Quantitative proteomics of Saccharomyces cerevisiae vacuoles and stress responses in Lactococcus lactis
Wiederhold, Elena

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
The little monk: I understand your bitterness. You are thinking of certain exceptional powers (...).
Galileo: Just say instrument of torture.
Abstract

Knowledge of the subcellular localization of proteins is indispensable to understand their physiological roles. In the past decade at least eighteen studies have been performed to analyze the protein content of isolated organelles from *Saccharomyces cerevisiae*. Here, we integrate the data sets and compare them with other large scale studies on protein localization and abundance. We evaluate the completeness and reliability of the organelle proteomics studies. Reliability depends on the purity of the organelle preparations, which unavoidably contain (small) amounts of contaminants from different locations. Quantitative proteomics methods can be used to distinguish between true organelar constituents and contaminants. Completeness is compromised when loosely or dynamically associated proteins are lost during organelle preparation, and also depends on the sensitivity of the analytical methods for protein detection. There is a clear trend in the data from the eighteen organelle proteomics studies showing that proteins of low abundance frequently escape detection. Proteins with unknown function or cellular abundance are also infrequently detected, indicating that these proteins may not be expressed under the conditions used. We discuss that the yeast organelle proteomics studies provide powerful lead data for further detailed studies, and that methodological advances in organelle preparation and in protein detection may help to improve the completeness and reliability of the data.
V. CHAPTER: Proteomics of *S. cerevisiae* organelles

**Introduction**

The yeast *Saccharomyces cerevisiae* is a widely used monocellular eukaryotic model organism. *S. cerevisiae* was the first eukaryotic organism of which the complete genome was sequenced (Goffeau et al., 1996). The genome contains 6,608 open reading frames, of which 5,797 encode polypeptides. 4,666 proteins have annotated functions in the yeast genome database SGD.

*S. cerevisiae* has intracellular compartments typical for eukaryotic cells (Table 1). The subcellular localization of proteins in yeast has been investigated using genome-wide chromosomal GFP-tagging approaches (Kumar et al., 200; Huh et al., 2003), in addition to immuno-EM and other low-throughput methods. The genome-wide studies have provided an unprecedented wealth of information. Yet, for ~17% of the 5,797 gene products (939 proteins), many of which have unknown biological function, the subcellular localization has not been assigned. In addition, tagging approaches can result in mis-localizations. Consequently, complementary proteome-wide studies, which avoid tagging, are required. Mass spectrometry (MS)-based proteomics, combined with subcellular fractionation, provides such alternative methodology.

Yeast is uniquely suitable for this approach, because of its relatively low complexity, infrequent occurrence of gene isoforms, and limited extent of posttranslational modifications. In the past decade, eighteen studies of the protein content of isolated organelles by means of MS-based approaches (organelle proteomics) have been performed, covering all the subcellular compartments with the exceptions of the ER, the endosome and the cytosol (Table 2, and Athenstaedt et al., 1999; Schäfer et al., 2001; Navarre et al., 2002; Pflieger et al., 2002; Yi et al., 2002; Sickmann et al., 2003; Marelli et al., 2004; Ohlmeier et al., 2004; Prokisch et al., 2004; Inadome et al., 2005; Yin et al., 2005; Delom et al., 2006; Reinders et al., 2006; Zahedi et al., 2006; Sarry et al., 2007; Gauci et al.; Mosley et al., 2009; Wiederhold et al., 2009).

The protein composition of supramolecular assemblies from yeast, such as the nuclear pore complex, and the replication, transcription and translation machineries (Link et al., 1999; Rout et al., 2000; Alber et al., 2007a; Alber et al., 2007b) will not be discussed here, but the experimental design and the challenges are very similar to organelle proteomics.

In this review, we will give an overview of results of the organelle proteome studies, and try to answer the questions: How complete is our current view of organelle proteomes? And how reliable are the proteomics data? Completeness and reliability of the organelle proteome data are dictated to a large extent by methodology. Although the different studies have used a plethora of methods (Table 2), two steps are in common for all published organelle proteome analyses. First, pure organelles are isolated by biochemical fractionation methods. Second, the proteins in the purified organelles are identified. The reliability of organelle proteome analyses ultimately depends on the purity of the isolated organelles. In the ideal case no proteins from different cellular locations (contaminants) should be present in the final preparation. Defining whether or not an identified protein is a contaminant is challenging as proteins may have multiple destinations in a cell, and the distribution over the different localizations may vary throughout the cell cycle and as a function of metabolic and environmental conditions. It is also important to realize that within a single culture, cells of different age and cell cycle stages are present. Thus organelle proteomics studies report ‘averaged’ proteomes. The completeness of the organelle proteomes also depends on the quality of the organelle preparation, because loosely attached proteins may be lost during the isolation. In addition, completeness depends on the sensitivity of the analytical methods that are used to identify the proteins. These methods include extraction of the proteins from the organelle preparation, generation and separation of peptides, and (tandem) mass spectrometry analysis to identify the proteins.

Because organelle proteomics is so intimately dependent on biochemical methods for subcellular fractionation, we will give a brief overview of the methodology to isolate subcellular compartments. We will not discuss the methods for protein extraction, peptide generation and identification using
tandem MS-based techniques. The analytical methods continue to evolve at great pace because of ongoing technical advances. Consequently, the methodological details vary considerably between the different studies that were performed over a period of a decade. Recently, state of the art methods in proteomics were summarized in a comprehensive review by Speers & Wu (Speers & Wu, 2007). Mass spectrometry-based proteomics method for organellar proteomics have been reviewed by Andersen and Mann (Andersen & Mann, 2006) and Yates III et al., (Yates III et al., 2005). Methods dedicated to membrane proteomics have been reviewed by Wu et al., (Wu et al., 2003; Wu & Yates, 2003) and recent instrumental developments in mass spectrometry for proteomics have been summarized in Han et al., (Han et al., 2008).

**Isolating organelles**

All methods used to prepare different organelles have several steps in common: Cells are cultivated, disrupted and the cellular content including organelles is liberated. Then, the released organelles are separated into different populations according to their physical or chemical properties. In Table 2 details for the organelle isolation in the 18 proteomics studies are given. Before integrating the results of the 18 organellar proteomics studies into a global view of protein localization in yeast, we will discuss a few experimental details of the isolation procedures which may affect the ensemble interpretation.

**Considerations about growth conditions**

The content and/or the abundance of organelle proteomes in yeast depends on the growth conditions. Care must be taken when integrating the results from different studies to obtain a global view of protein localization because different growth conditions have been used. It also must be noted that different strains of *S. cerevisiae* have been used in the various organellar proteomics studies (Table 2).

For most organelle proteome analyses, yeast cultures were grown aerobically at 30°C on a rich medium (YPD, 1% yeast extract, 2% Bacto-peptone, 2% glucose) containing an excess of glucose as carbon source, and the cells were harvested in the mid-logarithmic phase. These are 'standard' growth conditions for *S. cerevisiae*, but some organelles require specific growth conditions in order to be obtained in sufficient amounts (Table 2 and discussed below).

