Expanding instrumental capabilities of MS-based proteomics

Advanced LC and MS methods for in-depth analysis in challenging applications

Leonie Francelle Waanders

Expanding instrumental capabilities of MS-based proteomics

Advanced LC and MS methods for in-depth analysis in challenging applications

Een wetenschappelijke proeve op het gebied van de
Natuurwetenschappen, Wiskunde en Informatica

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S. C. J. J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op

vrijdag 2 oktober 2009
om 10.30 uur precies
door

Leonie Francelle Waanders

geboren op 12 augustus 1980 te Enschede

Promotores:
Prof. dr. G. J. M. Pruijn
Prof. dr. M. Mann

Doctoral Thesis Committee:
Prof. dr. L. M. C. Buijdens
Prof. dr. A. H. M. Heck (Universiteit Utrecht)
Dr. E. Lasonder

Paranymphs:
Dr. M. Vermeulen
Drs. E. Schwartz

ISBN: 978-90-9024545-4

Cover design and layout: Leonie F. Waanders
Printed by Ridderprint, Ridderkerk, the Netherlands

The research in this thesis was performed at the Max Planck Institute for Biochemistry, Department of Proteomics and Signal Transduction, in Martinsried, Germany

Publication of this thesis was financially supported by Advion BioSciences Ltd.

'Yesterday is history,
tomorrow is fantasy.
Today is a gift,
that's why it's called present!'

Kung Fu Panda

Table of contents

Abbreviations

AGC	automatic gain control
BSA	bovine serum albumin
CID	collision-induced dissociation (fragmentation technique)
Da	Dalton (unit of protein size)
ECD	electron capture dissociation (fragmentation technique)
ELISA	enzyme-linked immuno sorbent assay
ESI	electrospray ionization
ETD	electron transfer dissociation (fragmentation technique)
FDR	false discovery rate
FT	Fourier-transform
FWHM	full weight at half maximum (unit of resolution)
HCD	higher collision dissociation (fragmentation technique)
ICAT	isotope-coded affinity tag (chemical labeling)
ICR	ion-cyclotron resonance (type of mass spectrometer)
ID	inner diameter
IRMPD	infrared multiphoton dissociation (fragmentation technique)
IT	ion trap (type of mass spectrometer)
LC	liquid chromatography
LCM	laser capture microdissection
LM	lock mass
LOD	limit of detection
LTQ-FT	linear ion trap - fourier transform mass spectrometer (type of mass spectrometer)
m/z	mass to charge ratio
MALDI	matrix-assisted laser desorption/ionization
MMD	maximum mass deviation
MRM	multireaction monitoring
MS	mass spectrometry
MS/MS	fragmentation
MuDPIT	multidimensional protein identification technology
MW	molecular weight
PC	peak capacity
PCM	polydimethylcyclosiloxane (background ions)
PCR	polymerase chain reaction
PMF	peptide mass fingerprinting
ppb	parts per billion
ppm	parts per million
PTM	post-translational modification
QQQ	triple quadrupole (type of mass spectrometer)
RPLC	reverse phase liquid chromatography
S/N	signal to noise
SCX	strong cation exchange chromatography
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	stable isotope labeling by amino acids in cell culture (metabolic labeling)
SIM	selected ion monitoring
SUMO	small ubiquitin like modifier (post-translational modification)
TOF	time-of-flight (type of mass spectrometer)
UPLC	ultrahigh pressure liquid chromatography

1 General introduction

From protein to proteome

For decades, biochemists and molecular biologists have been trying to understand the structure, location, and function of biomolecules in the cell. Small-scale experiments, focused on the activity of single genes or proteins, have so far been the norm. However, in the past decades it has become increasingly clear that such targeted experiments are insufficient to tackle the whole complexity of problems posed by biological systems. In order to obtain more comprehensive knowledge about biological systems like a cell or tissue, large-scale global analyses are required.

The system-level studies of biological networks received a boost from the Human Genome Project, the consortium that first sequenced all hereditary information encoded in the DNA of a human cell. The knowledge of the complete genome led to better understanding of genes and their context, as well as creating the opportunity to study the effect of disturbances like ligands, knock-downs of selected gene products, and mutations on the global gene expression level.

However, as the hereditary information encoded in the DNA is – in the overwhelming majority of cases – executed by the interplay of transcription, transcript processing, translation and post-translational processing, studying this process at the transcriptional level alone is frequently insufficient to elucidate the phenotypic reaction of a cell to a given stimulus or disturbance. As a result the demand increased for comprehensive studies of general effector molecules of biological processes themselves: polypeptides or proteins. Several consortia were founded with the aim to analyze their localization and abundance level by e.g. immunohistochemistry and western blotting[1].

In addition to large scale analysis with antibody-based techniques, Mass Spectrometry (MS) became increasingly popular for protein expression studies. Initially, MS was mainly used to characterize purified proteins from gel-bands, but gradually it started to be applied in combination with liquid chromatography, enabling both identification and quantitation of several hundreds of proteins in a single experiment. MS thereby contributed considerably to the shift from single protein analysis to 'proteomics'. The term 'proteome' was coined in 1994 and is defined as the complete set of proteins expressed under a certain condition in a certain biological system, whether organelle, cell, tissue or organism[2].

A major advantage of MS is its unbiased nature, which is in stark contrast to techniques like western blotting or Enzyme-Linked Immuno Sorbent Assay (ELISA), where antibodies are raised against known targets. Furthermore, mass spectrometry can confirm the primary sequence and can thereby also distinguish isoforms, genetic variations and co- or Post-Translational Modifications (PTMs) of proteins that frequently impact their function. Last but not least, MS allows examining large sets of proteins simultaneously and is thereby extremely fast in analyzing complex mixtures. In short, MS has become an indispensible research tool with the potential to broadly impact biology and laboratory medicine[3].

1.2 MS-based proteomics

The increasing popularity of MS in protein research is reflected in the growing number of peer-reviewed publications in this field (red and blue line in Figure 1.1). But before MS could become ubiquitous, major technological improvements were necessary. As depicted in Figure 1.1, these advances range from the creation of sensitive and high-resolution hybrid mass analyzers and the automated coupling of Liquid Chromatography (LC) systems, to the development of new 'soft' ionization techniques. The introduction of supplementary strategies such as stable isotope labeling, as well as the use of intelligent acquisition and data analysis software complement the hardware improvements and yield improved data output and quality.

With all those improvements MS-based proteomics is maturing into a robust and high-throughput, extremely accurate and sensitive platform that can be used with a variety of aims. It is for example currently possible to catalogue all proteins of an organelle or lower organism, to obtain information on specific protein-protein interactions, or to quantify time- or stimulus-dependent protein expression or activity changes. Many of those aims can be achieved using a popular proteomics workflow for protein identification and quantitation by LC-MS, which is shown in Box A.

In the next paragraphs I will shortly review the basics of presently used instrumentation and highlight the most important instrumental improvements that have contributed to the high confidence and high accurate plat-

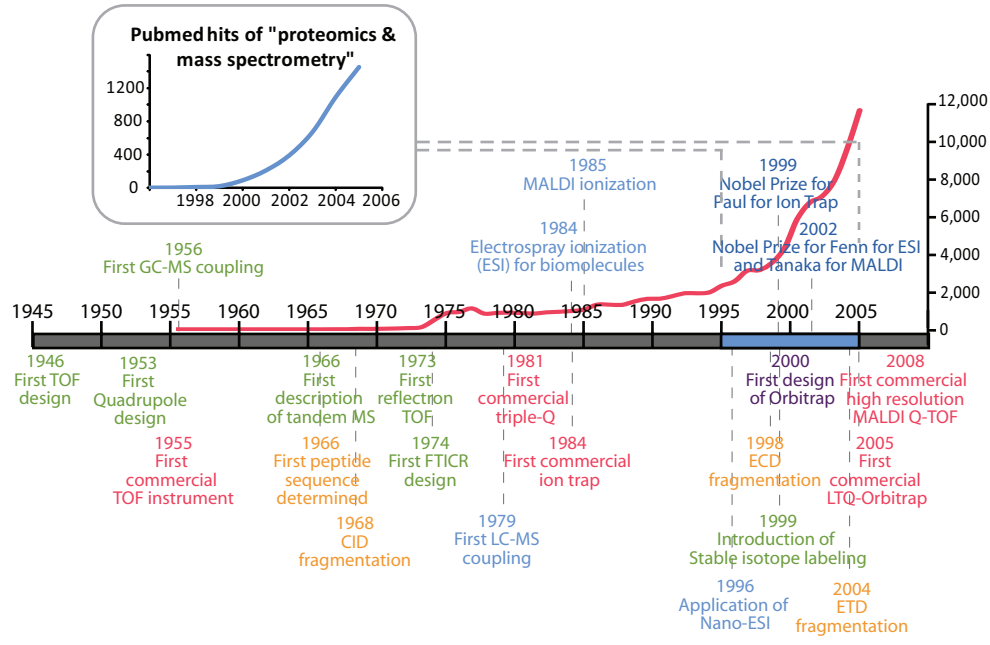


Figure 1.1 Chronological representation of major MS developments since the Second World War. The red line indicates the number of papers published and listed in Pubmed per year, using the term 'mass spectrometry' (retrieved via http://dan.corlan.net/cgi-bin/medline-trend?). The inset shows the number of peer-reviewed publications per year related to the search term 'proteomics and mass spectrometry'.

form of liquid chromatography tandem mass spectrometry (LC–MS/MS). Additionally, I will discuss four essential MS parameters and examine to what extent current instrument configurations fulfill the requirements posed by biological experiments. In the last part I will highlight a few research areas where further advances are required to expand MS capabilities. Those areas motivated the method developments I carried out during my doctoral study and which I will discuss in the succeeding chapters.

1.3 MS instrumentation

The basic idea behind mass spectrometry is to obtain information about the chemical composition of a compound by accurately determining the mass-to-charge ratio (m/z) of the analyte in ionized form. The charged condition is essential for accurate guidance, ion separation in electric and magnetic fields as well as detection of the ion. A mass spectrometer consists of three parts: 1) an ion source to transform the compound into gaseous ions, 2) a mass analyzer that carries out the actual ion separation, and 3) a detector that reads out the abundance of each ion and converts this signal into a mass spectrum.

Ionization sources

The analysis of biological samples was long hampered by the lack of good ionization sources. The chemical ionization methods existing since the 1960s were too harsh for the molecules involved and induced extensive fragmentation. The solution came in the 1980s in the form of electrospray (ESI) developed by Fenn and coworkers[4,5] and a complementary technique called Matrix-Assisted Laser Desorption/Ionization (MALDI) published by Hillenkamp and Karas[6,7]. Both ESI and MALDI, schematically represented in Figure 1.2, enabled ionization

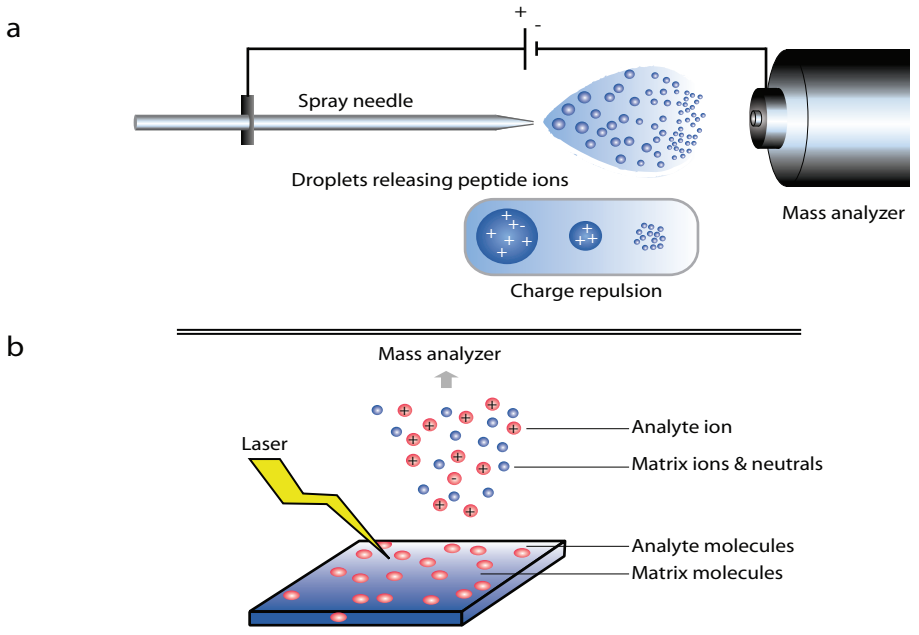


Figure 1.2 Schematic representation of ESI and MALDI, adapted from ref. [8].

of large biomolecules like proteins without substantial decomposition. In 2002, both inventions were awarded with the Nobel Prize in Chemistry, acknowledging their important contribution to chemical and biological research.

As shown in Figure 1.2a, ESI produces gaseous ions using a needle to spray a mist of tiny droplets containing charged analytes into a strong electric field. While the neutral solvent evaporates, the droplets undergo dispersion by charge repulsion until the ions are virtually free of solvent. The nature of this process creates mainly multiply charged ions for biomolecules larger than 1000 Dalton (Da). In MALDI, analytes are co-crystallized with a chemical matrix before being ionized by a laser beam, as is presented in Figure 1.2b. The matrix protects the biomolecules from the disruptive laser energy, and after ionization it transfers its charges to the analytes, resulting in predominantly singly charged ions entering the mass spectrometer.

Mass analyzers

Inside the mass spectrometer the ions are guided to a mass analyzer in which they are separated based on their m/z. In the past fifty years, various types of analyzers were developed: the first analyzer was of the Time-Of-Flight (TOF) type, followed by quadrupole, Ion Trap (IT), Fourier Transform Ion Cyclotron Resonance (FT-ICR) and finally orbitrap analyzers. Their basic designs are depicted in Figure 1.3 and main features are briefly explained below. Their differential performance in terms of sensitivity, mass resolution, mass accuracy and scanning speed will be explained and discussed in more detail below.

TOF mass analyzers separate ions spatially (Figure 1.3a). Inside the instrument, all ions are accelerated with a specific kinetic energy, inducing them to fly at a velocity proportional to their m/z. During the flight through a long tube the ions separate and are then scanned out by a detector. The m/z of every ion can be determined from its traveling time. Reflectors correct for differences in kinetic energy and longer flight tubes increase the separation efficiency and hence the resolution. Modern TOF instruments routinely achieve mass resolution of 10,000 (at present up to a maximum of 60,000) and mass accuracies in the low parts per million (ppm) range.

In quadrupole mass analyzers ions travel in between four parallel rods in an oscillating electric field (Figure 1.3b). Mass spectra are acquired by tuning the electrical field such that only ions with a specific m/z do not collide with the rods and reach the detector on a stable trajectory. Quadrupole instruments are very selective and relatively fast, but limited in resolution and mass accuracy. The maximum resolution, largely

Box A Popular proteomics workflow

Many years of development have led to a widely used workflow for the identification and quantitation of proteins in complex mixtures by MS in a high-throughput fashion. Typically the actual MS analysis is performed on peptides rather than on intact proteins. The reason for this is that peptides behave more homogeneously in LC, are easier to solubilize and result in more easily interpretable mass spectra than the proteins they derive from. Furthermore, sequencing a peptide is generally sufficient to identify a protein. Therefore, the protein or protein mixture is typically digested into peptides

by sequence-specific endoproteases and the MS measurement is performed on peptides, which are afterwards linked to the originating proteins. For this reason this MS workflow is commonly described as bottom-up or shotgun proteomics[9,10].

The basic workflow is depicted in Figure 1.A. Starting from tissue or cell material, the proteins need to be extracted by mechanical or chemical methods. Solubilization of as many proteins as possible and removal (or shearing) of DNA are imperative. The second step is meant to reduce the sample complexity – often at the expense of increasing MS measurement time. Fractionation can be done by a plethora of biochemical methods like 1D (or 2D) gel electrophoresis, cation exchange chromatography or gradient centrifugation. Depending on the sample complexity and the biological question, the fractionation can be extensive or can be omitted entirely. Subsequently, proteins are digested to peptides by site-specific proteases. The most commonly used protease is trypsin, cleaving C-terminal of arginines and lysines, since the resulting peptides are of similar length, generally well-ionizable and fragment easily in MS. Acidified tryptic digests carry typically one charge at the protonated N-terminal amine and one at the protonated R or K side chain at the C-terminus[11].

