first used to describe vesicles released by reticulocytes during their maturation into red blood cells [29]. Exosomes are generated by the invagination of membrane vesicles within maturing late endocytic organelles. They are secreted by a variety of cells after fusion of late endocytic multivesicular bodies with the plasma membrane [30]. The presence of major histocompatibility complex (MHC) class II molecules on the exosome membrane and the

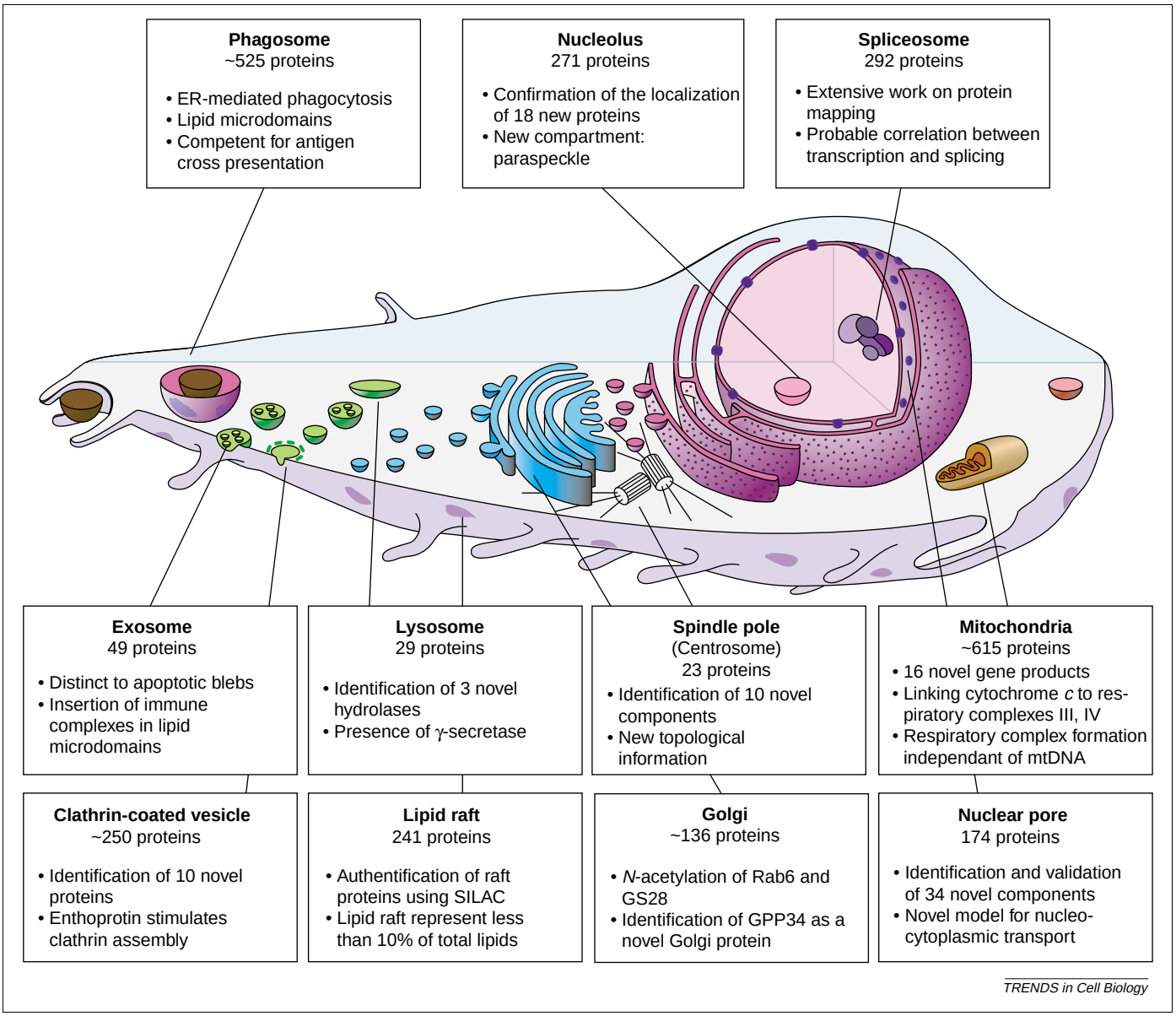

Figure 3. The cell map. Several studies using organelle proteomics have led to a better understanding of the molecular mechanisms involved in the regulation of the function of various cellular organelles. Values were extracted from the following references: phagosome [48,49,50,57; M. Desjardins, unpublished]; nucleolus [14–16]; spliceosome [12,13]; exosome [34–36]; lysosome [20,21,27]; spindle pole [58]; mitochondria [59–62]; Clathrin-coated vesicle [17]; lipid raft [63]; Golgi [40–42]; nuclear pore [10]. ER, endoplasmic reticulum; SILAC, stable-isotope labeling with amino acids in cell culture.