For the proteomics studies of peroxisomes cells were grown in conditions that stimulate fatty acid ß-oxidation (Schäfer et al., 2001; Yi et al., 2002; Marelli et al., 2004), which leads to an increase in the number of peroxisomes from 1-4 per cell (under standard conditions) to as many as 14 per cell, occupying 8-10% of the cytoplasmic volume (Veenhuis et al., 1987). Also mitochondrial morphology and abundance depends on growth conditions (Visser et al., 1995). For the mitochondrial proteome analyses, yeast was grown on the non-fermentable carbon sources in order to activate respiration and induce many small mitochondria (Pflieger et al., 2002; Sickmann et al., 2003; Prokisch et al., 2004; Reinders et al., 2006; Zahedi et al., 2006). When grown on glucose-containing medium, the abundance of lipid particles increases towards the stationary phase. For the proteome analysis, the cells were grown on glucose and harvested in the late exponential phase (Athenstaedt et al., 1999).

**Considerations about organelle isolation**

When aiming at an inventory of proteins from organelles, a sensitive technique as mass spectrometry will identify not only proteins that are truly located in the organelle of interest, but also contaminating proteins. It is important to realize that, for physical and biological reasons, it is not possible to obtain 100% pure organelle preparations. One of the challenges in organellar proteomics is to obtain as pure as possible preparations of subcellular compartments. On the other hand, genuine organellar residents may have multiple locations and may be hard to discriminate from contaminants. Another challenge is to obtain complete/intact organelles. Due to generally lengthy isolation procedures a fraction of the true resident proteins may be lost because they have a more dynamic or low affinity association with the organelle or beca-
### Table 1. Physical Properties of Yeast Organelles

Used abbreviations: +/-Carb, with and without carbohydrates; DAPI, 4',6-diamidino-2-phenylindole; DASPMI: dimethyl-aminostyryl-methylpyridinium iodine; GL, glycerol; IM, Inner membrane; IMS, Intermembrane space; OA, oleic acids; OM, Outer membrane.

<table>
<thead>
<tr>
<th>Organelle</th>
<th>Lipid-to-protein ratio[^i] mg/mg</th>
<th>Density[^ii] g/mL</th>
<th>Marker enzyme</th>
<th>Antibody</th>
<th>Specific fluorescent dye</th>
<th>Range of protein abundances copies per cell[^iii]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mitochondria</td>
<td>0.92 (OM) 0.17 (IM)</td>
<td>1.14-1.21[^iv]</td>
<td>Aconitase, fumarase (matrix); Cytochrome c peroxidase Ccp1 (IMS); Cytochrome c oxidase (IM)</td>
<td>Anti-Porin (OM); Anti-cytochrome c oxidase subunit III (IM)</td>
<td>DAPI (pink-white staining); Rhodamine; DASPMI</td>
<td>10^1-10^5</td>
</tr>
<tr>
<td>Peroxisomes</td>
<td>0.46</td>
<td>1.21-1.23 (GL+OA); 1.19-1.21 (GL) 1.16-1.17 (Nycodenz; OA); 1.19-1.2 (OA)</td>
<td>Catalase; Isocitrate lyase</td>
<td></td>
<td></td>
<td>10^2-10^4</td>
</tr>
<tr>
<td>Lipid particle</td>
<td>16.80</td>
<td>&lt;1.000</td>
<td>Delta(24)-sterol C-methyltransferase Erg6</td>
<td></td>
<td></td>
<td>10^-9-10^-4</td>
</tr>
<tr>
<td>Nucleus</td>
<td>0.11</td>
<td>&lt;1.000</td>
<td>RNA polymerase</td>
<td>Anti-nucleophorins</td>
<td></td>
<td>10^-10</td>
</tr>
<tr>
<td>Cell envelope</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasma membrane</td>
<td>0.65</td>
<td></td>
<td>Vanadate-sensitive ATPase</td>
<td>Anti-PMA1</td>
<td>FM4-64 and derivatives</td>
<td>10^-10</td>
</tr>
<tr>
<td>ER (light microsomes)</td>
<td>0.13</td>
<td>SER: 1.146 RER: 1.17-1.19</td>
<td>NADPH-cytochrome c reductase NCp1[^v]; Dolichol phosphate mannose synthase DPM1[^vi]</td>
<td>Anti-Dpm1 (membrane)</td>
<td></td>
<td>10^2-10^5</td>
</tr>
<tr>
<td>Golgi (heavy microsomes)</td>
<td>0.08</td>
<td>1.248[^vii] (sorbitol)</td>
<td>Cs: ER-to-Golgi t-SNARE Sed5; α-1,6-Mannosyl-transferase Och1; Medial/trans: α-1,3-Mannosyl-transferase Mnt1; Trans/TGN: α-1,2-Mannosyl-transferase Kex2; serine carboxypeptidase Kex1; Unclassified: α-glucane synthase; sorting receptor for HDEL-tagged proteins End2</td>
<td>Anti-Vps10 (late Golgi membrane)</td>
<td></td>
<td>10^-1-10^-5</td>
</tr>
<tr>
<td>Endosome</td>
<td>1.17[^viii] 1.12-1.16[^ix] (Nycodenz)</td>
<td></td>
<td>Anti-Pep12 (membrane)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacuole</td>
<td>0.58</td>
<td>1.030-1.075[^x] 1.011 (ficolli)</td>
<td>α-mannosidase AMS1 (membrane); Proteinase A PEP4; Proteinase B PRB1 (lumen)</td>
<td>Anti-alkaline phosphatase Pho8 (membrane); Anti-CPY (lumen); Anti-ATPase subunits α, β (membrane)</td>
<td>Neutral red; FM4-64 (vacuolar membrane)</td>
<td>10^-1-10^-5</td>
</tr>
</tbody>
</table>

---

[^i]: The lipid-to-protein ratio describes the sum of phospholipids and sterols (Zinser et al., 1991; Zinser et al., 1993). In case of heavy microsomal fraction, the phospholipid-to-protein ratio was taken from Kuchler et al., 1986.

[^ii]: The apparent density of organelles depends on the separation medium. Unless stated otherwise, the reported densities are for organelles in sucrose solution.

[^iii]: Although the enzyme is often used as ER marker, it is also present at the mitochondrial outer membrane (Braell, 1988).

[^iv]: The lipid-to-protein ratio describes the sum of phospholipids and sterols (Zinser et al., 1991; Zinser et al., 1993). In case of heavy microsomal fraction, the phospholipid-to-protein ratio was taken from Kuchler et al., 1986.

[^v]: The apparent density of organelles depends on the separation medium. Unless stated otherwise, the reported densities are for organelles in sucrose solution.

[^vi]: Although the enzyme is often used as ER marker, it is also present at the mitochondrial outer membrane (Braell, 1988).