The resulting peptide mixture is subsequently further separated by C18-Reverse Phase Liquid Chromatography (RP-LC) (discussed in paragraph 1.5). Eluting peptides are ionized (see paragraph 1.3) and guided into the mass spectrometer. There, first survey MS spectra are recorded in which the signal intensity is plotted against the mass to charge (m/z). To obtain information about the peptide sequence, the ions are subsequently isolated and fragmented, yielding fragmentation spectra, also referred to as MS/MS or MS^2 spectra. The standard technique to achieve fragmentation is Collision-Induced Dissociation (CID). This and other types of peptide fragmentation methods will be discussed in paragraph 1.3.

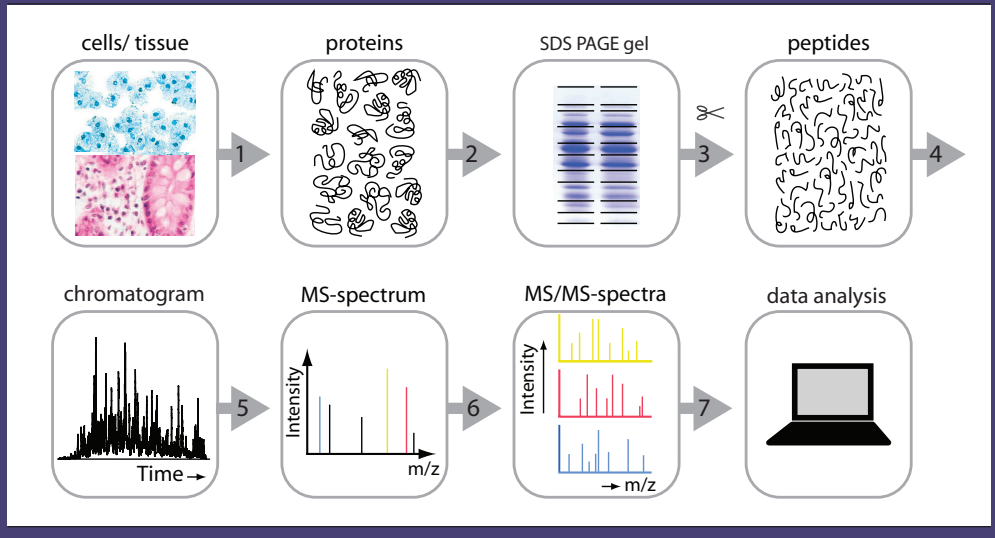


Figure 1.A Schematic workflow of a typical proteomic experiments designed for protein identification or quantitation. The basic steps are (1) protein extraction, (2) protein fractionation or enrichment, (3) digestion into peptides, (4) peptide fractionation and separation, (5) MS and (6) tandem MS spectra acquisition, and (7) data analysis by database searching.

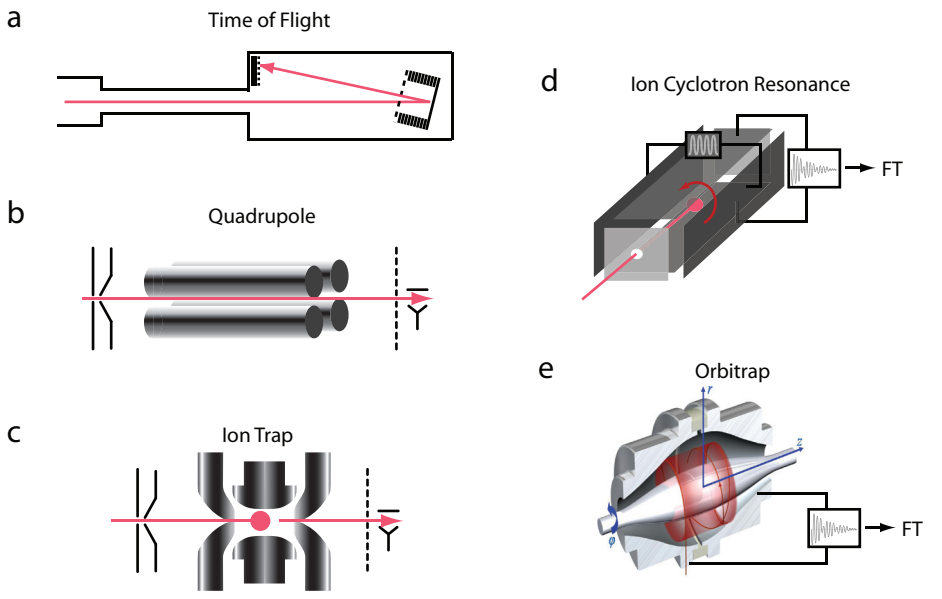


Figure 1.3 Schematic representation of the different mass analyzers used for proteomic research. Modified from ref. [12] and [13].

determined by rod length, is typically 1000[14] and mass accuracy is in the range of 0.5 Da.

Ion trap (IT) mass analyzers[15] function by the same physical principles as quadrupole instruments but – as the name implies – they are capable to confine and thereby accumulate ions (Figure 1.3c). Mass spectra are acquired by manipulating the potentials involved in a manner that successively destabilizes ions of specific m/z. Electron multiplier detectors, placed on either side of the rods, subsequently sense the ions ejected from the trap. Since the ion signal is multiplied by electron emission this analyzer type is extremely sensitive. Its scanning speed is another advantage, but also here the resolution and accuracy are comparatively low.

The idea of trapping ions has been further optimized in the Fourier transform mass analyzers, such as the ion cyclotron resonance (FT-ICR)[16] and the orbitrap[17]. In both instrument types, ion packets are cycling in orbits inside a detector cell, thereby generating repeated image currents induced by passing close to a pair of detector plates. (Figure 1.3d and e). The transient signal – a superposition of sine waves – is converted into a mass spectrum by application of a fast Fourier transformation. Apart from their geometic differences, the FT-ICR and the orbitrap mass analyzer differ in the way the oscillation of ions is induced. In FT-ICR instruments a so-called 'Penning Trap' cell is placed inside a large superconducting magnet and ions therefore move in a uniform magnetic field. The m/z of every ion is then directly proportional to its cyclotron frequency (recently reviewed in ref. [18]).

In the orbitrap analyzer the magnetic field is replaced by an electric field, where electrodes guide the ion motions in the orbitrap cell. The central electrode attracts the ions and thereby counterbalances the

Table 1.1 Main characteristics of modern (hybrid) mass spectrometers. Adapted from ref. [20].

	IT	Q-Q-Q	Q-Q-TOF	TOF-TOF	FTICR	IT-FTICR	IT-Orbitrap
Mass accuracy	Low	Low	Good	Good	Excellent	Excellent	Excellent
Resolving power	Low	Low	Good	Good	Very high	Very high	Very high
Sensitivity (LOD [1])	Good	High	High	High	Medium	Medium	Good
Dynamic range	Low	Good	Medium	Medium	Medium	Medium	Medium
ESI	Yes	Yes	Yes	No	Yes	Yes	Yes
MALDI	(Yes)	No	(Yes)	Yes	(Yes)	(Yes)	(Yes)
MS/MS capabilities	Yes	Yes	(Yes)	Yes	Yes	Yes	Yes
Additional capabilities	MS^n	Neutral loss, MRM [2]	Ion mobility	–	ECD, IRMPD [3]	ECD, MS^n IRMPD,	ETD, HCD, PQD, MS^n
Identification	+	+	++	++	+++	+++	+++
Quantitation	+	+++	+++	+++	++	++	++
Speed	+++	++	++	+++	++	+++	+++
Detection of PTMs	+	+	++	++	++	+++	+++

[1] LOD = Limit Of Detection, [2] MRM = Multiple Reaction Monitoring, [3] IRMPD = InfraRed MultiPhoton Dissociation

centrifugal force caused by the orbiting ions. The outer electrodes register the axial oscillations along the central electrode, from which the m/z of the ions can be determined[17].

In terms of resolution, Fourier transformation analyzers perform better than any other mass spectrometer. This is inherent to their detection principle: the resolving power is proportional to the number of orbits during data acquisition, of which there are many thousands[17,19]. This also explains why FT instruments can measure with various resolutions, which are selectable through the scan acquisition time. Drawbacks of these analyzers are the relatively long scanning times (up to 1 s) and the relatively low sensitivity: in complex mixtures they require millions of ions per scan.

Hybrid mass spectrometers

For accurate readout, the TOF, FT-ICR and the orbitrap require pulsed ion injection. Ion traps and quadrupoles can be used to select and/or accumulate ions of specific m/z and subsequently transfer them to a secondary analyzer selected from the group above. The creation of 'hybrid' instruments also allows optimal exploitation of the complementary features of the different analyzers. Today, practically all instruments on the proteomics market are hybrid mass spectrometers. Well known types are the triple-quadrupole (QQQ), Q-TOF, TOF-TOF, linear ion trap (LTQ) FT-ICR or LTQ-FT and the LTQ-Orbitrap analyzers. Table 1.1 provides an overview of the currently most advanced instruments and their characteristics.

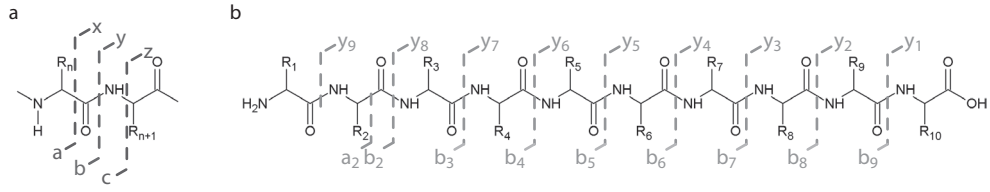


Figure 1.4 Roepstorff–Fohlmann–Biemann nomenclature for assigning fragment ions. Assignment is based on the cleavage site: (a) N-terminal fragments are named a-, b- or c-ions and C-terminal ions are x-, y-, or z-ions. (b) CID fragmentation leads predominantly to b- and y- ions. The location of the bond breakage with respect to the precursor backbone is indicated by the amino acid number after which fragmentation occurred.

Fragmentation types

Information about the chemical structure of an analyte is obtained by gas phase fragmentation inside the mass spectrometer. Typically, a single ion species is isolated and fragmented by collision with neutral gas molecules. This technique is named Collision-Induced Dissociation (CID)[21]. CID is a relatively slow process in which the protons and the energy can reposition prior to fragmentation, thus leading to predominant breakage of the weakest bonds. In normal peptides these are amide bonds, resulting in so-called b- and/or y-ions, depending on which side retains the proton (see Figure 1.4). In case of extremely labile peptide bonds – like N-terminal to proline or C-terminal to aspartate – the resulting b- or y-ion can take up to 80 to 90% of the total ion current. Such extremely abundant fragments often hamper the peptide identification (see Box B), because besides those ions not much else about the sequence is revealed. The same problem also occurs with peptides containing labile Post-Translational Modifications (PTMs) like phosphorylation or glycosylation, which tend to predominantly lose the phosphate or sugar moiety and hence lead to information-poor spectra.

To improve identification of peptides with labile modifications, two other fragmentation techniques were developed that produce MS/MS spectra complementary to CID (reviewed in [22]). These methods are based on the interaction of multiply charged polypeptides with free electrons (Electron Capture Dissociation or ECD) or with fluoranthene radical anions (Electron Transfer Dissociation or ETD). ECD and ETD mainly induce backbone breakage between C_α and N, resulting in c- and z-ions. These fast reactions involve little energy randomization and therefore leave peptide modifications untouched[23].

As the ionization efficiency is lower, ECD and ETD are slightly less sensitive than CID, especially for doubly charged ions. But since they produce a wider range of peaks of similar intensity, the information obtained from the spectra can be much higher. Hence, ETD/ECD fragmentation is very suitable for *de novo* sequencing (see Box B) and for analysis of large peptides or intact proteins[22].

Box B Peptide identification

For peptide identification, the total mass and a (ideally complete) set of fragment masses should be known. The mass of a peptide can be calculated by multiplying its m/z value with the charge state and subtracting the added proton masses. In rare cases, the intact peptide mass is sufficient for identification, but – as will be explained in paragraph 4.2 – often multiple peptides exist with the same nominal mass, therefore usually fragment spectra are needed to elucidate which peptide is detected.

There are two principle approaches for peptide identification, shown in Figure 1.B. The first approach (Figure 1.B–a) starts from fragmentation spectra and uses the knowledge that the predominant cleavages occur at defined positions along the peptide backbone. The difference between fragment ions should therefore be equal to the m/z of one or more amino acids. 'Sequence tags'[24] can be composed by reading out combinations of such fragment ion pairs. Together with the total peptide mass the sequence tags can identify the peptide in protein sequence databases like the International Protein Index (IPI). When the spectrum quality is so high that a complete sequence can be 'read out', it is called '*de novo* sequencing'.

Less spectrum quality-dependent and therefore more commonly applied is the second approach (Figure 1.B–b), in which the intact and fragment masses are compared to a peptide library, composed of peptides created by *in silico* digestion of all known proteins. Filtering this library with the detected intact peptide mass results in a list of candidates that are subsequently validated by matching the fragmentation data with the hypothetical MS/MS spectra. For this approach various complementary search engines exist, like Sequest, Mascot, Proteinprospector, and Xtandem – mainly differing in scoring algorithm.

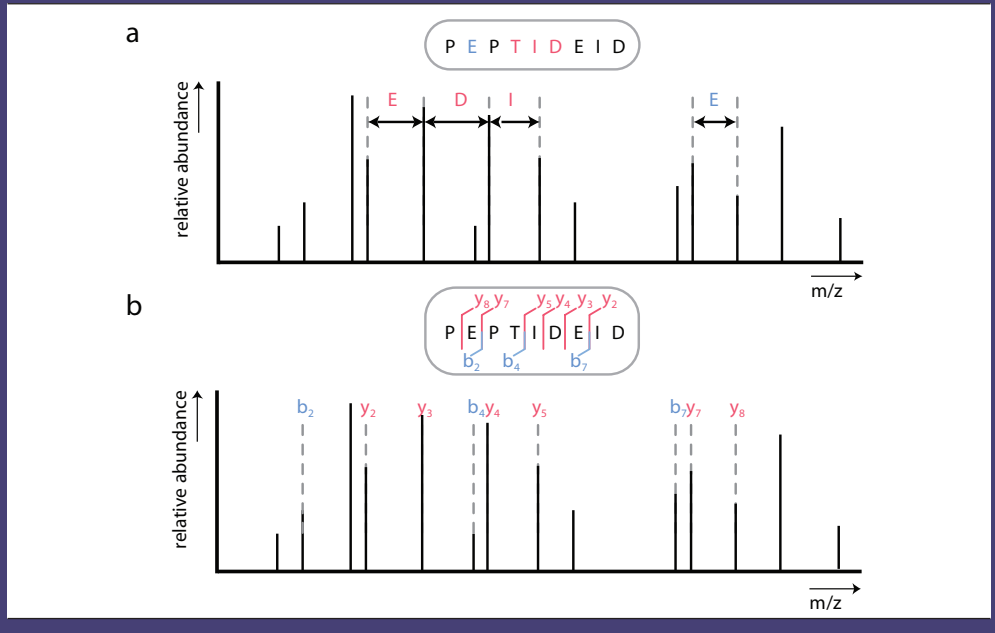


Figure 1.B Main peptide identification methods: (a) Finding 'amino acid sequence tags' in MS/MS spectra or (b) correlating match between real and theoretical spectra derived from *in silico* digested proteins. Adapted from ref. [8].