ability of exosomes derived from B lymphocytes to induce an antigen-specific T-cell response led to the proposal that exosomes might be involved in antigen presentation [31]. Dendritic cell (DC) exosomes displaying MHC class I and class II molecules can prime a specific cytotoxic T-lymphocyte response that can eradicate or suppress the growth of established murine tumors [32]. These properties are somehow related to another type of microvesicles, termed apoptotic blebs, which are released from apoptotic cells [33]. Proteomics analyses of DC-derived [34,35] and B-cell-derived [36] exosomes led to the identification of about 50 proteins. The presence of several proteins formally identified on endosomes confirmed the endocytic origin of exosomes and the fact that they are biochemically and morphologically distinct from apoptotic blebs [35]. Interestingly, although MHC class II molecules were found to be present in detergent-soluble regions at the plasma membrane, they were present, together with tetraspanins, in insoluble parts of exosomes, suggesting their incorporation in tetraspanin-containing detergent-resistant membrane microdomains during the formation of exosomes [36]. Thorough identification of the molecules present on exosomes is of interest considering the possibility of using these vesicles in the development of cancer vaccines [37].

The Golgi membrane

2-D gels are notoriously difficult for separating and characterizing integral membrane proteins [38]. Therefore, attempts using 1-D gels have been preferred in the characterization of integral membrane proteins by mass spectrometry. Hence, after purification of the hepatic Golgi complex by a recently developed methodology [39], Triton X-114 partitioning, an approach enabling the enrichment of proteins that have affinity for lipids, was used to

separate integral membrane proteins. Using this approach, 81 proteins were characterized by mass spectrometry. In addition, more than 40 sequence tags were identified that did not correspond to sequences found in the database. Remarkably, a prominent protein previously unassigned was characterized as a novel Golgi protein associated with the cytosolic surface of the organelle. Proteomics analyses by two separate groups identified this protein as GMx33/GPP34 [40,41,42]. Furthermore, *N*-acetylation of rab6 [42] and the SNARE GS28 (J.J.M. Bergeron, unpublished) were detected, representing the first evidence for such post-translational modifications of these proteins.

Two approaches have been pursued to assess whether a protein is truly associated with the Golgi membrane. The first involves the use of electron-microscopic immunolocalization either on cryosections of the isolated fraction or on isolated organelles to verify the location of the characterized protein. This has helped provide insight into several proteins the have been elucidated in the proteomics effort, including Golgi-located SNAREs and contaminants such as the endoplasmic reticulum (ER) protein, calnexin [42].

The phagosome

Phagocytosis is the process by which cells internalize large particles for degradation along the endocytic/phagocytic pathways. It plays key roles in the nutrition of unicellular organisms, the handling of apoptotic cells, tissue remodeling and the control of infection by intracellular pathogens [43,44]. Phagocytosis is initiated by the binding of particle ligands to a variety of receptors at the cell surface, enabling the reorganization of cytoskeletal elements and the recruitment of the plasma membrane to form extensions leading to particle engulfment. Newly formed phagosomes interact rapidly with organelles of the biosynthetic and endocytic apparatus, enabling their gradual transformation into phagolysosomes in which proteolytic degradation occurs. The proper maturation of phagosomes is pivotal for antigen processing and presentation at the cell surface, an essential mechanism by which the spread of infection by intracellular pathogens is restricted. Nevertheless, pathogens have evolved ways of avoiding the harsh environment of phagolysosomes, and survive in their hosts by subverting the molecular mechanisms associated with phagosome maturation. The molecular means by which pathogens act on their host cells is still largely unknown, although different mechanisms appear to be used by the various infectious agents [45].