[^vii]: The lipid-to-protein ratio describes the sum of phospholipids and sterols (Zinser et al., 1991; Zinser et al., 1993). In case of heavy microsomal fraction, the phospholipid-to-protein ratio was taken from Kuchler et al., 1986.

[^viii]: The apparent density of organelles depends on the separation medium. Unless stated otherwise, the reported densities are for organelles in sucrose solution.

[^ix]: Although the enzyme is often used as ER marker, it is also present at the mitochondrial outer membrane (Braell, 1988).

[^x]: The lipid-to-protein ratio describes the sum of phospholipids and sterols (Zinser et al., 1991; Zinser et al., 1993). In case of heavy microsomal fraction, the phospholipid-to-protein ratio was taken from Kuchler et al., 1986.
### Table 2. SUMMARY OF THE EIGHTEEN EXPERIMENTAL PROTEOMES DISCUSSED IN THE REVIEW

Used abbreviations: 2-DE, 2-dimensional gel electrophoresis; ATPase, adenosine nucleoside triphosphatase; BAC-PAGE, N,N'-bisacrylcylycystamine polyacrylamide gel electrophoresis; CHAPS, 3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonic acid; CTAB, cetyl-trimethylammonium bromide; ESI, electron spray ionization; EDTA, ethylenediaminetetraacetic acid; EM, electron microscopy; FTICR, Fourier transform ion cyclotron resonance; GC, gas chromatography; HA-tag, hemagglutinin-tag; ICAT, isotope-coded affinity tagging; IEF, iso-electric focusing; nLC, nano-flow liquid chromatography; LDS-PAGE, lithium dodecyl sulfate polyacrylamide gel electrophoresis; LOPIT, localization of organelle proteins by isotop tagging; MALDI, matrix-assisted laser desorption/ionization; MS/MS, tandem mass spectrometry; OM, outer membrane; PFP, peptide finger print; RP, reverse-phase; SAX, strong anion exchanger; SCX, strong cation exchanger; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TMD, transmembrane domain; WB, Western blot.

<table>
<thead>
<tr>
<th>Organelle</th>
<th>Strain/ Growth conditions</th>
<th>Spheroplasting and lysis</th>
<th>Purification separation enrichment</th>
<th>Stripping</th>
<th>Purity assessment</th>
<th>Proteins and/or peptides separation</th>
<th>MS analysis</th>
<th># of identified proteins (%TMD)</th>
<th>Detected abundance range, copies/cell</th>
<th>Validation</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MITOCHONDRIA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CW04 (n)</td>
<td>YPL 6</td>
<td>Zymolyase; osmotic lysis</td>
<td>Differential centrifugation</td>
<td>EM</td>
<td>SDS-PAGE</td>
<td>nLC-ESI-MSMS</td>
<td>179 (12%)</td>
<td>149-882,841</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>YPH499 (n)</td>
<td>YPG</td>
<td>Zymolyase; mechanical disruption</td>
<td>Differential centrifugation; Sucrose density gradient</td>
<td>Salt, shaving with trypsin</td>
<td>WB</td>
<td>2-DE; SCX-RP- nLC 64; SDS-PAGE 65</td>
<td>MALDI-MS/MS or nLC-ESI-MS/MS</td>
<td>750 (18%)</td>
<td>99-1,255,722</td>
<td>Protein import assay</td>
</tr>
<tr>
<td></td>
<td>BJ1991 (n)</td>
<td>YPD versus YPG</td>
<td>Zymolyase; mechanical disruption</td>
<td>Sucrose density gradient</td>
<td>EDTA</td>
<td>2-DE</td>
<td>PFP</td>
<td>251(6%)</td>
<td>149-1,018,216</td>
<td>Transcriptome</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>W303 (n), S1001 YPL, YPD, SCL and SCD 8</td>
<td>Nycodenz density gradient; FFE; Differential centrifugation</td>
<td>urea, thiourea, CHAPS</td>
<td>EM</td>
<td>RP-LC</td>
<td>ESI-MS/MS; LC-FTICR</td>
<td>546 (17%)</td>
<td>57-1,255,722</td>
<td>Co-enrichment with known marker using WB</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>YPH499 (n)</td>
<td>YPG</td>
<td>Zymolyase; mechanical disruption</td>
<td>Differential centrifugation; Sucrose density gradient</td>
<td>EDTA</td>
<td>WB</td>
<td>2-DE; SAX- or SCX-RP-Le 64; SDS-PAGE</td>
<td>MALDI-MS/MS or nLC-ESI-MS/MS</td>
<td>851 (18%)</td>
<td>99-1,255,722</td>
<td></td>
</tr>
<tr>
<td></td>
<td>YPH499 (n)</td>
<td>YPG</td>
<td>Zymolyase; mechanical disruption</td>
<td>Intact mitochondria as in ref. 2 OM: sucrose density gradient</td>
<td>EDTA trypsin</td>
<td>WB</td>
<td>2D-BAC/SDS; SCX-RP-Le 64; SDS-PAGE</td>
<td>MALDI-MS/MS; nLC-ESI-MS/MS</td>
<td>117 (20%)</td>
<td>147-882,841</td>
<td>Computational and biochemical analysis</td>
</tr>
<tr>
<td><strong>PEROXISOMES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BJ1991 (n)</td>
<td>BY4743 (2n) SCIM 8</td>
<td>Lyticase</td>
<td>Sucrose density gradient</td>
<td>Salt, pH</td>
<td>1D-SDS-PAGE</td>
<td>μLC- and nLC-ESI-MS/MS</td>
<td>45 (18%)</td>
<td>358-1,255,722</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>BJ1991 Δapo1 YNO 8</td>
<td>Zymolyase; mechanical disruption</td>
<td>Differential centrifugation; Nycodenz density gradient</td>
<td>EDTA</td>
<td>μLC</td>
<td>GC-ESI-MS/MS</td>
<td>44 (11%)</td>
<td>238-111,935</td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BY4743 (2n)</td>
<td>Zymolyase; mechanical disruption</td>
<td>Differential centrifugation; Nycodenz density gradient</td>
<td>pH</td>
<td>WB</td>
<td>ICAT-μLC</td>
<td>ESI-MS/MS</td>
<td>72 (21%)</td>
<td>238-268,717</td>
<td>GFP-tagging, mutant phenotype</td>
<td>9</td>
</tr>
<tr>
<td><strong>LIPID PARTICLES</strong></td>
<td>X2180-1A (n) FY1679 (2n) YPD</td>
<td>Zymolyase; osmotic lysis</td>
<td>Ficoll density gradient</td>
<td>Delipidation</td>
<td>Enzymatic assays, EM</td>
<td>1D-SDS-PAGE</td>
<td>nLC-ESI-MS/MS</td>
<td>18 (28%)</td>
<td>1,212-168,876</td>
<td>Lipids and sterols composition in deletion mutants</td>
<td>10</td>
</tr>
</tbody>
</table>