1.4 Important MS characteristics for proteomics

When mass spectrometry became available for protein analysis – with the advent of soft ionization techniques, as discussed in paragraph 1.3 – it posed new challenges to MS instrumentation. Since proteins, even when digested to peptides, are much larger than most other chemical substances, the instruments needed to enlarge their resolution and mass accuracy. Moreover, higher sensitivity and faster scanning speed were desired for the analysis of complex samples. In the following paragraphs I will explain these essential MS characteristics and discuss to what extent they have achieved the posed requirements.

The power of high resolution

The term 'resolution' describes the minimal distance between two signals that allows their discrimination. In the field of MS, mass resolution or mass resolving power is therefore defined as

$$\text{Resolution} = \frac{\text{mass}}{\Delta\text{mass}} \qquad (1)$$

As a measure of resolution, typically the Full Width at Half Maximum (FWHM) of a specific mass-to-charge (m/z) is given, as shown in Figure 1.5a. A higher resolving power results in sharper peaks and reveals more details – from charge states and isotope patterns to relative isotope abundances (Figure 1.5b)[25].

In proteomics, high-resolution mass

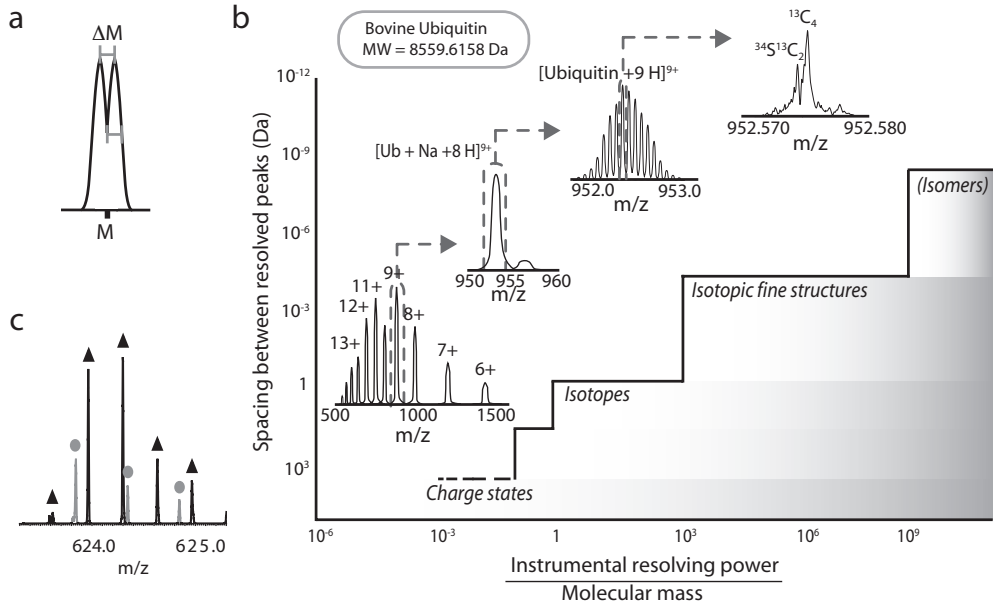


Figure 1.5 Definition (a) and impact of mass resolution. (b) An increase of instrumental resolving power enlarges the spacing between the peaks and boosts the information content. Depending on the level of resolution, charge states, adducts, isotopes or isotopic fine structures can be observed. (c) Typical region of a proteomics spectrum, with multiple peaks in the same nominal mass range. Part b adapted from [25].

analyzers are important, because often peptides of the same nominal mass co-elute (see Figure 1.5c) and might otherwise go unnoticed. Furthermore, a high resolving power raises the Signal-to-Noise ratio (S/N) and thereby increases sensitivity and quantitation accuracy.

What resolution is sufficient depends on the biological question as well as the nature of the sample and its complexity. In standard proteomic experiments where proteins are digested to peptides (as explained in Box A), peptides typically obtain two to four charges with ESI and a resolving power of about 5000 would suffice to determine the charge state and allow calculation of the total mass. However, the resolution also influences the mass accuracy (see following paragraph) and the distinction of co-eluting peptides with similar m/z. Hence, the highest possible resolving power is typically applied – up to 50,000 – 100,000 in FT instruments.

Highly accurate mass determination

Mass accuracy is defined as the deviation of the instrument reading from a known calculated mass. This parameter, which becomes better with higher resolution, is typically expressed relative to the calculated mass in parts per million (ppm) or parts per billion (ppb), or in absolute units in Dalton (Da).

In proteomics, mass accuracy is a major filter criterion for peptide identification. In databases containing all proteins of a specific organism, many peptides of similar mass exist. By applying stringent mass accuracy criteria the number of the peptide candidates can be appreciably reduced (see Figure 1.6) and this lessens the risk of miss-identification. Accurate mass determination is also essential for characterizing PTMs, since numerous modifications or amino acid substitutions induce similar mass shifts. Classical examples of PTMs that are difficult to differentiate are trimethylation and acetylation differing by only 36 mDa and phosphorylation and sulfation differing by 10 mDa.

To improve mass accuracy spectra are typically recalibrated; either by spiking in internal calibrants during acquisition or by post-acquisition spectra correction based on confidently identified peptides and/or charge state information. Current FT instrumentation can obtain average deviations in the range of 200 to 300 ppb after recalibration, which is sufficient to determine the elemental composition of most peptides smaller than 800 Da[26]. In comparison, TOF analyzers reach accuracies in the low ppm range but quadrupole and ion trap instruments are about two orders of magnitude less precise. Although a further 10-fold mass accuracy improvement would help to also determine the chemical composition of larger peptides and ones with lower S/N, much more does not seem necessary. For confident identification assignment of peptides accurate mass alone will remain insufficient, because many peptides have counterparts with an equal elemental composition but different amino acid composition (isomers) or order. According to calculations of He et al., there is a 50% chance to pick the right peptide if all but three residues are known[27]. Even more isomers are to be expected if PTMs are considered additionally.

For peptide identification it is generally sufficient to know the precise (intact) peptide mass and a set of less exact fragment masses. Consequently, fragmentation spectra can be acquired in low-resolution mass analyzers that have much faster scan speeds and higher sensitivity. In some hybrid instruments the fragmentation scans are even recorded in parallel to the high resolution and accurate survey scans. Obtaining more accurate fragmentation data is particularly desired, if (1) the protease specificity is unknown;

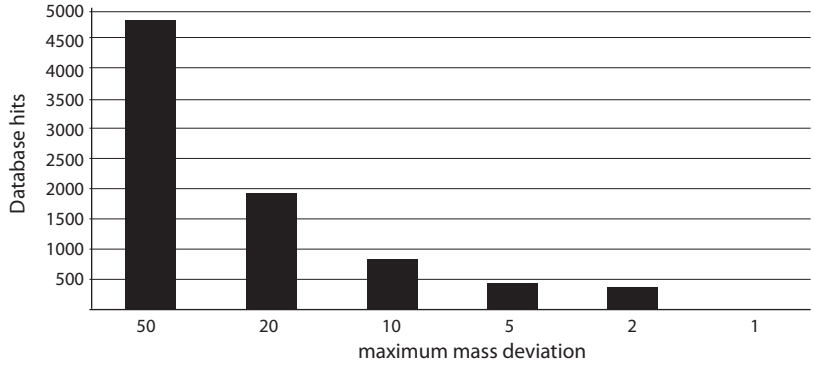


Figure 1.6 Number of candidate peptides returned from the NCBI-nr protein database upon searching a peptide mass (of 1190.7132 Da), demonstrating the filtering function of high mass accuracy. From [28].

(2) the identification is based on *de novo* sequencing or (3) the fragments are large and highly charged, as happens when analyzing large peptides or intact proteins. In those cases the assignment confidence can be improved considerably by recording the MS/MS spectra in high-resolution mass analyzers that allow charge state determination of all fragments and ppm range mass accuracies.

High sensitivity and dynamic range

A third essential MS parameter for proteomics is sensitivity, typically described as Limit Of Detection (LOD) referring to the minimal number of ions required for a conclusive signal. The importance of this parameter is evident: identification is only possible if the ion is visible. To analyze low abundant ions, the ionization process, the transmission and the detection of ions need to be highly efficient.

The ionization process efficiency depends on the peptide sequence. Aebersold and coworkers found that well-ionizing peptides contain on average more histidines, have a higher total or average charge and are relatively hydrophobic[29]. Although the inherent peptide ionization efficiency can not be changed, a reduction of the sample complexity may however influence the ionization positively. This is thought to be due to peptides competing for available protons, and this 'suppression effect' would be smaller in less complex samples[3].

Inside the mass analyzer, two aspects of sensitivity are important. The first is the total number of ions needed to produce an information-rich spectrum. Here ion traps are unsurpassed, partly due to their electron multiplier detectors. They can produce informative spectra from less than 500 ions, whereas the FTICR needs at least two orders of magnitude more analytes. The second essential aspect is the number of different ion species that can be observed simultaneously. With highly complex samples, as are typical for proteomic experiments, low intensity ions are likely to be masked by more abundant species. Therefore a high 'dynamic range' is desired, with dynamic range defined as the maximum detectable intensity difference between the ion species. For most modern instruments dynamic range is in the order of 10^3 or 10^4 per scan, a number typically limited in complex ways by the analyzer. Measuring larger ion populations in the analyzer may improve the dynamic range, but simultaneously induces 'space charging',

a mutual ion repulsion that negatively affects mass accuracy.

Since the dynamic range of complex biological samples may exceed 10^{10} to 10^{12}, it is evident that without additional adjustments proteins of the lowest abundance will not be detected, let alone be quantified. Unfortunately, many regulators of disease specific cellular processes, that could be interesting biomarkers, are expressed in very low copy numbers per cell, or trace concentration in body fluids, and are therefore difficult to identify (Figure 1.7).

The most common strategy to improve identification of low abundant species is to reduce sample complexity by separation or specific enrichment. LC is currently the standard separation method used in combination with MS, but also other techniques like gel electrophoresis, isoelectric focusing or strong cation exchange chromatography are applied to achieve further fractionation.

Aside from separation, special methods exist to improve the sensitivity or dynamic range of mass spectrometers. An interesting example is the so-called Multi Reaction Monitoring (MRM) technique in triple quadrupole (QQQ) instruments. In this setup, the first and third quadrupole of the instrument are preset to only detect specific pairs of precursor peptide and fragment ions, while the second quadrupole serves as a collision cell for fragmentation. Thus, only preselected ions are detected while the rest of the peptide signals are ignored. This high

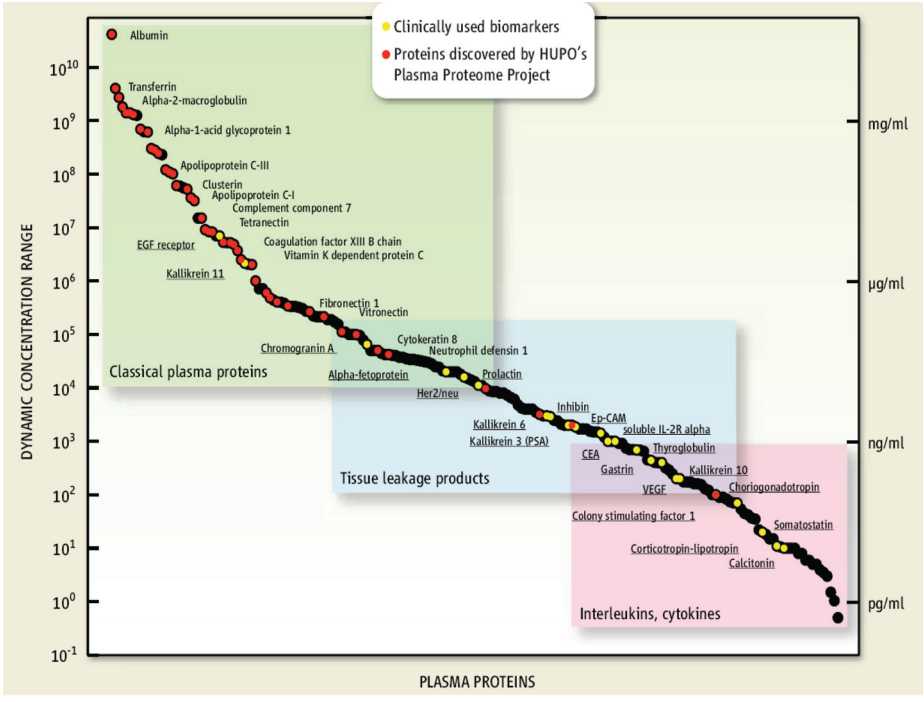


Figure 1.7 Dynamic range of proteins in blood plasma, demonstrating more than 10 orders of concentration difference between the most and least abundant proteins. Many clinically used biomarkers are expressed at very low levels. From ref. [30].

selectivity renders the QQQ a very sensitive instrument. In the LTQ-FT and LTQ-Orbitrap the dynamic range can be improved by so-called 'selective ion monitoring' or SIM-scanning, where a narrow m/z range is monitored. The analyzer is in this case only filled with a subpopulation of ions, boosting the intensity for and the S/N ratio of ions in that mass range. Other options are: 1) to selectively destabilize abundant ions in the ion trap in order to allow more low-intensity ions inside the analyzer; (2) to scan alternating with high and low numbers of ions (in which the scan with fewer ions is used to calculate improved mass accuracy of the intense species) as is done in TOF instruments; or (3) to improve the mass analyzer geometry to enable higher field strength. Unfortunately, no option is trivial or generally applicable. It is therefore very likely that the limiting dynamic range will increasingly become a bottleneck in MS analysis[31] and future efforts should focus on improving this MS characteristic.

Fast scanning instruments

For confident peptide identification, the mass spectrometer should attempt to obtain sequence data for every detectable ion. To do this automatically most instruments use a data-dependent precursor selection, in which the most intense peptide peaks are isolated in turn, based on the previous MS scan and sequenced. However, due to the co-elution of ions in most LC-MS experiments the mass spectrometer can currently not scan fast enough to sequence all peptides in a single run. This limitation is commonly referred to as 'undersampling'[20]. Undersampling is at present – together with the limited dynamic range – one of the main issues in MS, because many peaks are not targeted for fragmentation and thereby 'lost'. Understandably, many scientists are trying to overcome this problem. One method is to run the samples multiple times thereby increasing the chance that all peptides are at least sequenced once by semi-random peak selection[32]. Subsequently the identifications from the various measurements are combined. Another approach is to use information of previous runs and transfer peak identities from them, solely based on accurate mass, intensity and retention time (without MS/MS)[33]. Yet another approach to this problem is to employ algorithms targeting specifically those ions in a second run that were not sequenced in the first run[34].

A completely different way of dealing with undersampling is to fragment all ions at once and collect alternating full and fragmentation scans. With this method many data points are generated per second, which allows accurate matching of the MS and MS/MS profile to assign the fragments to the intact peptide ions based on the overlapping elution time profile[35].

Which will be the best strategy in the long-term remains unclear. MS vendors are attempting to build faster scanning instruments and researchers are developing smarter software to ensure for instance that peptide peaks are only selected once, that known contaminants are ignored and to check that MS/MS spectra contain sufficient information to result in identification.

1.5 Developments in liquid chromatography

Since 1969, Reverse Phase Liquid Chromatography (RPLC) has been combined with MS[36]. The chromatographic separation is based on hydrophobic interactions with a stationary phase, consisting of hydrocarbon chains (predominantly octadecyl, C18) covalently bound to silica particles. In proteomics, peptides (or proteins) are loaded onto RP-columns under aqueous, acidic buffer conditions and are sequentially eluted by an increasing percentage of organic solvent like acetonitrile or methanol. Analytes elute in order of increasing hydrophobicity in linear or segmented gradients[37].

LC can readily be coupled to MS, because the peptides are already solubilized upon separation and need only to be ionized after the separation. By adding an ESI emitter to the end of the column, the eluting peptides are electrosprayed and subjected to direct MS analysis. This is often referred to as 'on-line LC-MS'. A typical nanoLC-MS setup is displayed in Figure 1.8.