To understand further the function of phagosomes at the molecular level and how they are compromised by intracellular pathogens, we have initiated the proteomics analysis of latex bead-containing phagosomes. The unique advantage of this approach is that the use of low-density latex microparticles enables the isolation of highly enriched phagosome preparations, away from all the other cell organelles, by a single centrifugation step on a sucrose gradient [46,47]. These phagosome preparations display key features of genuine phagosomes/phagolysosomes, such as their fusogenic properties and the high content of hydrolases needed for antigen processing [48]. So far, 520 proteins have been identified in these phagosome preparations using MALDI-TOF MS and nanoelectrospray MS/MS, of which 150 have been published [48,49]. Although several of these proteins (e.g. hydrolases and the proton pump ATPase) were expected in a compartment involved in the killing and degradation of microorganisms, the nature of some of the other proteins, including several proteins of the ER, revealed unexpected features of phagosomes (Figure 4).

Although textbooks indicate that phagosomes originate from invaginations of the plasma membrane (PM), which accordingly contributes most of the membrane of newly formed phagosomes, we were able to show that a significant part of the phagosome membrane is made of ER [50]. This membrane is recruited onto nascent phagosomes containing latex particles, as well as *Salmonella*, *Leishmania* or red blood cells, by a process that appears to involve its direct fusion with the PM, underneath phagocytic cups [50,51]. As surprising as this might have seemed, data in the literature supported the possible fusion of ER with the PM. Indeed, Rothman's group found that vesicles containing the PM t-SNARE Sso1/Sec9c could fuse with vesicles displaying the ER v-SNARE sec22, a process that follows the principles described by the SNARE hypothesis [52].

The surprising finding that ER contributes to phagosome formation raised the question of whether this membrane is used simply to spare the PM or whether it contributes to some of the functional properties of the phagosome [44].

ER-mediated phagocytosis: a new membrane for new functions

As mentioned above, a key aspect of phagocytosis is its role in the degradation and processing of peptides derived from intracellular pathogens for antigen presentation and the elaboration of an efficient adaptive immune response. Peptides derived from intracellular pathogens are normally presented on MHC class II molecules in the phagosome lumen, in contrast to peptides present in the cell cytoplasm, which are normally loaded in the ER and presented at the cell surface on MHC class I molecules [53]. Despite this apparent segregation, several studies have reported that antigens from intracellular pathogens can be presented on MHC class I molecules and trigger a CD8 + T-cell response. Because antigens from intracellular pathogens are normally presented on MHC class II molecules, triggering a CD4 + T-cell response, their presentation on class I molecules is referred to as cross-presentation [54]. Faced with these results, models had to be proposed, within the cellular context, for phagosomes originating from the PM and unrelated to ER. Thus, a process allowing for the transfer of peptides from the phagosome lumen to the cell cytoplasm (for further processing by the proteasome) and then to the ER lumen, where class I molecules are loaded, was proposed [55]. This process would require the retrotranslocation of peptides from the phagosome lumen to the cytoplasm, the ubiquitination and proteosomal degradation of these proteins in the cytoplasm, the translocation of the processed peptides in the ER lumen for loading on MHC class I molecules and the delivery of the loaded peptides to the cell surface (through the secretory pathway). Proteomics analysis of