**References:**
1. (Pflieger et al., 2002); 2. (Sickmann et al., 2003); 3. (Ohlmeier et al., 2004); 4. (Prokisch et al., 2004); 5. (Reinders et al., 2006); 6. (Zahedi et al., 2006); 7. (Schäfer et al., 2001); 8. (Yi et al., 2002); 9. (Marelli et al., 2004); 10. (Marelli et al., 2004); 11. (Mosley et al., 2009); 12. (Gauci et al., 2009); 13. (Inadome et al., 2005); 14. (Sarry et al., 2007); 15. (Wiederhold et al., 2009); 16. (Yin et al., 2005); 17. (Navarre et al., 2002); 18. (Delom et al., 2006).
<table>
<thead>
<tr>
<th>Region</th>
<th>Strain</th>
<th>Methodology</th>
<th>Modification</th>
<th>Technique</th>
<th>ESI-MS/MS</th>
<th>Ion Trap MS/MS</th>
<th>OPLS-DA</th>
<th>Ion Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUCLEUS</td>
<td>BY 4741 (n)</td>
<td>Zymolyase; mechanical disruption</td>
<td>DNAse I, heparin;</td>
<td>SCX-RP-nLC</td>
<td>ESI-MS/MS</td>
<td>2698 (11%)</td>
<td>49-1,255,722</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>BY 4743 (n)</td>
<td>Zymolyase, lyticase, Glucanex; mechanical disruption</td>
<td>DNAse I, heparin;</td>
<td>Peptide IEF; SCX-RP-nLC; 1D-SDS-PAGE; MS/MS</td>
<td>ESI-MS/MS</td>
<td>1889 (5%)</td>
<td>41-1,255,722</td>
<td>12</td>
</tr>
<tr>
<td>GOLGI</td>
<td>CjY119 (n)</td>
<td>Lyticase; osmotic shock</td>
<td>Differential centrifugation; Immunoprecipitation</td>
<td>1D-SDS-PAGE; MALDI-MS/MS</td>
<td>52 (48%)</td>
<td>238-82,660</td>
<td>Knock-out mutants analysis; HA-tagging</td>
<td>13</td>
</tr>
<tr>
<td>VACUOLE LUMEN</td>
<td>Sey6210 (n)</td>
<td>Zymolyase; mechanical disruption</td>
<td>Sucrose/sorbitol/ficoll density gradient</td>
<td>Proteinase K; 2-DE</td>
<td>MALDI-MS/MS; nLC-IT-MS/MS; ESI-MS/MS</td>
<td>118 (5%)</td>
<td>172-1,018,216</td>
<td>Internal</td>
</tr>
<tr>
<td>VACUOLE MEMBRANE</td>
<td>W303 (n)</td>
<td>Zymolyase, lyticase, Glucanex; osmotic lysis</td>
<td>Ficoll density gradient</td>
<td>EDTA, pH; SCX-RP-nLC</td>
<td>MALDI-MS/MS</td>
<td>148 (57%)</td>
<td>57-314,034</td>
<td>15</td>
</tr>
<tr>
<td>CELL WALL</td>
<td>FY833 (n)</td>
<td>Direct mechanical disruption</td>
<td>Acid precipitation;</td>
<td>SAX</td>
<td>nLC-ESI-MSMS</td>
<td>19 (5%)</td>
<td>600-42,000</td>
<td>Phenotypic cell-wall assay using knock-outs</td>
</tr>
<tr>
<td>PLASMA MEMBRANE</td>
<td>FY1679 (n)</td>
<td>Direct mechanical disruption</td>
<td>Acid precipitation;</td>
<td>CTAB/SDS PAGE</td>
<td>MALDI-MS/MS</td>
<td>48 (19%)</td>
<td>377-1,255,722</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>FY1679 (28°C)</td>
<td>Direct mechanical disruption</td>
<td>Acid precipitation;</td>
<td>ATPase assays; SDS PAGE</td>
<td>MALDI-MS/MS</td>
<td>86 (21%)</td>
<td>166-1,255,722</td>
<td>18</td>
</tr>
</tbody>
</table>

a) (n) or (2n) behind the strain name indicates haploid or diploid strain, respectively
b) Unless indicated otherwise, the cells were grown aerobically at 30°C in YPD medium containing 1% yeast extract, 2% Bacto-peptone, 2% glucose
c) YPL: 1% yeast extract, 2% Bacto-peptone, 0.5% lactic acid
d) YPG: 1% yeast extract, 2% bactopeptone, 3% glycerol, pH 5.0

e) Digestion was carried out using four proteases: trypsin, chymotrypsin, Glu-C, subtilisin
f) SCL and SCD media are complete synthetic media supplemented with lactic acid or glucose, respectively
g) YNO: 0.1% yeast extract, 0.5% ammonium sulfate, 1.7 g/L yeast nitrogen base, 0.02% Tween 40, 0.3% glucose, 0.1% oleic acid
h) SCIM: 0.5% yeast extract, 0.1% peptone, 0.79 g/L complete synthetic medium mix, 0.5% ammonium sulfate, 1.7 g/L yeast nitrogen base, 0.1% Tween 40, 0.1% glucose, 0.15% oleic add
use they are prone to proteolytic degradation.

In all of the eighteen organelle proteomics studies the crude cell lysates were subjected to differential centrifugation to obtain enriched crude organelles (Chapter II, Fig. 1). Differential centrifugation can involve multiple subsequent centrifugation steps that are performed at increasing g-forces. Organelles can be further separated from each other on the basis of their physical properties by the means of density gradient centrifugation or free-flow electrophoresis; or based on their antigenic properties by immuno-techniques (Luers et al., 1998; Inadome et al., 2005). Density gradient centrifugation uses differences in buoyant density of the particles for separation (Table 1), and was utilized in all yeast organelle proteomics studies performed thus far (Chapter II, Fig. 1 and Table 2), except for the proteomes of the Golgi (where immuno-precipitation was used) and the cell wall (differential centrifugation only). In the case of peroxisomes, density gradient centrifugation was used in combination with immunologically based pullouts (Marelli et al., 2004).

Sub-organelle preparations

Membrane enclosed organelles such as mitochondria, peroxisomes, vacuoles, ER, Golgi and nuclei can be fractionated further into sub-compartments. The internal (luminal) content can be separated from the membrane upon osmotic and/or mechanical lysis of the organelles and the subsequent differential centrifugation. Because membrane proteins are usually of a lower abundance than luminal enzymes, enrichment of the membrane fraction of an organelle is important for a comprehensive identification of these proteins. In only two cases (vacuoles and mitochondria) have suborganelar proteomes been studied separately (Zinser et al., 1991; Zahedi et al., 2006; Sarry et al., 2007; Wiederhold et al., 2009).