Peak capacity

The chromatography performance is defined by the separation power or Peak Capacity (PC) and the sample capacity of the column. The PC is an estimate of the maximum number of peaks that can be separated in time on a given column. Under the premise that the peak width does not change as a function of retention time, PC is equal to:

$$PC = 1 + \frac{T_{grad}}{4\sigma} \qquad \text{(from ref. [38])}$$

with T_{grad} being the total gradient time and σ the average peak width standard deviation[38]. $4 \cdot \sigma$ is generally considered adequately separated. The PC is proportional to the amount of interaction possibilities between the sample and the column material and is therefore directly related to the column length and the particle size, as reviewed in ref. [39]. As the flow inside the column is predominantly unidirectional, the width of the column is theoretically does not influence its separation power. This is in contrast to the sample capacity, which

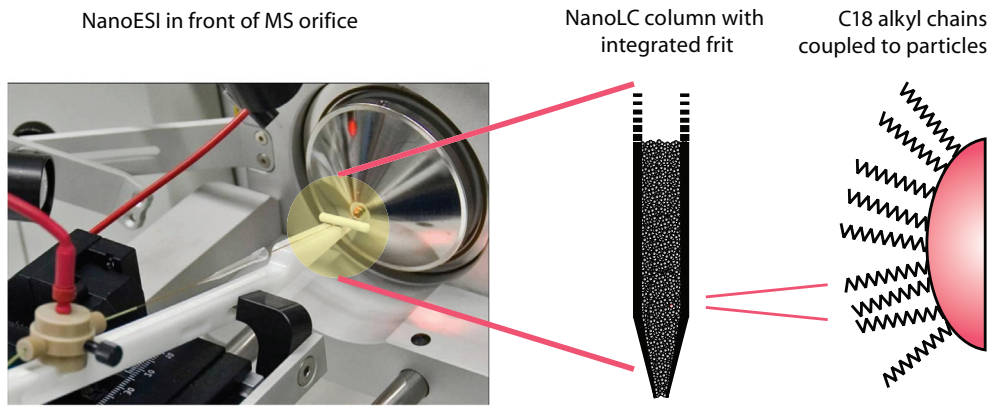


Figure 1.8 Typical nanoLC-MS setup, with a column placed in front of the mass spectrometer orifice. The column is packed with small particles that are retained inside by a polymerized frit, or, as shown here, with an integrated emitter[40]. The silica particles within the column are covered with covalently bound C18-alkyl chains, with which the peptides interact via hydrophobic interactions.

is proportional to the Inner Diameter (ID). Overloading a column can result in drastic peak broadening and the sample amount should therefore match the column size.

Increased sensitivity by nanoLC

LC can also be used for sample concentration purposes. Typically, peptides elute in several seconds and the elution volume is usually much smaller than the original sample volume. Given that ESI is a concentration dependent process, with the signal intensity directly proportional to the analyte concentration, the LC separation results in higher peaks compared to direct sample injection.

The concentration dependence of ESI has also stimulated the miniaturization of LC to nanoLC. In smaller columns peptides elute in approximately the same time, but in much smaller volumes. The gain in sensitivity due to the decrease in column size is proportional to the square of the ID ratio and approximately inversely proportional to the flow rate[41] as can be seen in Table 1.2. Smith and coworkers demonstrated this correlation experimentally[42].

Currently, most laboratories use columns of approximately 10 to 30 cm length, 50 to 100 µm ID and, packed with 3 µm particles, which are run at flow rates in the range of 200 to 300 nL/min. Under those conditions, HPLC pumps can deliver a stable and continuous flow and the samples can be loaded relatively fast. To reflect the scale of the solvent flow rate used, this separation is referred to as nanoLC when the column ID is between 10 and 100 µm[43]. Using flow rates lower than 500 nL/min in electrospray is also noted as nanoESI.

Further reduction of the column ID and flow rate are actively explored to gain further sensitivity. However, the detrimental effect of difficult to eliminate void volumes on chromatographic performance increases dramatically with a reduction of the flow rate and their avoidance therefore is a major focus of development.

Table 1.2 ESI sensitivity is proportional to sample concentration and inversely proportional to column ID. Column miniaturization therefore leads to significant increases in sensitivity. Adapted from ref. [41].

Column ID (mm)	Flow rate (µL/min)	Relative sensitivity increase (fold)
4.6	1000	1
2	200	5.3
1	50	21
0.5	12	84
0.2	2	529
0.1	0.5	2,116
0.075	0.25	3,762
0.05	0.12	8,464
0.025	0.03	33,856
0.01	0.005	211,600

Nanocolumns are generally constructed by packing a fritted fused silica capillary with a slurry of RP-particles The majority of columns are densely packed with small particles (of 1.7, 3 or 5 µm) under nitrogen pressure. Porous frits can be made by ceramic sintering, polymerization or by using unions containing stainless steel screens[44]. To facilitate ESI, a needle with small ID is placed behind the frit. Columns can also be made from tapered capillaries, where the tip functions as an integrated emitter. Ishihama described a method to create a frit by building an arch of particles in the tip[40]. The advantage of the latter method is that the post-column volume is minimized, thereby preventing any post-column mixing effects that could negatively affect the separation power. Instead of using particles, a column can also be produced as a monolith - an in situ po-

lymerized, continuous material with pores of different sizes. Although these columns perform well with higher flow rates, their suitability for proteomics is still not clearly established[43].

Ultrahigh pressure LC

To improve the separation power, columns can be extended in length or be packed with smaller particles, at the expense of increasing the backpressure, as shown in Table 1.3. Conventional HPLC equipment performs best at system pressures below 300 bar and thereby constrains column length and particle size. Recently, special equipment was developed to employ higher pressures in order to obtain increased separation efficiencies. So called 'ultrahigh pressure' LC (UPLC) systems can withstand pressures up to 600 bar. These instruments are increasingly used for fast screening in pharmaceutical companies and can separate peptides with reported PCs of >700 using short gradients (35 min) and relatively high flow rates (1 ml/min)[45–46].

Two dimensional LC

Another approach to increase the separation power is by applying two orthogonal LC types in sequence. The best-known of such approach was developed by Yates and co-workers and named Multi Dimensional Protein Identification Technology (MuDPIT)[47]. They placed a short Strong Cation eXchange (SCX) column in front of the column with standard RP-material, loaded a complex sample and then used fractions of different salt concentrations to transfer a subset of peptides from the SCX column to the RP-column. The standard organic gradient only elutes the RP-bound analytes which are subsequently analyzed by MS. Although this method is theoretically very powerful, its deployment in high throughput context remains difficult due to practical problems such as crystallizing salts.

Further 2D-LC strategies have been proposed, for instance the combination of size exclusion chromatography with RP-LC[48], however, none of these methods has proven to be robust or superior to 1D-LC yet.

Table 1.3 Column backpressure and separation efficiency per cm column length are proportional to particle size. As pressure constraints limit the column length, smaller particles do not necessarily result in a higher peak capacity. Adapted from ref. [41].

Particle size (µm)	Backpressure (bar/cm)	Separation efficiency (plates/cm)	At 400 bar limit		
			Max. column length (cm)	Max. number of plates	Max. peak capacity
1.7	28	294	14	4116	171
3	9.2	167	43	7181	299
5	3.3	100	121	12100	504

1.6 Accompanying technology – stable isotope labeling strategies

Until a few years ago, MS was predominantly used to identify (semi) purified proteins. With the advent of online nanoLC–MS and automated search engines, identification became more straightforward and the focus shifted to quantitation[49]. Consequently, researchers have developed various methods to compare protein abundances, for example, by counting the number of identified peptides[50], by correlating the number of identified peptides to theoretically observable peptides[51], or by comparing the peak intensity or peak area of peptide signals[52].

More accurate quantitation is possible with stable isotope strategies in MS, since the exchange of normal carbon or nitrogen atoms by their heavy analog does not affect any parameter besides the m/z readout in the mass spectrometer[49]. Peptide mixtures with different labels are mixed prior to LC–MS and the relative intensity of every peptide is directly read out as they are visible as doublets. Labeling is performed on either protein and peptide level, using metabolic or chemical strategies. Well-known chemical labeling strategies are Isotope-Coded Affinity Tag (ICAT), $^{18}H_2O$, or iTRAQ, and they attach a label respectively before, during or after digestion[53–55]. Metabolic labeling is performed by feeding cells in culture or organisms with a medium in which respectively all nitrogen atoms (^{15}N labeling) or specifically one or two amino acids (termed Stable Isotope Labeling by Amino acids in Cell culture or SILAC) are exchanged with their heavy isotope variant[56]. Protein quantitation by metabolic labeling strategies is more accurate because experimental error is minimized by mixing 'heavy' and 'light' cell populations directly after lysis and the labeling efficiency can be better controlled. Unfortunately, these metabolic methods are mainly restricted to systems that grow and proliferate, like cells or animals.

For human tissue samples only chemical labeling strategies are applicable, but most of these are not specific or are difficult to drive to completion – factors that can seriously complicate quantitation and identification[49]. Probably the best performing approach at present is iTRAQ labeling that allows comparison of four or eight samples simultaneously by use of isobaric reagents. The different reagents become visible upon fragmentation and can then be read out in the fragment mass range from 113 to 121 m/z. However, iTRAQ quantitation is not as accurate as other labeling strategies since every peptide is typically only fragmented once whereas MS intensity can be averaged over multiple survey scans.

Alternatively, one can refrain from labeling and compare the extracted ion chromatograms of the same peptides measured in consecutive runs. This type of quantitation is becoming relatively accurate and is applied under the name 'label-free quantitation'.

In this thesis we applied both SILAC labeling and label-free quantitation.

1.7 In-depth analysis in more challenging applications

As discussed above, the rapid development of proteomics technology has enabled in-depth analysis of many types of biological samples. However, for more demanding scenarios such as proteins carrying extensive PTMs or proteins available in limited amounts only, standard workflows (like the one introduced in Box A) provide insufficient depth of analysis and alternative strategies need to be explored. Below examples of challenging research fields are described, in which MS has high potential but currently only few applications.

Intact protein characterization

PTMs often influence the structure of a protein and thereby its function. Although far from being completely understood, it is known that various PTMs are essential in signaling pathways, that they are involved in practically all cellular processes, and that slight disturbances in biological systems regulating these modifications can have remarkable impact. It was shown for example, that various diseases are initiated or influenced by altered modification levels[57-59].

In the majority of cases, MS can determine both the type of PTM and the site of modification by the effective mass difference. However, the popular protocol involving *in silico* assembly of peptides identified from a proteolytic digest (see Box A) is not suited to explore the interactions between multiple PTMs, since information about combinatorial effects is frequently lost. Here intact protein characterization can provide invaluable additional information.

Analysis of intact proteins by MS without prior digestion is commonly termed 'top-down' mass spectrometry. The difference between the earlier introduced bottom-up and top-down MS is presented schematically in Figure 1.9. Under the premise that the modified protein form is detectable, top-down MS enables characterization of all PTM combinations as well as determination of the modification sites. In addition, it can reveal the stoichiometry of the differentially modified protein species because the ionization efficiencies of modified and unmodified proteins are very similar.

The detection of distinct protein isoforms as a result of differential splicing is in princi-

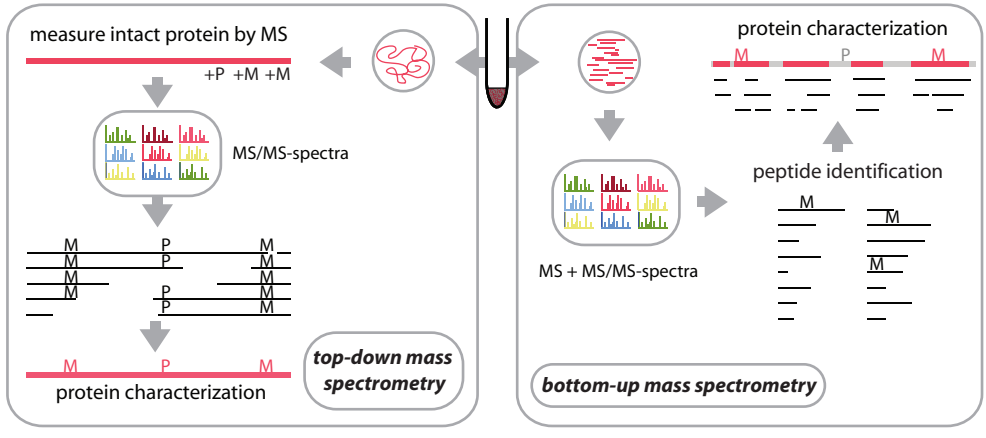


Figure 1.9 Schematic comparison of top-down and bottom-up mass spectrometry.

ple also achievable by MS on intact proteins[60], whereas isoform-identification by bottom-up approaches depends on the presence of unique peptides. The latter generally requires high sequence coverage, and sometimes there are no unique peptides due to the protease chosen or as a result of the combination of splice isoforms present in the sample.

The basic workflow for intact protein analysis (see Figure 1.10) is similar to the bottom-up approach, but lacks the proteolytic step. Protein purification is generally more extensive and, in case of pure proteins or low-complexity sample mixtures, any RP-chromatography may be omitted. High resolution and high mass accuracy are essential for top-down characterization with ESI, to accurately determine the molecular mass of the intact protein and of its fragments[31]. As the molecular mass of a protein ($M_{protein}$) is defined as:

$$M_{protein} = (m/z_{monoisotopic} \cdot z) - z \cdot M_{H^+} \qquad (3)$$

the m/z of a monoisotopic peak (composed only of ^{12}C atoms) and its charge state (z) need to be known. Charge state recognition requires only medium resolution (see Figure 1.11), but to identify the monoisotopic peak, which is typically of low intensity for proteins, high S/N and resolving power are vital. Ideally this peak is detected in the spectrum directly (Figure 1.11c), but if it the protein is so large that this isotope cannot be seen, its mass can be estimated by overlaying the spectrum with a theoretically modeled isotope distribution. Additionally, high mass accuracy is important, because small m/z inaccuracies result in considerable protein mass errors upon multiplication with the number of protein charges. As described in paragraph 1.4, high-resolution MS/MS readout is also indispensible for accurate and confident fragment ion assignment. Without charge state information the risk of miss-assignment becomes unacceptably high, since the list of theoretical fragments used to assign the observed peaks, expands with every additional charge the fragments are considered to have.

This demand for high resolution implies that top-down characterization with ESI can best be performed on FT instruments. Although the suitability of FTICR instruments has been demonstrated for many years, the LTQ-Orbitrap was first used for intact protein analysis in our study described in Chapter 2.

Characterization of extremely small sample amounts

A completely different, but equally challenging research area is the field of clinical proteomics, in which patient material is studied to discover new therapeutic targets as well as novel diagnostic or prognostic markers. Unfortunately, human samples are extremely heterogeneous, usually difficult to

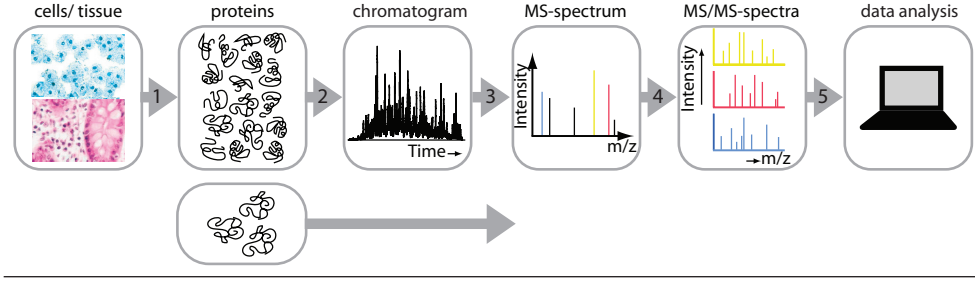


Figure 1.10 Workflow for intact protein (LC-) MS analysis. Steps are (1) protein extraction and (2) separation, (3) MS and (4) MS/MS acquisition and (5) data analysis. Depending on protein purity step 2 and 3 can be omitted.