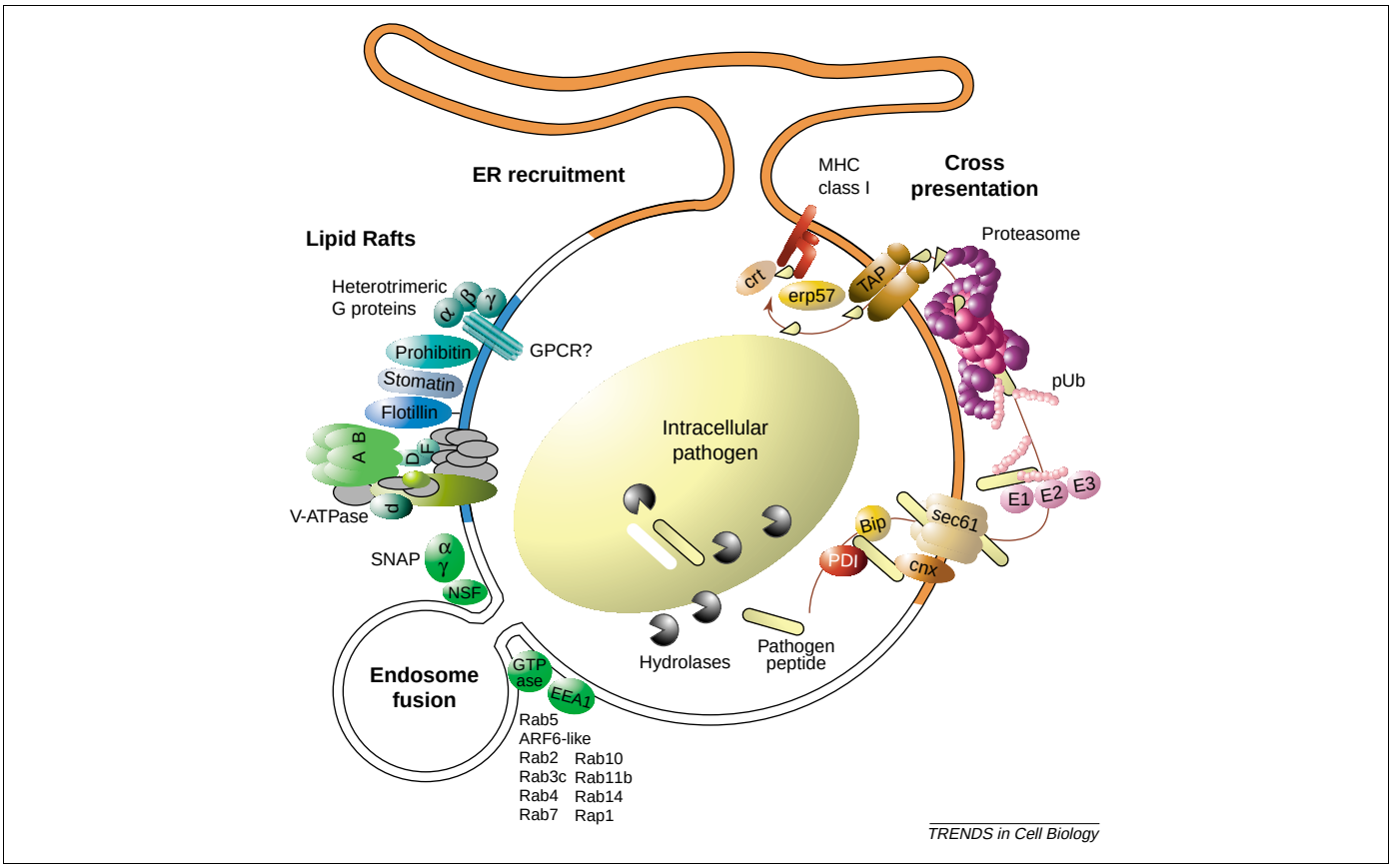

Figure 4. A new face for phagosomes. Proteomics analyses have identified over 500 proteins present in highly enriched phagosome preparations. Three new concepts in the biology of phagosomes validated by thorough cell biology studies are presented. The formation of phagosomes at the cell surface was shown to involve the direct recruitment of the endoplasmic reticulum [50]. This process was further shown to confer phagosomes with new functional properties, including the ability to process exogenous peptides for antigen cross-presentation [49]. Accordingly, our model proposes that peptides from intracellular pathogens would be generated by the direct and sequential action of various hydrolases in the phagosome lumen. These peptides would then be retrotranslocated to the cytoplasmic side of phagosomes by a chaperone-aided process involving sec61. The proteosomal degradation machinery, including ligases and proteasomes, would then associate transiently to phagosomes to further process the retrotranslocated peptides into shorter peptides that would then be translocated back to the phagosome lumen where loading on MHC class I molecules would occur. Lastly, proteomics analyses led to the characterization of lipid microdomains on the phagosome membrane, indicating that specialized functions, such as membrane fusion, might occur of specific foci of the phagosome membrane [57]. GPCRs, G-protein-coupled receptor.

latex-bead-containing phagosomes shed some light on the possible mechanisms involved in cross-presentation. MS/MS analysis indicated that proteins involved in each of the steps of the cross-presentation pathway are present on purified phagosomes [49]. Detailed studies confirmed that phagosomes are competent organelles for exogenous peptide cross-presentation, and functional for peptide retrotranslocation, ubiquitination and proteosomal degradation, as well as peptide translocation back to the phagosome lumen where loading on MHC class I molecules occurs [49,56].

