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.

Download date: 01-12-2018
The philosopher: Your Highness, ladies and gentlemen, I am just asking myself where all this may lead.

Galileo: I would suggest that as scientists it is not for us to ask where the truth may lead us.

The philosopher furiously: Signor Galilei, the truth may lead us to absolutely anything.
Mass spectrometry-based proteomics is a widely used technology for high-throughput identification of proteins in protein mixtures. Qualitative identification alone is not sufficient to answer current/modern biological questions, and quantitative methods are required to determine absolute and relative amounts of proteins, time- and condition-dependent changes in protein levels and modifications, and cellular localization of proteins. The present thesis demonstrates in Chapters III and VI two different applications of quantitative proteomics. In one case the iTRAQ-based quantitative proteomics was applied to determine specific co-enrichment of proteins along with vacuoles during subcellular fractionation (Chapter III). The quantitative information was used to discriminate between true vacuolar residents and contaminating proteins. In the second application, the physiological response of the bacterium *Lactococcus lactis* was studies, where the relative changes in protein abundance upon production of human CFTR were followed by quantitative proteomics (Chapter VI). Of course, the proteomics methods used were specifically adapted for each application, and dictated by the available instrumentation, but in this final chapter the question is addressed whether we can make general recommendations for an optimal quantitative proteomic procedure. What did we learn from the work described here?

**Sample preparation**

The first step of any mass spectrometry-based proteomic analysis is the isolation of a protein sample and generally requires the protein mixture to be extracted and purified from cells/tissues. For subcellular proteomics, cells are broken and the desired subcellular proteome (*e.g.*, an organelle) is purified from the lysate. Eukaryotic cells and organisms are far more complex than bacteria and sample preparation usually requires multiple steps of fractionation in order to obtain the proteome of interest. In Chapter II, an optimized protocol for isolations of highly pure vacuoles from *Saccharomyces cerevisiae* is described, and the most crucial steps in the isolation procedure are emphasized. Additionally, some important physical and biochemical features of yeast organelles, that can play a role in successful organelle proteomics, are described in Chapter II.

Often, regardless of the origin of the sample (organelles or bacterial cells), the soluble and membrane fractions are separated by differential centrifugation and the fraction of interest is subjected to the proteomic analysis. As described in Chapter VI, we suggest analyzing both the membrane and soluble fractions separately, which provides a more complete picture of the proteome rather than only a fraction of it. To facilitate the detection of (low-abundant) membrane proteins, the membrane fraction can be stripped of soluble proteins by salt/urea/pH treatment. Stripping of the membrane is recommended, but the stripped-off soluble proteins should be combined with the soluble fraction for completeness.

**Number of replicates and control samples**

Sample preparation is often the most labor intensive and time-consuming part of a proteomics analysis, and therefore it is important to plan carefully at a very early stage how many and what kind of samples are needed. Especially when working with plants (or higher eukaryotes in general), which need long growing times, a sufficient number of biological replicates has to be grown. For the statistical analyses described in Chapters III and VI, at least three biologically independent samples are needed as empirically determined. But for convenience of labeling, four replicates might be considered (for an example of a labeling scheme see Chapter VI).

The inclusion of technical replicates to estimate the variation in the iTRAQ labeling procedure may be omitted because, as shown in Chapter VI, the technical variation at this stage of the experiment is very low. What we learn from the Chapter VI is that following time-dependent changes in the proteome (in both the control and treated samples) allows for more insight in the...
physiological response than comparison at only a single time point of the control and treated proteomes. Therefore, we recommend measuring and comparing multiple time points of the control and treated samples.

Sample processing and protein identification

Once the samples have been prepared, the proteins have to be identified and quantified by mass-spectrometry-based methods. Out of many ways of protein analysis by MS-based proteomics, we have chosen the bottom-up approach that involves enzymatic protein digestion to generate peptides, peptide separation by strong cation exchange and reverse phase chromatography and tandem MS analysis using MALDI-TOF/TOF. After the peptides have been analyzed in the mass spectrometer, the spectra have to be matched to peptide sequences and the corresponding proteins can be identified. In our laboratory, two database search engines Mascot and Paragon are used for peptide identification. How can the user make a decision which search engine to use, and does it matter? In Chapter VII, the Mascot and Paragon database searching engines were compared in terms of their performance to match peptides to measured spectra. Although additional experiments are required to complete the comparison, it became evident that Mascot might be more reliable for assigning peptides to spectra of moderate quality. Although the total number of detected proteins did not differ significantly between Paragon and Mascot, Paragon sometimes assigns different and evidently wrong peptides to spectra, and therefore possibly also different and wrong proteins. Wrong identifications might have a huge impact on quantification of peptides and proteins. Therefore, we suggest using Mascot for protein identification, until we have found out why the two search engines perform differently.

Protein quantification and statistical analysis

When the proteins have been confidently identified and quantified, the researcher faces the question: How to determine the threshold between different groups of interest, e.g. changed and unchanged relative abundance? To answer this question robust statistical methods are required. Using iTRAQ reagents, each spectrum contains information about the relative peptide abundance expressed as fold-change or ratio. Because at least two peptides are required for confident protein identification, the peptide ratios need to be averaged in order to obtain protein ratios. Simple (unweighted) averaging introduces large variations due to the large dynamic range of measured spectra. Weighted averaging as used in ProteinPilot and described in Chapter VI allowed for a robust quantification of identified proteins. Prior to quantification, the data have to be normalized. In contrast to normalization at the protein level (the procedure described in ProteinPilot), we empirically determined that normalization at the peptide level was more robust (Chapters III and VI).

Validation of proteomics data

Although a careful statistical analysis provides very reliable results, a fraction of false positives will be present. Therefore, complementary methods are often used to
evaluate proteomics data in order to obtain more confidence. When 22 novel proteins were found in the vacuolar membrane proteome (Chapter III), there was still a possibility that some of these proteins were either statistically or biologically ‘false positives’, because proteins that are targeted to the vacuole for degradation are expected to co-enrich with intact vacuoles just like true vacuolar resident proteins. Therefore, subcellular localization of a sub-set of novel vacuolar proteins was evaluated by GFP fusions as presented in Chapter IV. Of course, the validation method can be different for other research questions, but the need for validation must be emphasized here.

**Outlook**

Mass spectrometry-based proteomics is an expanding and improving technology, pushing the boundaries in multiple research fields from systems biology to biomarkers discovery. In the last couple of years, quantitative proteomics became possible due to technological developments, and quantitation has become increasingly important. However, many limitations still exist when addressing the complexity and dynamic behavior of biological systems. New instrumentation, innovative bioinformatics and analytical concepts continue being developed, *e.g.* targeted approaches such as selected reaction-monitoring assay (Picotti et al., 2010). Also, other areas of *omics* research (*e.g.* metabolomics, genomics, lipidomics) contribute increasingly to our understanding of living organisms. In the future, an effective integration of all available data from different -omics researches will become more and more important.