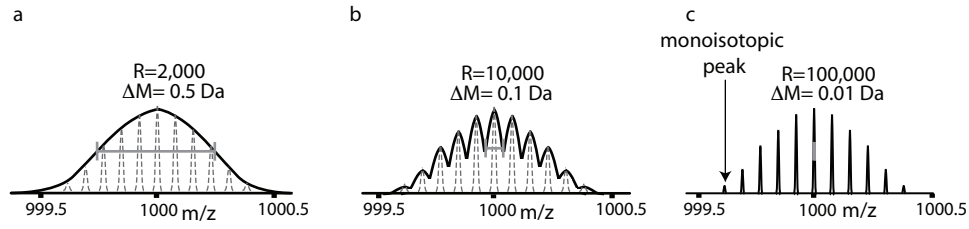


Figure 1.11 Importance of high resolution for identification of intact proteins. (a) A resolving power of 2000 FWHM is insufficient to distinguish individual peaks of the isotope cluster; (b) 10,000 FWHM enables charge state recognition; and (c) 100,000 FWHM additionally results in accurate mass determination.

obtain in vivo and often limited in amount. This stands in sharp contrast to fundamental biological research where cells of a single type are usually grown in culture dishes under well controlled and reproducible conditions. Such cells can be harvested with relative ease in large quantity to study fundamental processes like proliferation or gene transcription. However, a disadvantage of cultured cells is that they may not accurately represent molecular events taking place in an intact organism. As Pan et al. recently demonstrated on the level of the proteome, long-term selection for fast cell proliferation induced significant metabolic changes compared to primary cells[61].

Although animals can be used to model humans in health and disease – addressing some limitations of cell culture systems – animal studies are also often restricted in terms of clinical use due to differences in longevity, exposure to different external influences, and different genetic predispositions. As a result the transfer of results from animal models to the actual human disorders often remains challenging.

Thus, many studies of medical relevance have no alternatives to the use of human samples. However, this comes with great heterogeneity, introduced by individual genetic variability (in addition to gender, ethnicity, etc.) and environmental factors like diet, climate, and health. To deduce convincing functional trends large cohorts of patients need to be screened. Furthermore the data need to be of sufficient depth to detect more than typical housekeeping proteins.

When analyzing body fluids (blood, saliva, cerebrospinal fluid, or urine) it is generally possible to collect sufficient material. However, when tissue samples are taken as biopsies, the protein amounts are often extremely low, preventing an in-depth analysis using standard protocols. Sample scarcity is predominantly a problem for protein analysis, since signal from RNA or DNA molecules can effectively be amplified by the well-established polymerase-chain reaction (PCR) technique and in principle 100 – 1000 cells are sufficient[62]. Proteome analyses currently need at least 30,000 – 100,000 cells, which correspond to approximately 3–10 microgram of proteins. Only recently attempts with lower amounts of cells (3,000) were published[63], in which 1003 proteins were identified by matching their retention time and mass to previously acquired samples.

To improve the analysis of those clinically relevant samples, sample losses during preparation steps need to be minimized and the LC-MS sensitivity needs to be increased. If in-depth proteomes could be recorded with small sample amounts and little or no fractionation, virtually unlimited possibilities emerge in the combination of histology and pathology and proteomics.

Outline of this thesis

The aim of this thesis was to develop and optimize LC-MS strategies to accurately characterize samples to more depth than possible hitherto and thereby expand the field of MS-based proteomics to novel applications. We chose three categories of proteins or proteomes that were challenging for different reasons:

1. Intact proteins characterization and quantitation

2. High speed characterization of low-complexity protein mixtures

3. Quantitative proteomic analysis from extremely small sample amounts

We start by investigating the capabilities of the (in 2005) novel LTQ-Orbitrap for top-down protein analysis. As described in Chapter 2, we observed that this type of mass spectrometer is extremely accurate, highly sensitive and well suited for direct characterization of small proteins and for localizing their modifications. In Chapter 3 we continued this analysis by evaluating whether SILAC labeling can be applied in top-down MS experiments to provide quantitation to this approach. Modeling demonstrated that isotope clusters of differentially labeled proteins would not overlap in survey spectra and we proved this experimentally by measuring a protein named Grb2 in heavy and light labeled states using online LC-MS. Averaged over 20 spectra, we determined the ratio of heavy versus light with a precision of 6%.

In Chapter 4 we switch to bottom-up MS and present a novel method intended to analyze simple protein (or peptide) mixtures with faster throughput (sample type 2). We revived the direct nanoESI by use of the Triversa Nanomate robot and developed a fast and in-depth characterization method with extreme sensitivity and mass precision. By use of selective ion monitoring (SIM) scans, higher m/z regions with typically low ion signals were boosted in intensity and hence resulted in more peptide identifications.

To increase analysis depth also in more complex proteomic studies, we attempted to collect fractions while running LC-MS. After a few iterations, we developed a system that we named RePlay. As described in Chapter 5, the RePlay is composed of a splitting valve and a long capture capillary in which a part of the LC-column effluent is stored. Fully automated control allows direct reanalysis of the stored effluent and thereby enables a second analysis in which information from the first run may be used.

In Chapter 6, we subsequently applied the novel chromatographic system and proved its usefulness for the analysis of minute amounts of tissue material (sample type 3). Using four hour gradients at a reduced flow rate we analyzed single islets of Langerhans. From these mini-organs composed of less than 4000 cells, we identified 2013 proteins. Additionally we quantified the protein expression levels after one day of glucose stimulation. Up- or downregulated proteins confirmed published reports on specific transcriptional or translational effects upon hyperglycemia, and in addition we found a novel proteins involved in islet response as a result of high glucose exposure.

Finally, Chapter 7 and 8 encompass the summary and recommendations regarding future work.

1.8 References

1 Ponten, F., Jirstrom, K., and Uhlen, M. (2008) The Human Protein Atlas - a tool for pathology. J Pathol 216(4), 387-393

2 Wasinger, V. C., et al. (1995) Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis 16(7), 1090-1094

3 Qian, W. J., et al. (2006) Advances and challenges in liquid chromatography - mass spectrometry-based proteomics profiling for clinical applications. Mol Cell Proteomics 5(10), 1727-1744

4 Yamashita, M. and Fenn, J. B. (1984) Electrospray ion-source - another variation on the free-jet theme. J Phys Chem 88(20), 4451-4459

5 Fenn, J. B., et al. (1989) Electrospray ionization for mass spectrometry of large biomolecules. Science 246(4926), 64-71

6 Karas, M., et al. (1987) Matrix-Assisted Ultraviolet-Laser Desorption of Nonvolatile Compounds. Int J Mass Spectrom and Ion Processes 78, 53-68

7 Karas, M. and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem 60(20), 2299-2301

8 Steen, H. and Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol 5(9), 699-711

9 Ge, Y., et al. (2002) Top-down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. J Am Chem Soc 124(4), 672-678

10 Spahr, C. S., et al. (2000) Simplification of complex peptide mixtures for proteomic analysis: reversible biotinylation of cysteinyl peptides. Electrophoresis 21(9), 1635-1650

11 Olsen, J. V., Ong, S. E., and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol Cell Proteomics 3(6), 608-614

12 Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. Nature 422(6928), 198-207

13 Scigelova, M. and Makarov, A. (2006) Orbitrap mass analyzer - overview and applications in proteomics. Proteomics 6 Suppl 2, 16-21

14 Holme, A. E., Thatcher, W. J. and Leck, J. H. (1972) An investigation of the factors determining maximum resolution in a quadrupole mass spectrometer. J of Phys E: Scientific Instruments 5, 429-433

15 Paul, W. and Steinwedel, H. (1953) Ein neues Massenspektrometer ohne Magnetfield. Zeitschrift für Naturforschung A 8(7), 448-450

16 Comisarow, M. B. and Marshall, A. G. (1973) Fourier transform ion cyclotron resonance spectroscopy. Chem Phys Lett 25(2), 282-283

17 Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. Anal Chem 72(6), 1156-1162

18 Marshall, A. G. and Hendrickson, C. L. (2008) High-Resolution Mass Spectrometers. Ann Rev of Anal Chem 1, 579-599

19 Marshall, A. G., (2000) Milestones in fourier transform ion cyclotron resonance mass spectrometry technique development. Int J Mass Spectrom 200(1-3), 331-356

20 Domon, B. and Aebersold, R. (2006) Mass spectrometry and protein analysis. Science 312(5771), 212-217

21 Hadden, W. F. and McLafferty, F. W. (1968) Metastable ion characteristics VII. Collision-induced metastables. J Am Chem Soc 90, 4745-4746

22 Zubarev, R. A., Zubarev, A. R. and Savitski, M. M. (2008) Electron capture/transfer versus collisionally activated/induced dissociations: solo or duet? J Am Soc Mass Spectrom 19(6), 753-761

23 Kelleher, N. L., et al. (1999) Localization of labile posttranslational modifications by electron capture dissociation: the case of gam-

ma-carboxyglutamic acid. Anal Chem 71(19), 4250-4253
24 Mann, M. and Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal Chem 66(24), 4390-4399
25 Balogh, M. P. (2004) Debating Resolution and Mass Accuracy. LC-GC Europe 17(3), 152-159
26 Graumann, J., et al. (2007) SILAC-labeling and proteome quantitation of mouse embryonic stem cells to a depth of 5111 proteins. Mol Cell Proteomics, 7(4), 672-683
27 He, F., et al., (2004) Theoretical and experimental prospects for protein identification based solely on accurate mass measurement. J Proteome Res 3(1), 61-67
28 Kuzdzal, S. A., et al. (2004) The Importance of Mass Accuracy in Peptide Mass Fingerprinting. Perkin Elmer Inc. [cited 2008-12-10] Available from http://las.perkinelmer.com/content/technicalinfo/tch_protofmassfingerprinting.pdf
29 Mallick, P., et al. (2007) Computational prediction of proteotypic peptides for quantitative proteomics. Nat Biotechnol 25(1), 125-131
30 Service, R. F. (2008) Proteomics - will biomarkers take off at last? Science 321(5897), 1760
31 Mann, M. and Kelleher, N. L. (2008) Special Feature: Precision proteomics - The case for high resolution and high mass accuracy. Proc Natl Acad Sci USA, 105(47), 18132-18138
32 Liu, H., Sadygov, R. G., and Yates, J. R. (2004) A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. Anal Chem 76(14), 4193-4201
33 Jacobs, J. M., et al. (2005) Ultra-sensitive, high throughput and quantitative proteomics measurements. Int J Mass Spectrom 240(3), 195-212
34 Picotti, P., Aebersold, R. and Domon, B. (2007) The implications of proteolytic background for shotgun proteomics. Mol Cell Proteomics 6(9), 1589-98
35 Wieghaus, A., et al. (2008) The Thermo Scientific Exactive benchtop LC/MS orbitrap mass spectrometer. [cited 2008-12-10] Available from http://www.thermo.com/eThermo/CMA/PDFs/Articles/articlesFile_7841.pdf
36 Esselman, W. J. and Clagett, C. O. (1969) Gas-Liquid Chromatography-Mass Spectrometry of Hydroxylated Octadecanols Derived from Hydroxylated Stearic Acids. J of Lipid Res 10(2), 234-239
37 Egas, D. A., Wirth M. J., (2008) Fundamentals of Protein Separations: 50 Years of Nanotechnology and Growing. Ann Rev of Anal Chem 1, 833-855
38 Giddings, J. C. (1967) Maximum Number of Components Resolvable by Gel Filtration and Other Elution Chromatographic Methods. Anal Chem 39(8), 1027-1028
39 Vissers, J. P. (1999) Recent developments in microcolumn liquid chromatography. J Chromatogr A 856(1-2), 117-43
40 Ishihama, Y., et al. (2002) Microcolumns with self-assembled particle frits for proteomics. J Chromatogr A, 979(1-2), 233-9
41 De Jong, A. (2007) Nanoflow LC. [cited 2008-12-10] Available from www.nanoseparations.com
42 Shen, Y., et al. (2002) High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. Anal Chem 74(16), 4235-49
43 Hernandez-Borges, J., et al., (2007) Recent applications in nanoliquid chromatography. J Sep Sci 30(11), 1589-1610
44 Cutillas, P. R., (2005) Principles of nanoflow liquid chromatography and applications to proteomics. Current Nanoscience 1(1), 65-71
45 Petersson, P., et al. (2008) Maximizing peak capacity and separation speed in liquid chromatography. J Sep Sci 31(13), 2346-2357
46 Wren, S. A. C. (2005) Peak capacity in gradient ultra performance liquid chromatography (UPLC). J Pharm Biomed Anal 38(2), 337-343
47 Washburn, M. P., Wolters, D., and Yates, J. R. (2001) Large-scale analysis of the yeast pro-

teome by multidimensional protein identification technology. Nat Biotechnol 19(3), 242–7
48 Bedani, F., Kok, W. T. and Janssen, H. G. (2006) A theoretical basis for parameter selection and instrument design in comprehensive size-exclusion chromatography x liquid chromatography. J Chrom A 1133(1–2), 126–134
49 Ong, S. E. and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. Nature Chem Biol 1, 252–262
50 Gao, J., et al. (2005) Guidelines for the routine application of the peptide hits technique. J Am Soc Mass Spectrom 16(8), 1231–1238
51 Ishihama, Y., et al. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics 4(9), 1265–1272
52 Chelius, D. and Bondarenko, P. V. (2002) Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. J Proteome Res 1(4), 317–323
53 Gygi, S. P., et al. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17(10), 994–999
54 Yao, X., Afonso, C., and Fenselau, C. (2003) Dissection of proteolytic ^{18}O labeling: endoprotease-catalyzed ^{16}O-to-^{18}O exchange of truncated peptide substrates. J Proteome Res 2(2), 147–152
55 Oda, Y., et al. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci USA 96(12), 6591–6596
56 Ong, S. E., et al., (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1(5), 376–386
57 Van Venrooij, W. J. and Pruijn, G. J. (2000) Citrullination: a small change for a protein with great consequences for rheumatoid arthritis. Arthritis Res 2(4), 249–251
58 Wood, D. D., et al., (1996) Acute multiple sclerosis (Marburg type) is associated with developmentally immature myelin basic protein. Ann Neurol 40(1), 18–24
59 Dias, W. B. and Hart, G. W. (2007) O-GlcNAc modification in diabetes and Alzheimer's disease. Mol Biosyst 3(11), 766–772
60 Siuti, N. and Kelleher, N. L. (2007) Decoding protein modifications using top-down mass spectrometry. Nature Meth 4(10), 817–821
61 Pan, C., et al. (2008) Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type specific functions. Mol Cell Proteomics 8(3), 443–450
62 Espina, V., et al. (2007) Laser capture microdissection technology. Expert Rev Mol Diagn 7(5), 647–657
63 Umar, A., et al. (2007) NanoLC-FT-ICR MS improves proteome coverage attainable for approximately 3000 laser-microdissected breast carcinoma cells. Proteomics 7(2), 323–329

2

Top-down protein sequencing and MS^3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer

Boris Macek*, Leonie Waanders*, Jesper V. Olsen and Matthias Mann

Mol Cell Proteomics 2006 (8), 1452–1459

** Shared first co-authorship*

Top-down proteomics, the analysis of intact proteins (instead of first digesting them to peptides) has the potential to become a powerful tool for mass spectrometric protein characterization. Requirements for extremely high mass resolution, accuracy and ability to efficiently fragment large ions have often limited top-down analyses to custom built FT-ICR mass analyzers.

Here we explore the hybrid linear ion trap (LTQ) – orbitrap, a novel, high performance and compact mass spectrometric analyzer, for top-down proteomics. Protein standards from 10 to 25 kDa were electrosprayed into the instrument using a nanoelectrospray chip. Resolving power of 60,000 was ample for isotope resolution of all protein charge states. We achieved absolute mass accuracies for intact proteins between 0.92 and 2.8 ppm employing the "lock mass" mode of operation. Fifty femtomole of cytochrome c applied to the chip resulted in spectra with excellent S/N and only low attomole sample consumption. Different protein charge states were dissociated in the LTQ and the sensitivity of the orbitrap allowed routine, high resolution and high mass accuracy fragment detection. This resulted in unambiguous charge state determination of fragment ions and identification of unmodified and modified proteins by database searching.