Concluding remarks

The approach of studying biological processes by focusing on single proteins or small groups of proteins is essential and remains the most efficient way to decipher the tremendous complexity of the molecular mechanisms involved in cell functions. However, it must be kept in mind that, in many cases, the information generated in this way is used to generate models that are necessarily simplistic views of much more elaborate processes. In this respect, several of the models presented in cell biology textbooks can be compared to images obtained using low-resolution digital cameras. The use of high-throughput proteomics will undoubtedly be useful to provide the higher level of resolution needed to generate the detailed, more global, view needed to fully comprehend the nature of subcellular interactions. Current approaches are mainly focusing on identifying proteins of the various organelles. This should pave the way to more integrated studies in which the features of these proteins will be investigated, including their post-translational modification, the level of their expression, the nature of their interacting partners and how all of these features are modulated in various conditions such as health and disease. In some cases, analysis of suborganelle fractions (e.g. lipid microdomains) might be needed to decrease the complexity of the sample further for meaningful mass spectrometry analysis. In most cases, bioinformatics tools will have to be developed to handle the large and intricate sets of data generated by high throughput approaches (Boxes 2 and 3).

However, organelles are not fixed entities but rather dynamic structures interacting with each other and remodeling themselves in response to various stimuli. Accordingly, it is unlikely that a definite fixed proteome can be assigned to any of the cell subcompartments. Analysis of the cell organelles in various conditions will be needed to understand the dynamic nature of integrated

Box 2. The tide is high and rising

Advances in high-throughput biotechnology over the past 15 years (depicted graphically in Figure II) have enabled the generation of large quantities of data. However, the gap between data and knowledge continues to widen. It is estimated that GenBank doubles in size every 2 years, whereas PubMed, a proxy for biological knowledge, doubles only every 20 years (http://www.aeiveos.com/~bradbury/InformationGrowth/index.html and http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html). For example, in a blood serum study (20 patients) of Alzheimer's disease conducted at Caprion Pharmaceuticals, 1 terabyte of mass spectrometry (MS) data were generated in 2 weeks. The task of managing, analyzing and visualizing such volumes of data requires the skills of the bioinformaticist, but even more critical is the participation of the cell biologist who must form and evaluate hypotheses based on these data. In this context, two key aspects of bioinformatics in MS-based proteomics can be highlighted. First, how bioinformatics enables the generation of proteomic data by overcoming technical hurdles. Second, how bioinformatics can enable the cell biologist to interpret proteomic data.

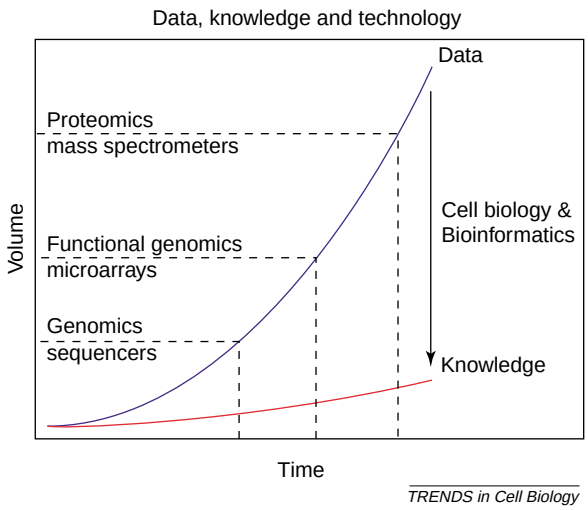

Figure II.

Box 3. Enabling data interpretation

In recent years there have been many advances in bioinformatics tools that enable high-throughput interpretation of proteomic data. These include tools for spectral analysis such as database searching [71] and *de novo* sequencing of novel spectra (http://www.bioinformaticssolutions.com). For high-throughput protein analysis, tools such as Gene Ontology [72], InterPro [73], homology clustering and genome alignment permit the cell biologist to organize thousands of proteins, automatically, into meaningful groups based on protein function, cellular process and cellular location. Furthermore, the clustering of previously sequenced and new peptide sequences along the genome permits the detection of novel exons and splice variants. Lastly, integrated tools for expression analysis such as that depicted in Figure III permit the investigator to visualize an entire study. In the top window, peptide maps from 10 normal (blue) and 10 disease (red) patients are displayed; in the bottom left window, the gel distribution of selected peptides is displayed across patients; and in the bottom-right window a scatterplot displays differential expression. With such tools, the investigator is able to relate peptide reproducibility, molecular weight and differential expression.