Organelle proteomes: Organelle profiling

The proteomes of eight subcellular compartments (further referred to as experimental organelle proteomes) have been analyzed by mass spectrometry-based techniques in eighteen separate studies (Table 2). The eight experimental organelle proteomes combine results of the corresponding individual studies dedicated to the same organelle. These are proteomes of the cell wall (Yin et al., 2005), the plasma membrane (Navarre et al., 2002; Delom et al., 2006), the Golgi apparatus (Inadome et al., 2005), the vacuoles (Sarry et al., 2007; Wiederhold et al., 2009), the nucleus (Gauci et al.; Mosley et al., 2009), the peroxisomes (Schäfer et al., 2001; Yi et al., 2002; Marelli et al., 2004), the mitochondria (Pflieger et al., 2002; Sickmann et al., 2003; Ohlmeier et al., 2004; Prokisch et al., 2004; Reinders et al., 2006; Zahedi et al., 2006) and the lipid particles (Athenstaedt et al., 1999) (Table 2, Supplemental Table 1).

Two questions arise in the light of these studies: how complete are these organelle proteomes, and how reliable are the results of these studies? In other words: what level of sensitivity has been achieved, and how selective were these studies? Sensitivity is determined by the analytical methods used, as well as by the isolation procedure, in which proteins may be lost. Selectivity is determined by the quality of the organelle preparation: How many contaminating proteins from different cellular locations were present in the purified organelles? To be able to answer these questions accurately, we would need to know the subcellular localization of all protein in yeast. But, of course, we do not have this information, because determining the subcellular localizations was the purpose of organelle proteomics. Fortunately, for yeast we can use a good approximation, because of the wealth of information organized in SGD, a manually curated database which integrates information on all genes and gene products of S. cerevisiae.

Defining the reference proteomes

For the purpose of this review, and based on the annotation of subcellular locations from the yeast genome database SGD, we define that proteins can be located in eleven different subcellular locations: cell wall, plasma membrane, cytosol, nucleus, mitochondrion, peroxisomes, lipid particle, endosome, ER, Golgi, vacuole (Supplemental Table 2). We define the
content of each reference proteome as ‘all proteins that according to the annotations in the SGD are present in the subcellular location’. The annotations for localization in the SGD are derived from a variety of genome-wide high-throughput studies (Kumar et al., 2002; Huh et al., 2003), low-throughput studies on individual proteins (e.g. immuno-EM), and global computational predictions. Localization data obtained in the eighteen mass spectrometry based organelle proteomics studies had not been included in the SGD at the time of our analysis, apart from the mitochondrial proteomics data, the implications of which will be discussed below.

It is possible that a protein has SGD annotations for localization in more than one organelle. Multiple annotated locations of a protein could be biologically relevant, as some proteins reside in more than one subcellular location. But dual or multiple localizations may also be artifacts of the methods used, e.g. a tagged protein may be targeted to the wrong location. So, we emphasize that the reference proteomes are not complete, and contain mistakes. Nonetheless, by-and-large they give us a good approximation of the numbers of proteins to expect. Because we have to take into account dual/multiple localizations, proteins with more than one annotated localization in the SGD were counted in each separate reference organelle proteome.

The reference organelle proteomes defined in this way have very different sizes (Fig. 1, large pie diagram, bottom right). The two smallest annotated proteomes harbor only 35 (lipid particles) and 58 proteins (peroxisomes). In contrast, the two largest reference proteomes contain 1,820 and 1,870 proteins (cytosol and nucleus, respectively), of which 576 proteins have annotations for both organelles. The mitochondrial proteome with 1,056 proteins is the third largest, containing 594 proteins with unique mitochondrial annotation. Other organelle proteomes are of intermediate size, containing between 100 and 300 annotated proteins. For many of the proteins in the reference proteomes, not only their subcellular localizations, but also their copy numbers in the cell has been determined.

The copy numbers are largely based on a genome-wide study using TAP-fusions in conjunction with quantitative Western blot analysis, which yielded copy number estimates for 3,837 out of the 5,797 yeast proteins (Ghaemmaghami et al., 2003). Again, just like the annotation of subcellular localizations, the annotated copy numbers for some of the proteins may be incorrect (for instance when the tag interferes with targeting, it is also likely to affect the copy number).

However, for the purpose of a global analysis of the organelle proteomes, the data provide us with extremely useful information. We distributed all quantified proteins over 5 abundance bins (copy numbers: <500; 500-2,000; 2,000-5,000; 5,000-20,000; 20,000-2,000,000). The bin widths were chosen to have approximately equal numbers of proteins per bin. The small pie charts in the Fig. 1 show the distribution of cellular localizations of the proteins in each abundance bin. The distribution varies among the bins: The reference peroxisomal and nuclear proteomes are slightly biased towards lower copy numbers (only 75 nuclear proteins and no peroxisomal proteins are present in the cell at a copy number >20,000). The bias for low abundance proteins in the nuclear proteome may be caused by the presence of many low copy-number transcription factors and regulatory proteins. The low abundance of peroxisomal proteins is caused in part by the growth conditions used in the studies to determine abundance: peroxisomal proliferation was not induced (with oleic acid). The reference proteomes for the cytosol, the cell wall and lipid particles have relatively large fraction of high copy number proteins, i.e. highly abundant glycolytic and other metabolic enzymes, and free ribosomes of the cytosol.

The protein abundances range from $10^1$-$10^6$ molecules per cell in four of the reference proteomes (cell wall, plasma membrane, mitochondria, cytosol), from $10^1$-$10^5$ in three organelles (vacuole, nucleus, and Golgi), from $10^1$-$10^4$ molecules per cell in lipid particles and ER, and from $10^2$-$10^5$ for endosomal proteins.
Figure 1. ABUNDANCE, ANNOTATED LOCALIZATION AND EXPERIMENTAL COVERAGE OF YEAST PROTEINS
The yeast proteins were split into six groups according to their cellular copy numbers. The top six bars represent the abundance groups and show the percentage of proteins identified by experimental proteomics studies in each group. The pie diagrams to the right of the bars show the annotated localizations of the proteins (identified plus unidentified) in each abundance group. The total number of unique proteins per group is indicated as the “SUM”, and the number of identified proteins is indicated in parentheses. Because proteins with multiple annotations were counted more than once, the sum of the proteins from all the segments of each pie is not the same as the SUM of unique proteins.

The bottom bar shows the percentage of identified proteins for the entire yeast proteome, and the large pie at the bottom-right shows the annotated localizations of all 5,797 yeast proteins. The localization of the 3516 yeast proteins identified in at least one proteome study, and the localization of the 2,281 unidentified proteins are shown in the bottom pies (left and middle, respectively).

Each color indicates a subcellular location: orange - cytosol; pink - plasma membrane; beige - cell wall; red - mitochondria; green - vacuole; light violet - endosome; violet - Golgi apparatus; purple - endoplasmic reticulum; grey - nucleus; blue - peroxisome; magenta - lipid particles; yellow - no localization assigned; black - others (e.g. extracellular).

The reference peroxisomal proteome has the narrowest range covering only 2 orders of magnitude from $10^2$-10^4 molecules per cell (Table 1).

