Galileo *almost friendly*: And why must it be physics? Why not horse-breeding?
Ludovico: My mother thinks that a little of science is necessary. All the world takes a drop of science with their wine nowadays, you know.
Each living organism is built up of diverse types of molecules, the complete collections of which ("bodies") are described as "omes", such as metabol-ome (metabolites), lipid-ome (lipids), gen-ome (DNA), transcript-ome (RNA) and prote-ome (proteins). The study of proteomes extents beyond the question of expression (presence, absence, quantity of each protein) but also addresses cellular localization, interactions, activity, modifications and different isoforms. A large collection of different techniques can contribute to the study of proteomes, including array-based approaches and microscopic imaging, and mass spectrometry-based approaches. Mass spectrometry (MS) is a high-throughput method with the capacity to identify very small amounts proteins in complex mixtures (Aebersold & Mann, 2003). Since the establishment of mass spectrometry for proteome studies, the term "proteomics" largely refers to MS-based proteomics. MS-based proteomics is not limited to one general approach but is rather a collection of different techniques with strength for a particular application (Han et al., 2008). The most popular strategy in proteomics is the bottom-up approach in which proteins are cleaved into peptides, the peptides are analyzed and the information on the peptides is used to infer information on the proteins. In contrast the top-down approach analyses whole proteins. Often the proteomic studies have been qualitative, aiming at determining of which proteins are present in the sample, providing a snap-shot overview of the proteome. However, proteomes in contrast to genomes are dynamic and change in time, space and in response to stimuli (Andersen et al., 2005). Quantitative proteomics can provide insights into the dynamic changes of proteomes and aims at determining how much of each protein is present in the sample at a particular time, and how the abundances change over time. Whereas the absolute quantification determines the total amounts of proteins in the sample expressed in, e.g. pmol/µL, during relative quantification the amounts of proteins between different samples, e.g. control and treated cells, are compared and expressed as fold-change or ratio. There are mainly two different approaches for protein quantification: label-based (stable isotope label-}

**ITRAQ is a valuable tool for proteomics**

One very commonly used type of chemical reagent for label-based quantification is iTRAQ (isobaric Tags for Relative and Absolute Quantitation). The iTRAQ reagents chemically modify the primary amines, i.e. the N-termini of peptides and proteins and lysine-residues, and sometimes also tyrosine-residues. ITRAQ is a very convenient chemical modifier of peptides because each peptide will be labeled at the N-terminus regardless the peptide sequence and therefore does not rely on the presence of internal lysine and/or tyrosine residues. There are two sets of iTRAQ reagents available: 4plex allowing for labeling of four different samples, and 8plex with eight different reagents. Each reagent consists of a reactive group which reacts with a peptide, a balance and a reporter group (Fig. 1).
The iTRAQ reagents are isobaric which means all reagents from the same set, i.e. four-plex or eight-plex, have the same total mass which is 145 m/z and 305 m/z, for four-plex and eight-plex, respectively. But the masses of the reporter group are different for each reagent within the four-plex or eight-plex: 114 through 117 m/z for 4plex, and 113 through 119 plus 121 m/z for eight-plex. The masses of the balance group are different accordingly, 31 through 28 m/z for four-plex and 192 m/z through 186 m/z plus 184 m/z for eight-plex.

The peptides created upon trypsinization of proteins from the control and treated samples are differentially labeled with iTRAQ (Fig. 2). The labeled peptides are mixed together and subjected to MS. In the MS mode, the peptides with the same sequence derived from the control and treated samples are detected at the same m/z (left-hand spectrum) due to isobaric nature of the labels. Upon fragmentation in the MS/MS mode, the reporter groups of different iTRAQ labels are cleaved off and due to their different masses are detected separately (red and green peaks). These areas of these peaks are measured and, when compared with each other, reveal the differences in abundance of the peptides in the samples that are compared. The peptide sequence is determined based on the peptide fragments (black peaks, right-hand spectrum). Therefore, identification and quantification are obtained from the same spectrum.

**Outline of the thesis**

The present thesis describes two large-scale quantitative proteomics studies using iTRAQ labeling of peptides, which were subsequently analyzed by MALDI-MS/MS. In addition, sample preparation, data analysis, and validation of the outcome of the proteomics studies are discussed. Sample preparation is a crucial step in proteomics because the purity, homogeneity, and reproducibility are important factors in the interpretation of the data. In Chapter II, procedures for preparation of different organ-
elles from the yeast *Saccharomyces cerevisiae* are described. In addition in Chapter II the optimization of the isolation procedure of vacuoles and vacuolar membranes from *Saccharomyces cerevisiae* is described, and critical aspects of the procedure are discussed. It is shown that vacuolar membranes largely devoid of other subcellular impurities can be obtained, which is a prerequisite for organelle proteome analysis.