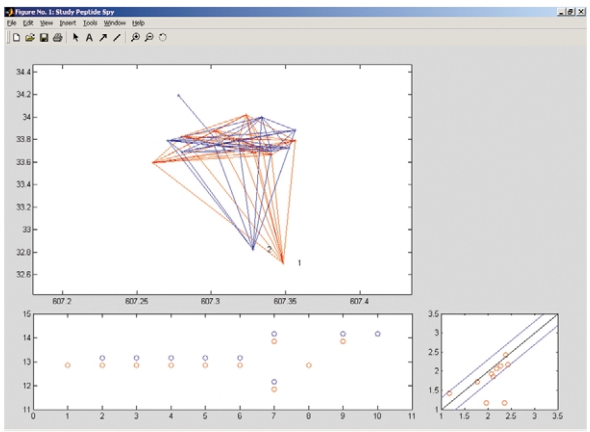

Figure III.

cell functions. Indeed, who could have imagined that from a single preparation routinely used to generate one data point (e.g. the presence of a protein by western blot), scientists can identify several hundred proteins, measure their level of expression and evaluate some of their post-translational features? Despite recent successes, proteomics is not the Holy Grail. Any identified protein remains a name on a list, or at best a potential resident marker. Because the concept of 'pure' organelle does not exist, demonstrating the presence of these proteins on the organelle by other approaches, such as immunofluorescence or tagged expression, is still an important step. Clearly, considering the large number of proteins that can now be identified in any complex sample, each of them cannot be validated individually. However, a degree of confidence builds up as more and more proteins are validated. Hopefully, proteomics analysis will soon become a tool available to all cell biologists, allowing the emergence of high-resolution models of cell function in health and disease.

References

1 Jung, E. *et al.* (2000) Proteomics meets cell biology: the establishment of subcellular proteomes. *Electrophoresis* 21, 3369–3377
2 Huber, L.A. (2003) Is proteomics heading in the wrong direction? *Nat. Rev. Mol. Cell Biol.* 4, 74–80
3 Patterson, S.D. and Aebersold, R.H. (2003) Proteomics: the first decade and beyond. *Nat. Genet.* 33 (Suppl.), 311–323
4 Palade, G.E. (1966) Structure and function at the cellular level. *JAMA.* 198, 815–825
5 De Duve, C. (1963) The Lysosome concept. In *Ciba Foundation Symposium* (de Reuck, A.V.S. and Cameron, M.P., eds), pp. 1–35, Churchill Livingstone
6 Huber, L.A. *et al.* (2003) Organelle proteomics: implications for subcellular fractionation in proteomics. *Circ. Res.* 92, 962–968
7 Blackstock, W.P. and Weir, M.P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17, 121–127
8 Gadal, O. and Nehrbass, U. (2002) Nuclear structure and intranuclear retention of premature RNAs. *J. Struct. Biol.* 140, 140–146
9 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
10 Rout, M.P. *et al.* (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* 148, 635–651
11 Staley, J.P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315–326
12 Neubauer, G. *et al.* (1998) Mass spectrometry and EST-database

searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.* 20, 46–50
13 Rappsilber, J. *et al.* (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res.* 12, 1231–1245
14 Andersen, J.S. *et al.* (2002) Directed proteomic analysis of the human nucleolus. *Curr. Biol.* 12, 1–11
15 Fox, A.H. *et al.* (2002) Paraspeckles: a novel nuclear domain. *Curr. Biol.* 12, 13–25
16 Dundr, M. and Misteli, T. (2002) Nucleolomics: an inventory of the nucleolus. *Mol. Cell.* 9, 5–7
17 Wasiak, S. *et al.* (2002) Enthoprotin: a novel clathrin-associated protein identified through subcellular proteomics. *J. Cell Biol.* 158, 855–862
18 De Camilli, P. *et al.* (2002) The ENTH domain. *FEBS Lett.* 513, 11–18
19 de Duve, C. (1983) Lysosomes revisited. *Eur. J. Biochem.* 137, 391–397
20 Journet, A. *et al.* (2002) Proteomic analysis of human lysosomes: application to monocytic and breast cancer cells. *Proteomics* 2, 1026–1040
21 Journet, A. *et al.* (2000) Towards a human repertoire of monocytic lysosomal proteins. *Electrophoresis* 21, 3411–3419
22 Ni, J. *et al.* (1998) Cystatin F is a glycosylated human low molecular weight cysteine proteinase inhibitor. *J. Biol. Chem.* 273, 24797–24804
23 Nathanson, C.M. *et al.* (2002) Regulated expression and intracellular localization of cystatin F in human U937 cells. *Eur. J. Biochem.* 269, 5502–5511
24 Veal, E. *et al.* (2000) The secreted glycoprotein CREG enhances differentiation of NTERA-2 human embryonal carcinoma cells. *Oncogene* 19, 2120–2128
25 Wunderlich, W. *et al.* (2001) A novel 14-kilodalton protein interacts with the mitogen-activated protein kinase scaffold mp1 on a late endosomal/lysosomal compartment. *J. Cell Biol.* 152, 765–776
26 Teis, D. *et al.* (2002) Localization of the MP1–MAPK scaffold complex to endosomes is mediated by p14 and required for signal transduction. *Dev. Cell* 3, 803–814
27 Pasternak, S.H. *et al.* (2003) Presenilin, nicastrin and beta-amyloid production in the endosomal/ lysosomal system: toward a unifying model of Alzheimer's disease. *J. Biol. Chem.* (in press)
28 Pan, B.T. *et al.* (1985) Electron microscopic evidence for externalization of the transferrin receptor in vesicular form in sheep reticulocytes. *J. Cell Biol.* 101, 942–948
29 Johnstone, R.M. *et al.* (1987) Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). *J. Biol. Chem.* 262, 9412–9420
30 Denzer, K. *et al.* (2000) Exosome: from internal vesicle of the multivesicular body to intercellular signaling device. *J. Cell Sci.* 113, 3365–3374
31 Raposo, G. *et al.* (1996) B lymphocytes secrete antigen-presenting vesicles. *J. Exp. Med.* 183, 1161–1172
32 Zitvogel, L. *et al.* (1998) Eradication of established murine tumors using a novel cell-free vaccine: dendritic cell-derived exosomes. *Nat. Med.* 4, 594–600
33 Casciola-Rosen, L.A. *et al.* (1994) Autoantigens targeted in systemic lupus erythematosus are clustered in two populations of surface structures on apoptotic keratinocytes. *J. Exp. Med.* 179, 1317–1330
34 Thery, C. *et al.* (1999) Molecular characterization of dendritic cell-derived exosomes. Selective accumulation of the heat shock protein hsc73. *J. Cell Biol.* 147, 599–610
35 Thery, C. *et al.* (2001) Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles. *J. Immunol.* 166, 7309–7318
36 Wubbolts, R. *et al.* (2003) Proteomic and biochemical analyses of human B cell-derived exosomes. Potential implications for their function and multivesicular body formation. *J. Biol. Chem.* 278, 10963–10972
37 Wolfers, J. *et al.* (2001) Tumor-derived exosomes are a source of shared tumor rejection antigens for CTL cross-priming. *Nat. Med.* 7, 297–303
38 Santoni, V. *et al.* (2000) Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. *Electrophoresis* 21, 3329–3344
39 Dominguez, M. *et al.* (1999) Fusogenic domains of golgi membranes are sequestered into specialized regions of the stack that can be released by mechanical fragmentation. *J. Cell Biol.* 145, 673–688
40 Wu, C.C. *et al.* (2000) Proteomic analysis of two functional states of the Golgi complex in mammary epithelial cells. *Traffic* 1, 769–782
41 Wu, C.C. *et al.* (2000) GMx33: a novel family of trans-Golgi proteins identified by proteomics. *Traffic* 1, 963–975