**Overall experimental coverage**

The eighteen proteomics studies of yeast subcellular compartments have collectively led to coverage of 61% of the proteins from the entire predicted yeast proteome (Athenstaedt et al., 1999; Schäfer et al., 2001; Navarre et al., 2002; Pfieger et al., 2002; Yi et al., 2002; Sickmann et al., 2003; Marelli et al., 2004; Ohlmeier et al., 2004; Prokisch et al., 2004; Inadome et al., 2005; Yin et al., 2005; Delom et al., 2006; Reinders et al., 2006; Zahedi et al., 2006; Sarry et al., 2007; Gauci et al.; Mosley et al., 2009; Wiederhold et al., 2009). In other words, 3,516 different proteins were found in at least one of 18 individual proteome analyses. It is noteworthy that approximately 80% (4,517 proteins) of the yeast proteome has been detected under the standard growth conditions by chromosomal TAP- and GFP-fusions (Huh et al., 2003; Ghaemmaghami et al., 2003).

The lower fraction of identified proteins by the ensemble of proteomics studies may have several causes: (1) Some large subcellular compartments have not been analyzed in dedicated proteomics studies (cytosol, ER, endosome); (2) loosely attached proteins may be lost in the organellar isolation procedures; (3) some proteins escape identification in the proteomics studies because of their sequence properties. For example, small and hydrophobic proteins may not yield suitable peptides for identification by mass spectrometry when the common proteolytic enzyme trypsin is used. Trypsin cleaves C-terminally of arginine and lysine residues, and small/hydrophobic protein may lack these residues at suitable positions.

As indicated in the Fig. 1, the fraction of identified proteins in the collective proteome studies increases with protein abundances. Whereas only 54% of yeast proteins with a copy number <500 are covered by the proteome analyses, nearly 95% of the proteins with copy number >20,000 have been identified. These numbers indicate that abundance is the dominant determinant for protein identification.

The coverage of the proteomes of the individual organelles, expressed as the percentage of identified proteins of each reference proteome, varies considerably (Fig. 2). The reference proteomes with the highest coverage are the mitochondrion (88%) and peroxisomes (84%). The proteomes of the nucleus, Golgi, ER and lipid particles are covered to 70-75%. The vacuolar and cytosolic proteomes with 60% and 65% coverage, respectively, are close to the average of 61%. The proteomes of the cell wall (46% coverage), the endosome (50% coverage) and the plasma membrane (51% coverage) are covered below average. The low coverage of the cell wall proteome is caused by the experimental design, which was aimed at identification of covalently bound (mainly GPI-anchored) proteins. Soluble cell wall proteins were discarded by applying harsh extraction methods (Yin et al., 2005). The proteins without known localization were covered to only 16%. The low coverage may indicate that these proteins are not usually expressed under the conditions tested. Just
like for the entire yeast proteome, the coverage of proteins in the individual organelles tends to be better for more abundant proteins (Fig. 2). 2,281 proteins (~1/3 of the predicted proteome) still remain to be identified by mass spectrometry. The largest group of unidentified proteins (785) consists of proteins without known localization, of which the majority does not have assigned biological and/or molecular function. The abundance of unidentified proteins is biased toward low copy numbers: more than 1/3 (700 proteins) of the unidentified proteins have copy numbers of less than 2,000 per cell, and for more than a half (1,233 proteins) the copy number could not be determined. It seems that all high-throughput methods (both mass spectrometry-based proteomics and tagging techniques such as GFP- and TAP-fusions) have so far failed to record these proteins. The reason could be that the copy numbers are below the detection limit (~40 copies per cell), or that these proteins are not expressed under the conditions used. Among the unidentified proteins there are many regulatory proteins that are involved in signal transduction, such as GTPases, and transcription and translation factors, which might be expressed only under specific conditions.

**Figure 2. COVERAGE OF THE ORGANELLAR PROTEOMES AND SENSITIVITY OF THE PROTEOMICS STUDIES**

Analogous to the presentation in Fig. 1, the coverage of each organellar proteome is depicted as the total coverage of the reference proteome (the last bar in each small diagram), or for each abundance group separately (first six bars, from left to right represent copy numbers: <500; 500-2,000; 2,000-5,000; 5,000-20,000; 20,000-2,000,000; and no abundance data available). The coverage is presented as the fraction of proteins in each reference proteome that was identified by any of the eighteen experimental proteome studies. The fraction of each reference proteome that is covered by the dedicated organelle proteome study is indicated as the striped area of the total organelle coverage and represents the apparent sensitivity of the experimental proteomics experiments.

**Sensitivity of experimental proteomes**

The coverage of the reference proteomes described above is the sum of all the proteins identified in the different experimental studies. It does not imply that all of the identified proteins were found in experimental studies aimed at the organelle of interest; some proteins may have been
found only as contaminants in experimental studies aimed at studying different organelles. If the reference proteomes did not contain mistakes, and the sensitivity of the analytical techniques was infinite, each experimental proteome study would find all proteins of the corresponding reference proteome. In reality, the overlap is never complete, because the reference proteomes are not perfect, and the sensitivity of the analytical techniques is limited. For the eight cellular locations that have been targeted experimentally by proteomics studies, the overlap (apparent sensitivity) is indicated in Fig. 2 as striped areas of the each organelle reference proteome. Clearly, there is a great variety in the apparent sensitivities of the different proteomics studies. The dedicated investigations of the plasma membrane, the cell wall and Golgi found less than 1/4 of their respective reference proteomes (32 of 270 proteins for plasma membrane; 19 of 110 proteins for the cell wall, and 41 of 199 proteins Golgi). The reference proteomes of lipid particles and vacuoles are moderately covered by the dedicated proteome studies with about 40-50% coverage. The majority (over 70%) of the reference mitochondrial (881 of 1,056 proteins), peroxisomal (45 of 58 proteins) and nuclear proteins (1,303 of 1,870 proteins) were found in the dedicated experimental studies which indicates a high sensitivity.

Specificity of experimental proteomes

From Fig. 2 it is clear that for all organelles, fewer proteins have been identified by dedicated studies (striped parts of the bars) than by the ensemble of all eighteen organelle proteome studies (total bars). Strikingly, in some organelles the majority of covered proteins was found in different studies than the one(s) targeted at the specific organelle (i.e. endosome, cytosol, ER). The extra proteins found in the ensemble could be contaminants from proteome studies dedicated at other organelles, or could indicate incomplete or wrong database annotations of the reference proteomes. Here we discuss some striking examples, which highlight the difficulty of distinguishing between the two possibilities.

One way to obtain an indication about specificity is to calculate the fraction of proteins identified in each of the eight experimental proteomes that was indeed annotated in the SGD as localized in the organelle of interest. In Fig. 3 this apparent specificity is indicated. The apparent specificity ranges from 100% (all proteins identified in the lipid particle and cell wall proteomes have corresponding annotations) to less than 30% (more than 70% of the identified proteins in the plasma membrane proteome have annotations for different cellular locations).