Protein fragments were further isolated and fragmented in the LTQ, followed by analysis of MS3 fragments in the orbitrap, localizing modifications to part of the sequence and helping to identify the protein with these small peptide-like fragments. Given the ready availability and ease of operation of the LTQ-Orbitrap, it may have significant impact on top-down proteomics.

Introduction

Major goals in every mass-spectrometry-based proteomics experiment are protein identification and characterization. Almost invariably, proteins are enzymatically degraded to peptides, which are much more amenable to mass spectrometric investigation[1]. Further advantages of this 'bottom-up proteomics' approach are that one protein generates many peptides, providing many opportunities to identify or quantify it. However, identified peptides rarely cover the whole sequence of a given protein often leading to difficulties in protein characterization, particularly in determination of posttranslational modifications (PTMs). In the alternative approach, termed 'top-down proteomics', intact proteins are ionized, physically fragmented and analyzed in the mass spectrometer (reviewed in [2-4]). Since this approach starts from MS-detection of the intact, fully modified protein, it has the potential for full protein characterization. Although analysis of intact proteins has been reported for almost all mass analyzers, to date only one, the Fourier-transform ion cyclotron resonance (FT-ICR) analyzer, has sufficient resolving power and mass accuracy to efficiently analyze large protein ions. In addition, several methods especially useful for fragmentation of whole proteins have

been developed for the FT-ICR analyzers, such as infrared multiphoton dissociation (IRMPD), sustained off-resonance irradiation (SORI) and, in particular, electron-capture dissociation (ECD), which is nonergodic in nature and in some cases can cleave almost any peptide bond in proteins of up to 50 kDa[5]. A similar fragmentation method, electron transfer dissociation (ETD), was recently introduced for ion traps and, in combination with charge state reduction, shows great promise for top-down proteomics on these mass analyzers[6].

Collision-induced dissociation (CID) fragments proteins efficiently in ion traps, but these mass spectrometers lack sufficient resolution to resolve large protein fragment ions and their charge states. With the hybrid ion trap-FTICR mass spectrometer (LTQ-FT, Thermo Electron) it is possible to fragment large peptides or even protein ions in the ion trap and detect them with high resolution and accuracy in the FT-ICR[7,8]. However, in our experience, and as showed here, the LTQ-FT is less suitable for detection of fragments produced in the LTQ due to lower sensitivity and time-of-flight effects. On the other hand, the ions can be fragmented in the ICR cell using methods like ECD and IRMPD. Furthermore, even though the LTQ-FT is a commercial and robust instrument, the necessity for a high magnetic field detector and relatively high maintenance costs tend to limit its use to specialized laboratories.

In 2005 a new hybrid mass spectrometer, the LTQ-Orbitrap (Thermo Fisher Scientific), was introduced[9]. It consists of a linear quadrupole ion trap (LTQ) coupled to a novel mass analyzer, the orbitrap, invented by Makarov[10-12]. In the orbitrap, ion packages circle between two concentric electrodes and their axial motion is detected, as in the FT-ICR, by recording their image currents followed by Fourier Transformation of the time-domain signal to obtain the mass spectrum. Importantly, the orbitrap is very compact and requires no magnetic field or special maintenance. LTQ and orbitrap are coupled via the C-trap, an intermediate RF-only storage device, which can also be used to store background ions of known composition. When analyte ions are added and analyzed together with this 'lock mass', sub-ppm mass accuracy for peptides is achievable[13].

No systematic study dealing with intact proteins and their fragments in the LTQ-Orbitrap has been reported so far. In this study, we have explored the utility of the LTQ-Orbitrap mass spectrometer for top-down analysis of proteins ranging in mass from 10 to 25 kDa. Our results show that the instrument is capable of routinely achieving high sensitivity (attomole to femtomole range), high mass accuracy (low ppm) and isotope resolution of small proteins. In addition, selected multiply charged protein ions can be successfully fragmented in two stages, MS/MS and MS^3, in the linear ion trap and their fragments transferred and measured in the orbitrap. We demonstrate ready identification of modified and unmodified proteins by MS/MS and MS^3 data.

Materials and methods

All protein standards were obtained from Sigma. The following proteins were analyzed: bovine cytochrome c, bovine α-crystallin A and B chain, bovine β-lactoglobulin, bovine β-casein, and recombinant human ubiquitin. Proteins were dissolved in methanol/water with 0.5% formic acid immediately before analysis. Sample concentration ranged from 100 fmol/μL (acquisition of whole protein spectra), to 1 pmol/μL (acquisition of MS^n spectra). A sample volume of 0.5 μL was delivered to the mass spectrometer using a Nanomate 100 system (Advion Biosciences). A Nanomate low-flow 2.5 μm ESI chip was used as static nanoelectrospray emitter,

providing a stable flow of 20–30 nL/min. Voltages of 1.3–1.6 kV were applied to the chip through the Nanomate power supply, while the mass spectrometer source voltage was set to zero. Samples were infused using nitrogen gas at a pressures of 0.7–1.0 psi.

LTQ-Orbitrap Mass Spectrometry

Measurements were performed on a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific), in the positive ion mode. Electron multiplier gain, FT and storage transmission, as well as mass accuracy (FT) were calibrated immediately before measurements according to manufacturer's instructions. Collision-induced dissociation (CID) of selected protein charge states was performed in the LTQ and fragments were subsequently transferred and measured in the orbitrap. Selected protein charge states (typically the most abundant ones) were isolated with a width of m/z = 6 to 10 and activated for 30 ms using 30% normalized collision energy and an activation q of 0.25. The instrument was controlled using TunePlus 2.0 (beta 3) and the acquired spectra were evaluated using Xcalibur 2.0 software.

The orbitrap automated gain control (AGC) targets were set to $2x10^6$ charges for full scan, and $2x10^5$ for MS^n scan. Protein mass spectra were acquired at a resolving power of 60,000 and MS^n spectra were acquired at 60,000, 30,000 or 15,000. Lock-mass option was enabled in all measurements unless otherwise stated and polydimethylcyclosiloxane (PCM) background ions (at m/z 445.120025 and 429.088735) were used for real-time internal calibration as described previously[13]. Unless otherwise stated, all orbitrap scans consisted of 10 microscans (see below).

Data analysis

Protein masses were determined either by deconvolution using the integrated Xcalibur Extract software (Thermo Fisher Scientific) or by direct calculation from the peak positions and charge states. Expected protein fragment masses were calculated using PILGrinder software (developed in-house by Peter Mortensen), Protein Prospector software (http://prospector.ucsf.edu/)[14] or GPMAW (Lighthouse data)[15]. MS/MS spectra were searched with the web-based ProSight PTM (https://prosightptm.scs.uiuc.edu/)[16] against either the human Uniprot or a custom bovine database in 'Absolute Mass' mode. A wide tolerance of up to 2000 Da was used for the protein mass, to allow for differences between measured and theoretical masses due to protein modification. Fragment ion mass tolerance was in all cases set to 5 ppm and at least 5 matched fragments were required for protein identification.

MS^3 spectra were searched against NCBInr protein database (date 2005/04/15; 2440425 sequences) using the Mascot search engine (Matrix Science)[17]. Search criteria were: no enzyme specificity, precursor mass tolerance 5 ppm, and fragment mass tolerance 0.01 Da. Since Mascot cannot handle MS^3 data, we manually added the mass of H_2O (18.0106 Da) to all precursor ions that gave good quality MS^n spectra but did not result in protein identification. This formally turns b-type ions or internal fragments into peptide precursors. To match MS^3 fragments by Mascot when the precursor was a b-ion, y-ions were allowed to match only with H_2O loss.

Results and discussion

The utility of the LTQ-Orbitrap mass spectrometer for top-down analysis of proteins was assessed at three levels: (a) high-accuracy MS measurement of the whole protein in the orbitrap; (b) MS/MS (CID) fragmentation of a single charge state of the protein of interest in the LTQ, with subsequent detection of resulting fragments in the orbitrap; and

(c) MS^3 of selected CID fragment(s) in the LTQ and their detection in the orbitrap.

First we wanted to establish optimal conditions for protein measurement on the LTQ-Orbitrap. Nanoelectrospray[18,19] is commonly used in top-down proteomics because it allows detailed investigation of complex samples for extended periods of time and because it is very sensitive. Here we employed a newly developed very low flow nanoelectrospray chip (Nanomate, 2.5 µm nozzle inner diameter, Advion Biosciences), which supports stable flow rates in the true nanoelectrospray range of 20 to 30 nl/min. This allowed us to routinely use a volume of 0.5 µl and still acquire data for more than 15 minutes, more than sufficient time for all measurement sequences.

We investigated optimal parameters for acquisition of whole protein spectra in the orbitrap mass analyzer. The best sensitivity, S/N and accuracy were obtained when each scan consisted of 10 microscans. In this regime, transients of 10 consecutive microscans are added to form a final transient on which Fourier transform is performed. Data acquisition time was between 10 and 30 s at the resolution chosen (see below), so it was still very short compared to total available spray time. At the concentrations used, fill times for the 106 target value chosen was between 0.2 to 4 s, comparable to the transient time of 750 ms.

In non-mass resolved mode, target values of up to 10^7 are possible in the LTQ part of the instrument, much higher than the limit of about 10^6 of the C-trap. However, in mass resolved mode, for example when storing a charge state for subsequent dissociation, only about 10^5 ions can be accumulated. In this mode, the ability to sequentially fill the C-trap would be useful as pointed out previously[13].

Measurement of Intact Proteins

Sensitivity

At the intact protein level, well-defined "envelopes" arising from detection of multiple charge states were routinely obtained in the orbitrap for all investigated proteins at concentrations of 100-500 fmol/µL and total protein amounts of less than 250 fmol (Table 2.1). The lowest amount analyzed in this study was 50 fmol of cytochrome c (Figure 2.1). Note that this was the total protein amount used for measurement, and that high S/N protein mass spectra were obtained even after one MS scan (10 microscans), on a population of about $2x10^7$ ions (low attomole range) and acquisition time of 30 s. The same amount of cytochrome c was detected on the LTQ-FT instrument under the same measurement conditions, albeit with lower intensity and S/N ratio (data not shown).

High sensitivity measurements at the whole protein level have been reported before on FT-ICR instrumentation. For example, McLafferty and co-workers have reported sub-attomole detection of 8-30 kDa proteins using capillary electrophoresis separation coupled with FT-ICR mass detection[20], and Smith's group has reported low zeptomole detection of proteins in the same mass range[21]. However, these high sensitivities were not achieved routinely and required either up-front analyte separation and concentration, or extensive instrument optimization. The static nanoelectrospray measurements in this study show that the orbitrap sensitivity for proteins compares favorably with other instrumentation routinely used in top-down proteomics.

Resolution

All orbitrap protein spectra presented here were recorded at 60,000 resolution at m/z 400, which is the specified resolving power of the orbitrap mass analyzer.

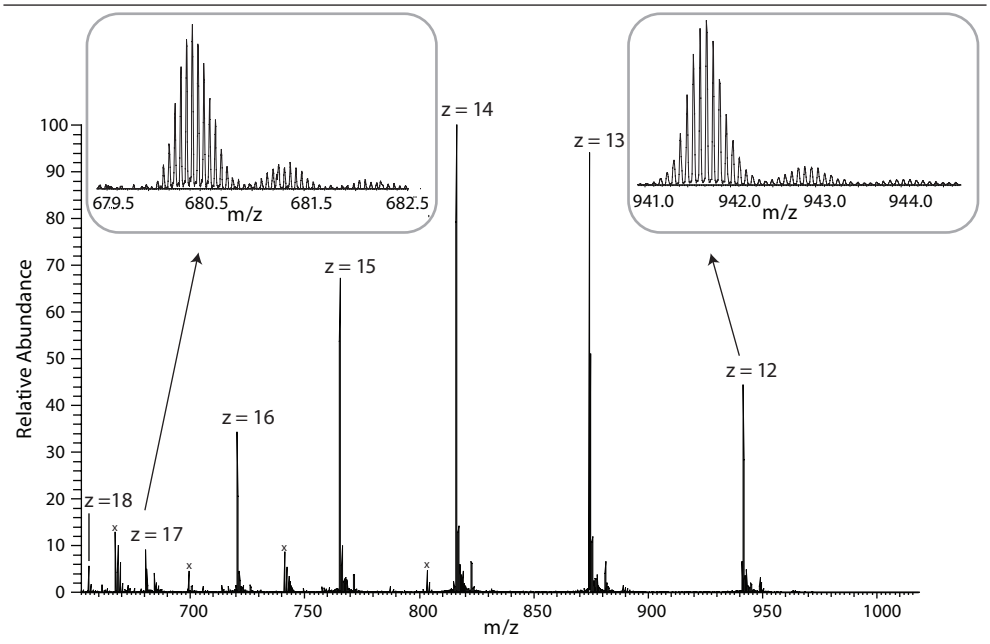


Figure 2.1 Single-scan mass spectrum of cytochrome c (0.5 µL of a 100 fmol/µL solution) acquired in the orbitrap mass analyzer at AGC target of $2x10^6$ charges, and resolution 60,000. The scan consisted of 10 microscans and took a total time of 29 s to acquire. Inserts show isotope resolution of protein charge states across the mass range. 'x' marks breakdown products or impurities in the sample.

This resolution requires 'one second' measurements (750 ms transients), whereas the maximum possible measurement time of 1.8 s leads to a resolution of 100,000. Although this value is less than the maximum possible with high magnetic field FT-ICR analyzers, it was sufficient for isotope resolution of all investigated proteins (Figure 2.1 and 2.2) and provided a good duty cycle. Furthermore, resolution in the orbitrap is inversely proportional to the square root of the m/z value, rather than to the m/z value directly as it is in the FT-ICR[10], resulting in slower decrease of observed resolution across the mass range. In practice this means that at m/z values above 1111.1 resolution power of 60,000 (m/z 400, 750 ms transient) of the orbitrap analyzer exceeds resolution power of 100,000 of the 7 Tesla LTQ-FTICR instrument. This unique feature is useful in analysis of intact protein mass spectra, where protein charge states are often observed above m/z 1000. It has to be noted, however, that longer transient acquisition in the LTQ-FT instrument leads to significantly higher resolution (specified to up to 500,000), whereas in the orbitrap mass analyzer the transient acquisition time is limited to 1.8 s and consequently its resolving power to about 100,000. Isotope resolution directly enables the detection of modifications such as disulfide bridges (Δm=2 Da) or deamidation (Δm=1 Da) of small proteins, whereas the ability to resolve isotopic clusters of different protein charge states across the mass range is important for proper charge state assignment in complex spectra caused by overlapping protein charge distributions.

Mass accuracy

For high mass accuracy, orbitrap spectra in this study were acquired in the 'lock-mass' mode of operation using background polydimethylcyclosiloxane (PCM) ions from the ambient air as internal calibrants, as described recently[13]. In the lock-mass mode, each orbitrap scan is preceded by a short LTQ scan in which a defined number of calibrant ions is injected into the C-trap. Analyte ions are then added into the C-trap and injected together with the calibrant ion into the orbitrap. The calibrant ion is recognized on-the-fly and masses of all ions are corrected in real time. For the intact protein measurements, the PCM ion at m/z 445.120025 was used as the reference, whereas for the MS^n measurements the corresponding neutral loss ($-CH_4$) PCM ion at m/z 429.088738 was used instead, due to its higher abundance in the MS^n spectra. This improved mass accuracy up to three fold (see Table 2.1).

Masses and standard deviations for each protein standard were calculated from up to 10 members of each isotopic distribution in at least two different charge states and two different scans. For larger proteins, where the monoisotopic peak could not readily be determined, the isotope state was calculated using the top-fitting method described by Zubarev et al.[22], with the cutoff level of h=50%. Alternatively, the protein mass was determined from the protein spectra deconvoluted by Xcalibur software. As expected, all protein standards were measured with very high mass accuracy, between 0.92 ppm (standard deviation, s.d. 0.79 ppm) to 2.80 ppm (s.d. 1.02 ppm) (Table 2.1). The overall (combined) absolute mass accuracy was 2.25 ppm with average s.d. 1.46 ppm.