Chapter III deals with the proteomic study of the isolated and purified vacuolar membranes from *S. cerevisiae*. Although yeast vacuoles have been studied extensively using biochemical techniques, the knowledge about the vacuolar membrane proteome is not complete. The purpose of the study was the identification of novel vacuolar (membrane) proteins using mass spectrometry based proteomics. Despite a high purity of the vacuolar membranes achieved by the procedure described in Chapter II, no biochemical isolation procedure is absolutely free of contaminations. Quantitative proteomics using iTRAQ was used in this study to successfully distinguish between the true vacuolar residents and contaminations. When the relative abundances of proteins were compared in the pure and crude vacuolar fraction, the true vacuolar residents were enriched (increased abundance) in the pure fraction, whereas the contaminants were depleted (decreased abundance) from the pure vacuolar fraction. Applying quantitative proteomics in conjunction with a robust statistical analysis, it was possible to identify 22 novel vacuolar proteins that had not been assigned before to any subcellular location in yeast.

The twenty-two novel vacuolar proteins may be true vacuolar residents (proteins which fulfill their cellular function in the organelle), but could also be proteins that fulfill their functions elsewhere, but that are degraded in vacuoles and therefore are physically present in the vacuole. To further evaluate the novel vacuolar proteins, GFP-tagging and microscopic imaging for eleven proteins was performed which is described in Chapter IV. Nine of eleven evaluated proteins did indeed localize to the vacuolar membrane or vacuolar periphery, one localized to the ER, and for one protein the GFP-tagged gene product could not be detected. Overall, the study using GFP as a marker for subcellular localization of proteins in yeast was a valuable validation of the previous proteome study.

Organelle proteomics is a high-throughput approach, which allows for a bulk identification of proteins localized to an organelle. Knowledge of organelle-specific proteomes, such as the vacuolar membrane proteome (Chapter III) is important for understanding organelle functions. In the past decade, multiple research groups contributed to obtaining proteome “maps” of different organelles form yeast. In Chapter V, eighteen high-throughput studies aiming at organelle proteomes using mass spectrometry are summarized, among which our vacuolar proteome study of Chapter III. The reliability and completeness of the proteomic analyses are discussed. Related to the reliability of proteome analyses, quantitative proteomics technologies are discussed which allow for a reliable identification of organelle-specific proteins.

In Chapter VI, a different application of quantitative proteomics is presented, in which the physiological response to stress is studied in the bacterium *Lactococcus lactis*. Biochemical and biophysical studies on isolated proteins require that large amounts of the functional proteins can be produced and purified. Because the natural abundance of membrane proteins is often low, membrane proteins are usually (heterologously) overproduced in microorganisms. However, the over-expression of membrane proteins is often toxic to the expression host. The work presented in Chapter VI identifies bottlenecks in membrane protein over-production by studying the physiological effects in *L. lactis* upon heterologous expression of the human CFTR. The membrane and soluble proteomes from the strains bearing either an empty expression vector or the vector with the gene encoding His-tagged human CFTR were determined. Time-dependent relative changes in protein abundance after induction of the expression of CFTR in the two strains were determined based on iTRAQ technology. Groups of differentially expressed proteins were determined using statistical analysis (False Discovery Rate (FDR) and iterative Group...
I. CHAPTER: Introduction

Various cellular processes and biochemical pathways were found to be affected by the CFTR over-expression, particularly the stress response, ribosomal biogenesis and recycling, nitrogen and sugar metabolisms. Overall, outcome of the study indicates that changes associated with over-production of CFTR affect multiple processes within the cell rather than a few specific “players”.

In (quantitative) proteomics each measured MS/MS spectrum needs to be matched to a peptide sequence (Fig. 2). In Chapter VII two database searching engines Mascot and Paragon are empirically compared in terms of their reliable performance in peptide matching. Both the Paragon and Mascot Algorithms are available as integrated parts of the ProteinPilot software. Analyzing seven datasets that were used in Chapter VI, Mascot appeared as a more reliable spectrum matching engine. Further experiments are required to understand the reasons for matching inconsistencies with Mascot and Paragon.

Learning from the studies described in Chapters III, IV, VI, VII, an attempt is made in Chapter VIII to suggest an “optimal” strategy for quantitative proteomics using iTRAQ. Some critical aspects in quantitative proteomics such as the number of replicates, time scale, choice of reagents, search engine, approach for protein quantification and the statistical analysis are discussed.