**Figure 3. APPARENT SENSITIVITY AND SELECTIVITY OF EXPERIMENTAL ORGANELAR PROTEOMICS STUDIES**

The apparent sensitivity was calculated as percentage of proteins from the reference proteome that was identified by the dedicated experimental organelar proteome study. The apparent specificity was calculated as the fraction of proteins identified in the dedicated experimental proteomics studies that was annotated as localized in the same subcellular location.

The high apparent specificity of the lipid particle and cell wall proteomes are a direct consequence of the biochemical fractionation methods, which resulted in very clean preparations because of the characteristic physical properties of these subcellular isolates. But a low apparent specificity does not necessarily indicate low quality of the biochemical fractionation. For instance, when highly sensitive analytical
methods are used, the apparent specificity is likely to become lower, because low-abundant contaminants are more efficiently detected (e.g. in the case of the nuclear proteome). Also SGD annotations play a crucial role in the apparent specificity. For example, the outcome of the mitochondrial proteome analyses has been included in annotations for localization in the SGD, which obviously will increase the apparent specificity. In contrast, data form other proteome analysis have not (yet) been included in the SGD, which lowers the apparent specificity for those proteomics studies that identify a large number of new organelle proteins (for example the vacuolar proteome, see below). A rough indication about the overall specificity can also be drawn from the number of uniquely identified proteins in each of the eight experimental proteome analyses (Fig. 4). Nearly 74% of all identified proteins were found in one proteomics study only, and 22% in two experimental organelle proteomes. The remaining 4% of the identified proteins were found in 3-5 proteomes with a clear trend (approximately 2/3) towards a high abundance (copy number >20,000) for those proteins. Some of these proteins are likely non-specific contaminations, e.g. P-type ATPase Pma1 of the plasma membrane, ribosomal proteins and many glycolytic enzymes. But among the proteins present in multiple proteomes are also proteins that are known to have multiple localizations, e.g. proteins involved in signal transduction like GTPases Rho1, YPT1 and CDC42, or the protein kinase TOR2. No proteins - not even very highly abundant ones - were found in more than 5 organelle proteomes.

The coverage of the reference proteomes for ER, endosome and cytosol are above 50%, although no dedicated proteomics studies were performed for these organelles. This could be indicative of large amounts of contaminants in the organellar preparations. However, there are also alternative explanations for the detection of these proteins. The largest experimental organelle proteome, the nuclear proteome, contributed the most to their coverage. Almost all proteins (236 of 253 identified proteins with ER annotations, 42 of 52 endosomal and 1,079 of 1,174 cytosolic proteins) were identified in the nuclear proteomics studies. There is a straightforward biological explanation for the appearance of ER proteins in the

Figure 4. PROTEINS IDENTIFIED IN MULTIPLE EXPERIMENTAL PROTEOMES
Each of the 3,516 identified proteins was found in 1 to 5 experimental proteomes. The bars indicate the total number of proteins found in 1, 2, 3, 4 or 5 proteomes. The inset shows an enlargement of the bars for the proteins identified in 3, 4 or 5 proteomes. The colors indicate in which experimental proteomics studies the proteins were identified: pink - plasma membranes; beige - cell walls; red - mitochondria; green - vacuoles; violet - Golgi apparatus; grey - nuclei; blue - peroxisomes; magenta - lipid particles.

70
nuclear proteome, as the nuclear envelope is continuous with the ER. The perinuclear ER has general ER function and covers ~20% of the ER surface (Preuss et al., 1991; Strambio-de-Castillia et al., 1995). We took a more detailed look into the coverage of the reference ER proteome, and found that out of the 253 identified proteins 160 ER-annotated proteins were uniquely detected in the nuclear proteomes. As indicated above, finding these ER-annotated proteins in the experimental nuclear proteomes is biologically relevant. Of the remaining proteins, 62 were detected in 2 experimental proteomes, 9 proteins in 3 proteomes, and 5 proteins in 4 proteomes, mainly from the experimental nuclear, mitochondrial, vacuolar and peroxisomal proteomes. The annotated functions of the ER proteins found in multiple experimental proteomes are biased towards sterol and phospholipid biosynthesis and protein trafficking across the secretory system. The proteins found in multiple organelle proteomes are relatively low abundant proteins and, in the light of the mentioned functions, might be of biological relevance, i.e. these proteins have multiple localizations in the cell.

The proteins with endosomal annotation make up 14% of the experimental Golgi proteome and 5% of the vacuolar proteome, respectively. In the other experimental proteomes, endosomal proteins account for a very minor fraction (<1%, even though the absolute number in the nuclear proteome is large). The endosome is an integral part of the secretory machinery and shares many components of the vesicle fusion and protein trafficking complexes with vacuoles and the ER-Golgi network. It is not surprising therefore, that many of the proteins that are annotated as "endosomal" in the yeast database, and which are found in the Golgi and vacuolar proteomes, have functions in ER-to-Golgi transport. The identified endosomal proteins in the Golgi and vacuolar proteomes are therefore unlikely to be contaminants.

The experimental proteome of the plasma membrane has the lowest apparent sensitivity and specificity of all eight proteomes (Fig. 3). In two dedicated proteomics studies, 121 proteins could be identified of which 73 proteins are annotated as cytosolic in the SGD. Those 73 proteins are mostly high abundant contaminations such as ribosomal proteins and glycolytic enzymes.

**How to deal with contaminants?**

The above examples show the need to deal explicitly with contaminants, because no biochemical separation technique will yield 100% pure organelles. In all organelar proteomics studies the obtained list of identified proteins was evaluated very carefully by comparison with existing database and literature information. Proteins that had not previously been assigned to the studied organelle were either regarded as contaminants, or considered as novel proteins. An example is the peroxisomal proteome (Yi et al., 2002), where the authors could identify ~240 proteins. After careful analysis, only 46 of these proteins were assigned as peroxisomal based on prior knowledge about the peroxisomal content. However, when aiming at identification of new proteins not previously assigned to an organelle, a database- and literature-based evaluation may not be sufficient.

One way to deal with contaminants would be to obtain purer organelle preparations. The mitochondrial proteome which is of high apparent sensitivity and selectivity indicates that very good purity can be achieved. However, for physical and biological reasons (Binns et al., 2006; Kvam & Goldfarb, 2006; Perktold et al., 2007; Schumann & Subramani, 2008), it is unlikely that organelles can be isolated to an absolute purity and homogeneity; rather the final preparations are enriched to a variable extent depending on the fractionation methods. In many cases marker proteins for the organelle of interest were used to assess the purity (either by Western blotting or by activity assays, Table 2). Marker enzymes for different organelles are indicated in Table 1. Marker enzymes are useful for qualitative assessment of organelle enrichment, but the assays are not sufficiently sensitive to judge the purity of a preparation for MS purposes. Even when organelle
preparations appear completely pure based upon immunological detection of the markers, they may still be contaminated and these contaminations will be revealed by mass spectrometry due to a higher sensitivity of the method. The issue of contaminating proteins is particularly problematic, because proteins in cells and organelles are present in abundances that can vary by 6 orders of magnitude. Consequently, even if a very small fraction of a contaminating organelle is present in the final preparation, highly abundant proteins originating from the contaminating organelle may be as easily detectable by MS as the low abundant proteins from the organelle of interest.