Although FT-ICR MS is capable of achieving extremely high mass accuracies, there are only a few reports where accuracy was better than 10 ppm at the protein level. The main reason for this is the space-charging effect, which depends on the ion population in the ICR cell. Although several strategies have been proposed to overcome this effect[23,24], the best results were obtained when ion population was well controlled or when internal calibration was performed. For example, Smith and coworkers have used a dual ESI source to infuse γ-endorphin as internal calibrant and enabled "mass locking" to its doubly charged ion. In a high-throughput LC-FT-ICR analysis of intact proteins of the yeast large ribosomal subunit they have identified about a hundred proteins with average absolute accuracy of about 2.5 and s.d. of 2.0 ppm, in cases when protein monoiso-

Table 2.1 Masses and accuracies of intact proteins measured in the LTQ-Orbitrap (with and without lock mass (LM)), and amounts of proteins used for analysis. Two charge states, about 10 isotopes and two different scans were used for mass determination. Standard deviation was calculated from these 40 values.

	$MW_{theoretical}$ (Da)	$MW_{experimental}$ (Da)	Mass accuracy (ppm)		Amount measured (fmol)
			- LM	+ LM	
Cytochrome c	12224.209	12224.175	1.96 ± 1.33	2.16 ± 1.79	150
α-Crystallin (A)	19820.867	19820.838	6.63 ± 1.86	2.44 ± 1.50	250
α-Crystallin (B)	20067.393	20067.404	7.25 ± 1.69	2.78 ± 1.80	250
β-Lactoglobulin	18266.394	18266.396	4.72 ± 1.32	0.92 ± 0.79	150
β-Casein	23969.226	23969.198	6.50 ± 2.14	2.41 ± 1.84	150
Ubiquitin	10026.329	10026.484	4.15 ± 0.77	2.80 ± 1.02	150

topic mass could be measured[25].

While a detailed study of space-charging effects in the orbitrap mass analyzer has not yet been published, we have not observed systematic mass shifts as a function of ion number up to the limit imposed by the C-trap, of about 10^6 charges. Thus in our experiments mass accuracies for all six standard proteins were never worse than 7.5 ppm, even without lock mass option. In addition, protein masses were measured with extremely high precision, almost always better than two ppm (Table 2.1). Together, these data demonstrate that the LTQ-Orbitrap can routinely achieve extremely high mass accuracy at the protein level, without any software or hardware modification.

Analysis of posttranslational modifications (PTMs) at the protein level

One of the strengths of the top-down approach in MS-based proteomics is potentially comprehensive characterization of PTMs. Three of the proteins analyzed in this study, namely β-casein, and α-crystallin A and B chains are known to be phosphorylated. In case of β-casein, five phosphorylation sites have been reported, and the mass measured in the orbitrap corresponded to the mass of the protein form containing all five phosphate groups. In contrast, both A and B chains of α-crystallin were detected in several forms; A-chain was detected as unmodified, and modified with one phosphate group; B-chain was detected as unmodified, and modified with one, two and three phosphate groups (Figure 2.2). Stoichiometries of these modifications were in agreement with those observed in a previous MS study[26]. Note that mass differences arising from modifications were measured with high precision (10 mDa, or two significant decimal places), potentially enabling discrimination of modifications with the same nominal mass, such as acetylation and trimethylation (Δm = 35 mDa), at the intact protein level. In addition, it is clear from Figure 2.2 that sensitivity for detection of phosphoproteins was excellent in this case, since 250 fmole total protein mixture had been loaded, and the singly and doubly phosphorylated crystallin B chain make up less than 15 percent of this amount.

Measurement of Protein MS/MS Fragments in the orbitrap

Accurate protein mass alone would be of little use in the analysis of unknown proteins or protein mixtures. As discussed before, various fragmentation methods have therefore been applied to intact proteins (or their isolated charge states) in order to obtain at least partial information on protein primary structure. In the LTQ-Orbitrap fragmentation must be performed outside of the orbitrap mass analyzer, therefore two fragmentation types are currently possible. Proteins can be fragmented by nozzle-skimmer (in-source) dissociation, or collision-induced dissociation (CID) in the linear quadrupole ion trap (LTQ) with subsequent detection of fragment ions either in the ion trap or in the orbitrap. Our initial studies using nozzle-skimmer fragmentation resulted in very little fragmentation, without useful sequence information (data not shown), therefore all fragmentation was performed by CID in the LTQ.

As already reported in experiments on the LTQ-FTICR, intact proteins readily fragment in the LTQ under the same conditions normally used for peptide sequencing[7,8]. Various charge states of all analyzed proteins produced multiply charged fragment ions which were isotopically resolved and showed excellent S/N (Figures 2.3 and 2.4). As in the case of intact protein measurements, spectra were acquired as the sum of 10 microscans. Efficient transfer from the LTQ to the C-trap and orbitrap and absence of time-of-flight effects insured that fragments could be acquired in a single mass range. MS/MS

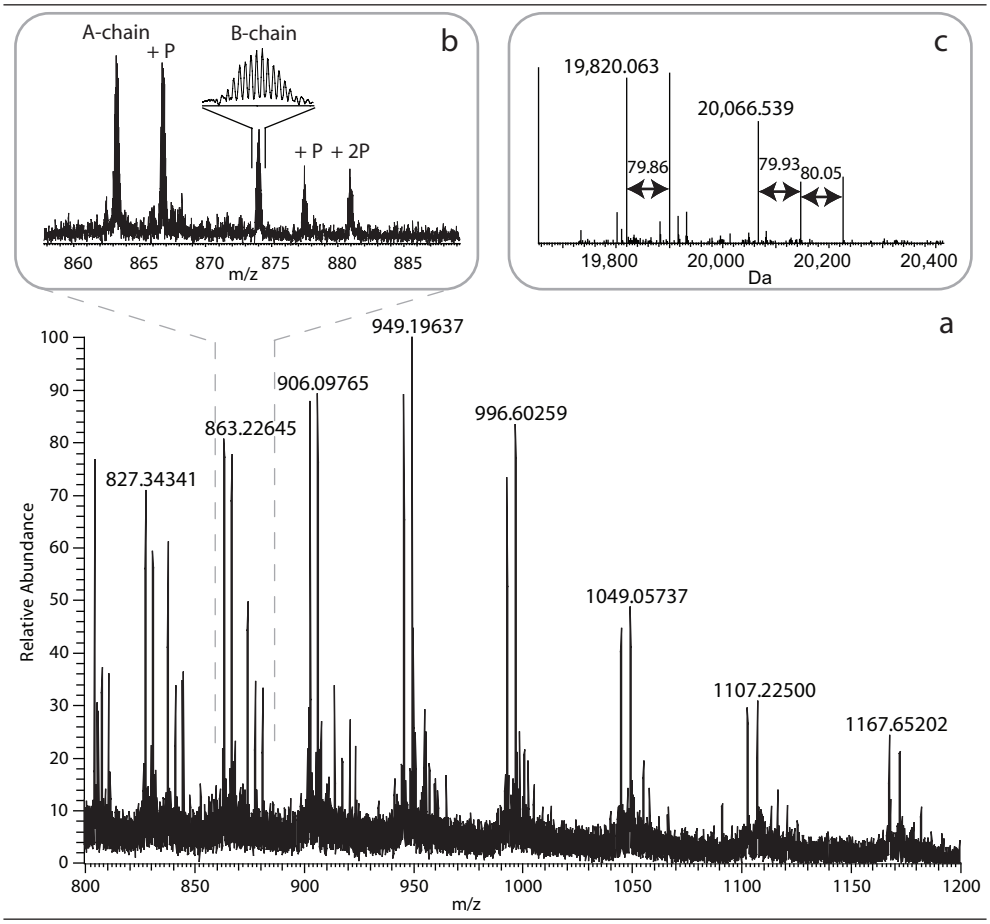


Figure 2.2 Mass spectrum of modified and unmodified α-crystallin A and B chains acquired in the orbitrap. (a) Appearance of the protein envelope at a total protein concentration of 500 fmol/µL; (b) Zoom-in of the $[M+23H]^{23+}$ charge state reveals presence of several phosphorylated forms of α-crystallin A and B chains; (c) Deconvoluted and de-isotoped spectrum of α-crystallin. The experimental values of the mass differences between modified forms are consistent with phosphorylation as the protein modification (theoretical mass difference 79.966).

fragments were calculated by ProSight (Single Protein Mode)[16]. Since ProSight calculates only b- and y-ions, internal fragments were calculated using Protein Prospector14 or GPMAW software[15]. As demonstrated for β-lactoglobulin and β-casein, fragments were measured with average absolute accuracy better than two ppm in the orbitrap.

CID of intact proteins has been performed in ion trap instruments before, producing multiply charged fragment ions (reviewed in [2]). Unfortunately, this had little application so far, due to the inability of ion traps to resolve charge states of fragment ions and necessity to perform charge state reduction by introduction of anions into the mass analyzer. The combination of a linear ion trap and the high resolution orbitrap analyzer now enables straightforward interpretation of protein fragment spectra without ion-ion reactions.

Fragmentation patterns

Figure 2.3a shows an MS/MS spectrum of β-lactoglobulin upon isolation and CID fragmentation of its $[M+15H]^{15+}$ charge state in the LTQ, and subsequent detection of fragment ions in the orbitrap. The fragmentation patterns of proteins in our study were in agreement with the ones previously reported for ion trap CID of intact proteins[7,8,27]. Cleavages C-terminal to charged residues, in particular Asp, Glu and Lys, as well as N-terminal to Pro, were the major fragmentation channels, although other cleavages were observed as well (Figures 3b and 4b). In MS/MS spectra of cytochrome c, ubiquitin, and β-lactoglobulin, complementary b/y fragments were observed. However, β-casein fragmented to approximately 70 residues from the N-terminus and 25 residues from the C-terminus, leaving the central portion of the molecule uncovered (Figure 2.4). Importantly, all detected b-ions formed by CID of its $[M+22H]^{22+}$ charge state that encompassed the five predicted phosphorylation sites indeed had all five

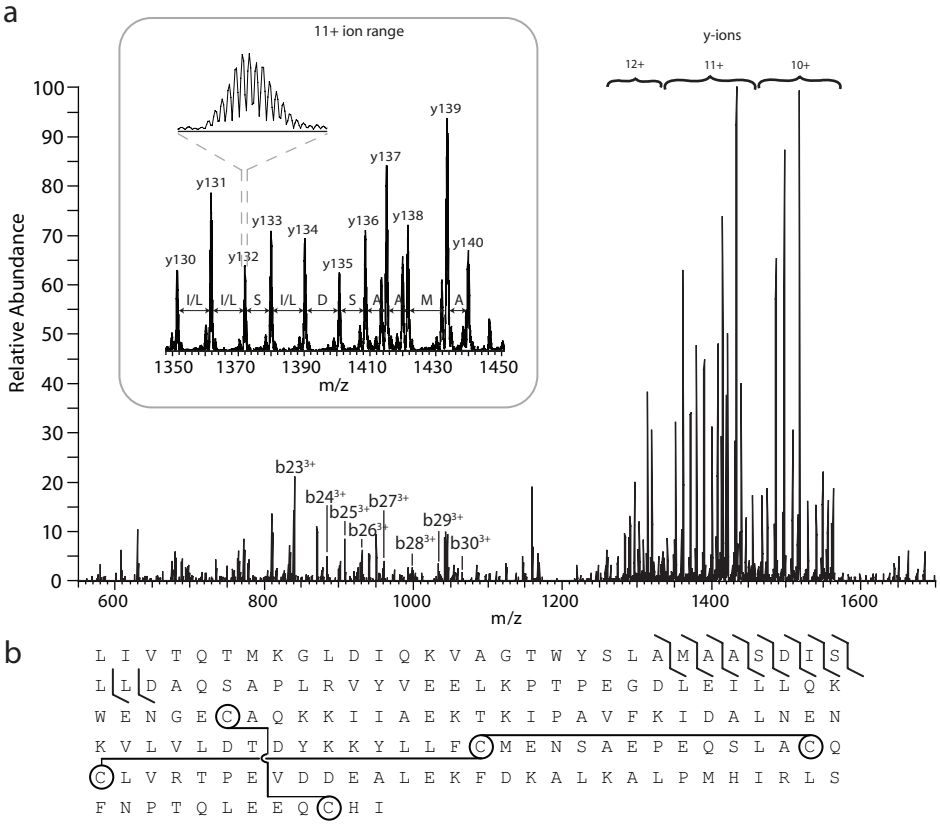


Figure 2.3 Collision-induced dissociation of β-lactoglobulin. (a) Single-scan fragmentation spectrum of the $[M+15H]^{15+}$ charge state of β-lactoglobulin acquired in the orbitrap at 30,000 resolution shows extensive fragmentation in the N-terminal region with a sequence tag easily detectable (insert); the scan consisted of ten microscans and took 12 s to acquire; (b) Sequence of the molecule with fragmentation pattern observed upon CID. The part of the molecule protected by the disulfide bonds is also indicated.

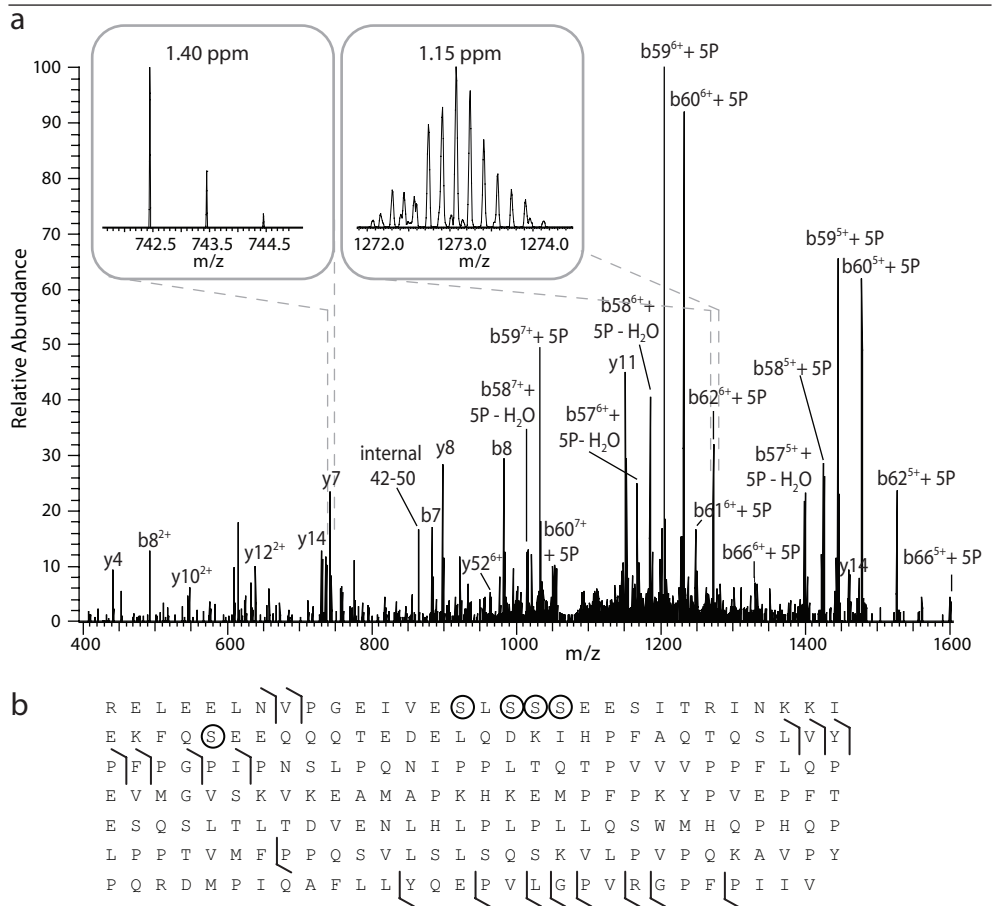


Figure 2.4 Collision-induced dissociation of β-casein. (a) Single-scan fragmentation spectrum of the $[M+22H]^{22+}$ charge state of β-casein analyzed in the orbitrap shows fragmentation at both termini. Note that all five phosphate groups remained attached to MS/MS fragments – neutral loss of phosphate upon CID was not observed. The insets show typical isotope patterns of small and large fragments as well as the obtained mass accuracy; the scan consisted of one microscan and took 1.12 s to acquire (at R=60,000); (b) Sequence of β-casein, with phosphorylation sites indicated by circles and fragmentation pattern observed upon CID.

phosphate groups attached. Together with low mass, unmodified b-ions, this locates all phosphosphorylation sites to between residues 9 and 57 of the sequence. Since the loss of phosphoric acid from the peptide backbone, commonly observed in 'bottom-up' approach, was not observed after CID of intact proteins, modified fragments could be fragmented further to give insight into the exact locations of protein modifications.