42 Bell, A. *et al.* (2001) Proteomics characterization of abundant Golgi membrane proteins. *J. Biol. Chem.* 276, 5152–5165
43 Underhill, D.M. and Ozinsky, A. (2002) Phagocytosis of microbes: complexity in action. *Annu. Rev. Immunol.* 20, 825–852
44 Desjardins, M. (2003) ER-mediated phagocytosis: a new membrane for new functions. *Nat. Rev. Immunol.* 3, 280–291
45 Duclos, S. and Desjardins, M. (2000) Subversion of a young phagosome: the survival strategies of intracellular pathogens. *Cell. Microbiol.* 2, 365–377
46 Wetzel, M.G. and Korn, E.D. (1969) Phagocytosis of latex beads by *Acanthamoeba castellanii* (Neff). 3. Isolation of the phagocytic vesicles and their membranes. *J. Cell Biol.* 43, 90–104
47 Desjardins, M. *et al.* (1994) Biogenesis of phagolysosomes proceeds through a sequential series of interactions with the endocytic apparatus. *J. Cell Biol.* 124, 677–688
48 Garin, J. *et al.* (2001) The phagosome proteome: insight into phagosome functions. *J. Cell Biol.* 152, 165–180
49 Houde, M. *et al.* (2003) Phagosomes are competent organelles for antigen cross-presentation. *Nature* 425, 402–406
50 Gagnon, E. *et al.* (2002) Endoplasmic reticulum-mediated phagocytosis is a mechanism of entry into macrophages. *Cell* 110, 119–131
51 Muller-Taubenberger, A. *et al.* (2001) Calreticulin and calnexin in the endoplasmic reticulum are important for phagocytosis. *EMBO J.* 20, 6772–6782
52 McNew, J.A. *et al.* (2000) Compartmental specificity of cellular membrane fusion encoded in SNARE proteins. *Nature* 407, 153–159
53 Watts, C. and Amigorena, S. (2001) Phagocytosis and antigen presentation. *Semin. Immunol.* 13, 373–379
54 Heath, W.R. and Carbone, F.R. (2001) Cross-presentation in viral immunity and self-tolerance. *Nat. Rev. Immunol.* 1, 126–134
55 Kovacsovics-Bankowski, M. and Rock, K.L. (1995) A phagosome-to-cytosol pathway for exogenous antigens presented on MHC class I molecules. *Science* 267, 243–246
56 Guermonprez, P. *et al.* (2003) ER–phagosome fusion defines an MHC class 1 cross-presentation compartment in dendritic cells. *Nature* 425, 397–402
57 Dermine, J.F. *et al.* (2001) Flotillin-1-enriched lipid raft domains accumulate on maturing phagosomes. *J. Biol. Chem.* 276, 18507–18512
58 Wigge, P.A. *et al.* (1998) Analysis of the *Saccharomyces* spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. *J. Cell Biol.* 141, 967–977
59 Lopez, M.F. *et al.* (2000) High-throughput profiling of the mitochondrial proteome using affinity fractionation and automation. *Electrophoresis* 21, 3427–3440
60 Brookes, P.S. *et al.* (2002) High-throughput two-dimensional blue-native electrophoresis: a tool for functional proteomics of mitochondria and signalling complexes. *Proteomics* 2, 969–977
61 Taylor, S.W. *et al.* (2003) Characterization of the human heart mitochondrial proteome. *Nat. Biotechnol.* 21, 281–286
62 Fountoulakis, M. and Schlaeger, E.J. (2003) The mitochondrial proteins of the neuroblastoma cell line IMR-32. *Electrophoresis* 24, 260–275
63 Foster, J.L. *et al.* (2003) Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5813–5818
64 Karas, M. and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular mass exceeding 10000 Daltons. *Anal. Chem.* 60, 2299–2301
65 Tanaka, K. *et al.* (1988) Protein and polymer analyses up to *m/z* 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2, 151–153
66 Fenn, J.B. *et al.* (1989) Electrospray ionization for the mass spectrometry of large biomolecules. *Science* 246, 64–71
67 Aebersold, R. and Mann, M. (2003) Mass-spectrometry-based proteomics. *Nature* 422, 198–207
68 O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250, 4007–4021
69 Gygi, S.P. *et al.* (2000) Evaluation of two-dimensional gel

electrophoresis-based proteome technology. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9390–9395
70 Wolters, D.A. *et al.* (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* 73, 5683–5690
71 Perkins, D.N. *et al.* (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567
72 Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29
73 Mulder, N.J. *et al.* (2003) The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res.* 31, 315–318