Instead of trying to avoid contaminants by using (apparently) very pure preparations, the contaminations can also be accounted for. One way of doing so is to use additional quality assessments, for instance GFP-tagged fusions. However, as mentioned above, GFP-tagged proteins are not always targeted to their native destinations which results in mis-localizations.

A robust approach to account for contaminants is to use relative quantification-based techniques (for reviews see (Ong & Mann, 2005; Yates III et al., 2005; Andersen & Mann, 2006)). The original idea of this approach (reviewed in (De Duve & Beautay, 1981)) is that proteins that are physically associated with an organelle become enriched upon purification of this organelle, for instance by density gradient centrifugation. The contaminants, in contrast, although still present and detectable by mass spectrometry, will become depleted upon purification. The specific enrichment or depletion can be quantified by comparing protein abundances in crude and pure preparations of the organelle of interest (Marelli et al., 2004; Dunkley et al., 2006; Wiederhold et al., 2009), or by comparing abundances in different fractions of a density gradient after centrifugation (Fig. 5). When the protein abundances in all of the fractions are compared, proteins with multiple localizations can be recognized, because they will be enriched in more than one part of the gradient. Enrichment in multiple fractions of a density gradient may also occur when the preparation of an organelle is heterogeneous. Such heterogeneity may be the result of the presence of slightly different organelle composition within a single cell, or because the final preparation contains organelles derived from cells of different age and cell cycle stage. For quantification isotope tags such as iTRAQ or ICAT can be used, as developed by the group of K. Lilley to assess the Localization of Organelle Proteins by Isotope Tagging (LOPIT) (Dunkley et al., 2004a; Dunkley et al., 2006; Sadowski et al., 2006).

Alternatively, label-free quantification can be performed by protein correlation profiling, which is based on the intensities of the peptide peaks in different MS analyses (Andersen et al., 2003; Wiese et al., 2007; Yan et al., 2008), using the Proteins Abundance Index (PAI) which represents the number of peptides identified divided by the number of theoretically observable peptides (Rappsilber et al., 2002; Ishihama et al., 2005), or spectral counting where the number of mass spectra identified for a protein is used as a measure of the protein’s abundance (Pang et al., 2002). Usage of the spectral count-based quantification for absolute and relative protein abundance determination in complex mixtures is described in Vogel et al., (Vogel & Marcotte, 2008). These different methods to account for contaminants are globally outlined in Fig. 5.

Label-based quantification in the yeast proteomic studies was successfully utilized for the vacuolar membrane and peroxisomal proteomes (Marelli et al., 2004; Wiederhold et al., 2009). The most valuable result of these studies is that many proteins with annotated localizations other then vacuolar or peroxisomal were found to be specifically co-enriched with vacuoles or peroxisomes which strongly indicates that they are part of the true proteomes of these organelles. Thus, the low apparent specificity of the peroxisomal and vacuolar proteome studies (Fig. 3) appear not to be caused by contaminants, but rather by incomplete database annotations for localization. Proteomics studies in which quantification was used to assess subcellular localization allow re-assignment
Figure 5. PRINCIPLES OF LOPIT AND LABEL-FREE QUANTIFICATION
The Localization of Organelle Proteins using Isotope Tagging (LOPIT) technology is based on the notion that biochemical purification of an organelle never results in an absolutely homogeneous preparation. Instead, proteins that are specifically associated with the organelle of interest (green dots) become enriched upon purification (top layer, “a”), and contaminants (red dots), although still present in the enriched fraction “a”, become depleted. The LOPIT technology employs labeling of peptides with stable isotopes such as ICAT or iTRAQ (right-hand side of the figure). For this, peptides from the enriched (“a”) and depleted (“b” and “c”) samples are labeled with reagents of different masses, mixed together and processed by liquid chromatography and mass spectrometry. The isotope labeling allows relative quantitation of the protein abundance in the enriched and depleted fractions simultaneously (“a”, “b”, “c”). For the label-free quantitative approach (left-hand side of the figure), peptides from the enriched (“a”) and depleted (“b” and “c”) fractions are processed by liquid chromatography and mass spectrometry separately without labeling or mixing. In this case, the protein abundance in each fraction is quantified based on the number of observed peptides divided by the number of observable peptides (Protein Abundance Index; PAI); the peptide peak intensity (Protein Correlation Profiling; PCP), or the spectral count (number of identified spectra for each protein).

Conclusion
Mass spectrometry-based proteomics is a powerful technique to study the subcellular localization of native, non-tagged proteins. The currently available yeast organellar proteomes have yielded a tremendous amount of valuable information. The proteomics studies complement high-throughput localization studies based on GFP-tagging and imaging. Both techniques have their advantages: proteomics studies do not require tagging, which may cause targeting failure; and GFP-tagging studies do not require subcellular fractionation. Subcellular fractionation of the cell lysate is an essential part of organellar proteomics. Robust methods to isolate subcellular proteomes are available and produce highly purified preparations. But the isolation procedures are lengthy and may lead to loss of proteins that are loosely attached or susceptible to degradation. Therefore, more rapid isolation procedures, e.g. based on affinity pullouts as for instance used for Golgi-isolation (Inadome et al., 2005), are desirable (Cristea et al., 2005).

No organellar preparation will be completely free of contaminants from different subcellular locations. Quantitative proteomics techniques that assay enrichment profiles allow for the
discrimination between contaminants and genuine organelar residents, and facilitate the assignment of cellular localizations to the identified proteins (Andersen et al., 2003; Foster et al., 2003; Schirmer et al., 2003; Marelli et al., 2004; Dunkley et al., 2006; Wiederhold et al., 2009). With the ongoing developments in quantification and identification methods it will become feasible to perform a global analysis of protein localization in yeasts from a single culture. One could envision that such an experiment would use a multidimensional fractionation scheme and extensive analysis of co-fractionation profiles that follow from quantitative mass spectrometry using e.g. the multiple reaction monitoring method (Kuhn et al., 2004; Fortin et al., 2009). We expect that combining these data with those from co-purification strategies where protein complexes are isolated using tagged proteins will provide a more complete and accurate description of protein localization. Ultimately, dynamics will also have to be included in the localization experiments, because proteins may be associated only temporarily, and dependent on the environmental conditions, with an organelle proteome (Andersen et al., 2005).

Acknowledgements
We acknowledge support from the Netherlands Organisation for Scientific Research NWO (VIDI fellowships to DJS and LMV; TopSubsidy to BP), the Netherlands Proteomics Centre, and the Zernike Institute for Advanced Materials.