It is well known that CID fragmentation of whole proteins depends largely on the protein structure (such as the positions of disulfide bonds) and size and that larger proteins (>20 kDa) tend to fragment mostly in terminal regions. It is both an advantage and a disadvantage of CID that it concentrates fragmentation products into a few preferred

channels. This increases the sensitivity for detecting these fragments but will often preclude complete characterization with single residue resolution. ECD can potentially cleave almost all peptide bonds in a protein[28], however it suffers from relatively low efficiency, which results in lower sensitivity and increased acquisition times. Therefore, these methods are complementary, rather than competing. ETD, a recently developed fragmentation technique for ion traps, would be very suitable for the LTQ-Orbitrap, since it would combine the advantages of nonergodic fragmentation (high sequence coverage, preservation of labile bonds, etc.), with the high resolution, sensitivity and accuracy of the orbitrap mass analyzer demonstrated here.

Protein identification using MS/MS data

High accuracy measurement – typically better than two ppm – and straightforward assignment of the charge states of fragment ions, together with the accurately determined protein mass were used for protein identification. Monoisotopic masses of the MS/MS fragments were submitted to ProSight PTM (Absolute Mass mode), currently the only publicly available search engine for top-down proteomics. MS/MS spectra of recombinant human ubiquitin were searched against the human Uniprot database, whereas all other proteins were searched against both custom-made bovine database and human Uniprot database. The protein mass tolerance was set to 2000 Da to allow for potential modifications, and a minimum of 5 matched fragments with accuracy of 5 ppm or better were required for a hit. These stringent criteria led to unambiguous identification of all analyzed proteins. Human ubiquitin was identified with 10 fragment ions (average absolute mass accuracy 2.45 ppm, s.d. 0.92 ppm), and probability score of 7.4×10^{-19}. As expected, its mass was 1597.42 Da higher than theoretical, corresponding to the difference of the Flag-tag (DYKDDDDKKLMV) plus Met at its N-terminus (Δm = 0.030 Da). In the case of cytochrome c, the measured mass differed from the theoretical by 616.172, which corresponds to the mass of the heme group (Δm = 0.006 Da). This not only shows the potential of the top-down approach to conclusively identify proteins, but to point to modifications not considered in the database, as well. α-Crystallin A and B chain, β-casein, and β-lactoglobulin were also identified as unique, high probability hits; however, their scores were not representative since the custom-made bovine database contained only eight entries. When searched against the human database MS/MS spectra of these proteins did not lead to any hits.

This demonstrates that high accuracy MS/MS spectra of whole proteins acquired in the orbitrap can routinely be used for high stringency database search in top-down proteomics, leading to high confidence identification and greatly constraining the nature of a possible modification.

Top-down Proteomics Using MS^3 on the orbitrap

We have recently shown that an additional level of peptide fragmentation (MS^3) is feasible on the LTQ-FT-ICR at high sensitivity and chromatographic time scale[29]. Together with a new scoring algorithm, this data significantly improved confidence of peptide identifications. With this background and the excellent quality of the MS/MS spectra, we decided to explore the feasibility of the LTQ-Orbitrap for top-down, MS^3 experiments.

Figure 2.5a shows the MS/MS spectrum of ubiquitin. The predominant ion is $y58^{7+}$ caused by proline directed cleavage. This ion was isolated and dissociated in the LTQ and resulting fragments analyzed in the orbitrap (Figure 2.5b). Even a single microscan acqui-

sition (0.6 s) led to a good S/N spectrum, which was further improved by summing 21 spectra (Figure 2.5b). The spectrum contains many informative ^b and y ions (^b refers to the MS^3 fragment ion generated from the end of the sequence truncated by the MS/MS cleavage, see[29]). To generate MS^3 spectra similar in quality to Figure 2.5b typically required selecting one of the dominant ions of the MS/MS spectrum and a total acquisition time less than one minute. Thus, several different MS^3 spectra can be acquired of a single protein loaded a few hundred femtomoles into the Nanomate. We also explored higher stages of MS^n. Figure 2.5c shows an MS^4 spectrum analyzed in the LTQ and demonstrates that interpretable spectra can still be obtained after four rounds of isolation and fragmentation.

Recently, Zabrouskov et al.[8], working on an LTQ-FT-ICR, used MS^3 of proteins for increased confidence of protein identification. Fragment and protein masses alone may not be sufficient to identify a protein in cases where overlapping protein envelopes lead to simultaneous fragmentation of two or more different proteins (such as in complex mixtures), or where CID of proteins results in low number of detected fragments, leading to multiple hits and/or low probability scores in database search. In these cases, unambiguous protein identification may require an additional level of structural information. Additionally, as mentioned above, MS^3 could help to pinpoint or restrict sites of modification in proteins.

Finally, we decided to investigate if small, peptide-like MS/MS fragments of proteins could be fragmented and used for protein identification in a manner similar to protein identification in the 'bottom-up' approach. The advantage of such a strategy would be that small fragments are less likely to be modified and that there are relatively few possible fragmentation channels. We chose a small fragment of β-casein which was easily determined to be doubly charged based on its isotope spacing (see Figure 2.4a). This fragment was accumulated in the LTQ, fragmented and analyzed in the orbitrap. Figure 2.6 shows a relatively simple MS^3 spectrum similar to the MS/MS spectrum of a small peptide. Several changes were made to the normal Mascot modus to be able to search this data. Since the MS^3 precursor could be a b or y ion of the protein, both possibilities were checked. Assuming a y-ion did not lead to any matches, therefore the mass of water was added to the precursor mass to formally convert it from b-ion to a peptide precursor. Mascot was directed to match only b or y-H_2O ions (these are the ^y MS^3 fragments generated from b-ion precursors[29]). Despite the short peptide length, a 'non-enzyme' specificity search by Mascot in the non-redundant database (NCBInr) yielded only two significant matches. The top match was located in the N-terminal region of the beta-casein precursor. A check of the database confirmed that Mascot had identified the eight N-terminal residues of the mature protein. The second significant hit had a related peptide sequence accounting for matched fragments despite the high mass accuracy, but was located close to the center of the protein sequence and was therefore discarded. Since the protein was modified with five phosphogroups, this example shows an interesting additional way of identifying modified proteins.

Conclusions

We here investigated whether a compact linear ion trap – orbitrap hybrid mass spectrometer (LTQ-Orbitrap) was suitable for top-down proteomics of small proteins. We interfaced the LTQ-Orbitrap to a low flow Nanomate static nanoelectrospray source providing more than 20 minutes measure-

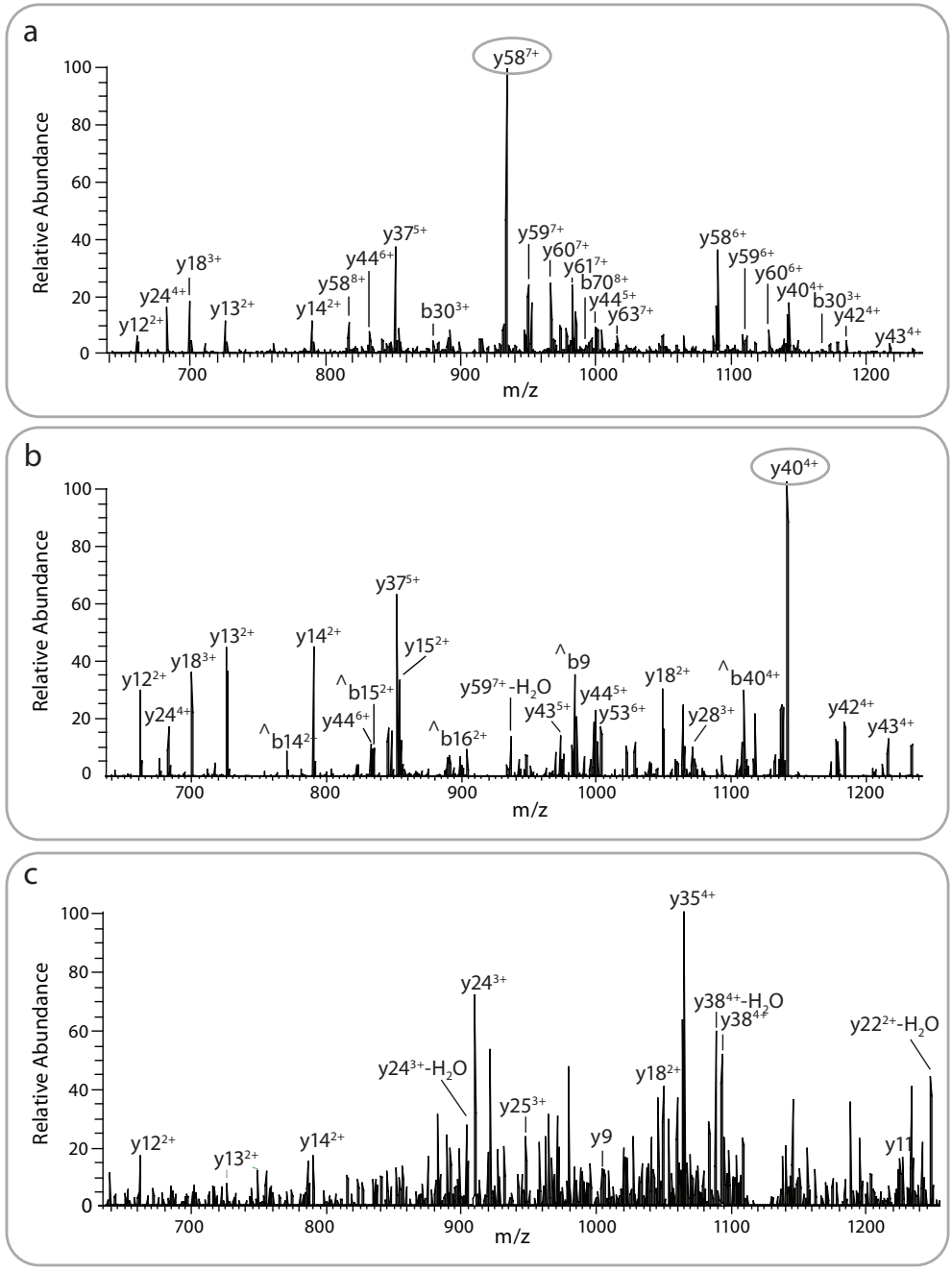


Figure 2.5 Multistage MS of the FLAG-tagged ubiquitin $[M+11H]^{11+}$ charge state in the LTQ-Orbitrap. (a) MS/MS fragmentation spectrum analyzed in the orbitrap; (b) MS^3 spectrum of the abundant, proline directed $y58^{7+}$ fragment ion analyzed in the orbitrap; (c) MS^4 spectrum (LTQ) of the proline directed $y40^{4+}$ fragment shown in panel b.

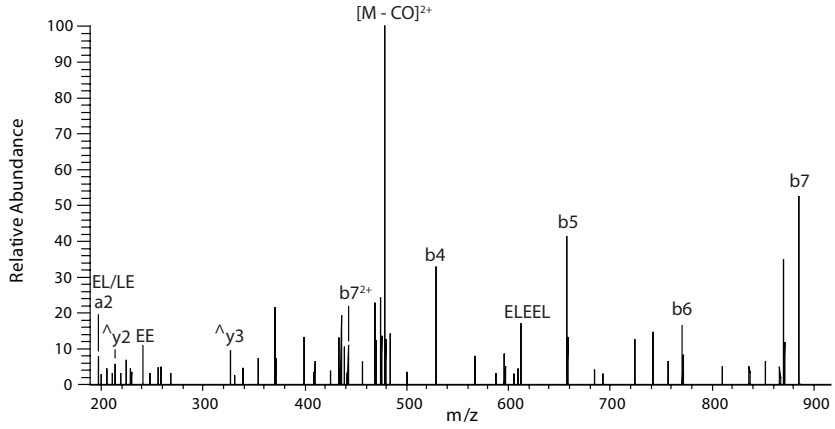


Figure 2.6 MS^3 spectrum of the $b8^{2+}$ ion precursor from β-casein depicted in Figure 2.4a. The spectrum was acquired at 15,000 resolution without the lock mass option (average absolute mass accuracy was 4 ppm for all matched fragments). Mascot search of this fragmentation spectrum identifies the peptide, RELEELNV, which is the N-terminal sequence of the mature β-casein protein.

ment time for half a microliter of solution. Sub-picomole amounts of intact proteins proved sufficient to obtain high quality MS, MS/MS and MS^3 spectra, with mass accuracies in the few ppm range. Intriguingly, database searches with the MS/MS data allow straightforward identification not only of unmodified but also of modified proteins. Since the molecular weight of proteins seldom corresponds to that calculated from the database sequence, this should be very valuable for top-down proteomics. In particular we show that small peptide-like fragments generated by protein MS/MS can be fragmented similar to peptides in the bottom-up approach, potentially combining advantages of both strategies. In the future, it will be interesting to employ this instrument with online protein separation as well as to develop new algorithms tailored to the strength of the data that it generates. This technology may then allow routine, high accuracy and high confidence analysis of the 'peptidome' and small protein component of the proteome.

References

1 Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. Nature 422, 198-207

2 Reid, G. E., and McLuckey, S. A. (2002) 'Top down' protein characterization via tandem mass spectrometry. J Mass Spectrom 37, 663-675

3 Kelleher, N. L. (2004) Top-down proteomics. Anal Chem 76, 197A-203A

4 Bogdanov, B., and Smith, R. D. (2005) Proteomics by FTICR mass spectrometry: top down and bottom up. Mass Spectrom Rev 24, 168-200

5 Ge, Y., et al. (2002) Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. J Am Chem Soc 124, 672-678

6 Coon, J. J., et al. (2005) Protein identification using sequential ion/ion reactions and tandem mass spectrometry. Proc Natl Acad Sci USA 102, 9463-9468

7 Wu, S. L., et al. (2004) A new and sensitive online liquid chromatography/mass spectrometric approach for top-down protein analysis: the comprehensive analysis of human growth hormone in an E. coli lysate using a hybrid LTQ/

FTICR mass spectrometer. Rapid Commun Mass Spectrom 18, 2201–2207

8 Zabrouskov, V., et al. (2005) New and Automated MS(n) Approaches for Top-Down Identification of Modified Proteins. J Am Soc Mass Spectrom 16 (12), 2027–2038

9 Makarov, A., et al. (2005) Dynamic Range of Mass Accuracy in FTMS. Proc. 53rd Conf. Am. Soc. Mass Spectrom, San Antonio, TX, 2005

10 Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal Chem 72, 1156–1162

11 Hardman, M., and Makarov, A. (2003) Interfacing the orbitrap mass analyzer to an electrospray ion source. Anal Chem 75, 1699–1705

12 Hu, Q. Z., et al. (2005) The orbitrap: a new mass spectrometer. J Mass Spectrom 40, 430–443

13 Olsen, J. V., et al. (2005) Parts per million mass accuracy on an orbitrap mass spectrometer via lock-mass injection into a C-trap. Mol Cell Proteomics 4, 2010–2021

14 Clauser, K. R., Baker, P., and Burlingame, A. L. (1999) Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Anal Chem 71, 2871–2882

15 Peri, S., Steen, H., and Pandey, A. (2001) GPMAW – a software tool for analyzing proteins and peptides. Trends Biochem Sci 26, 687–689

16 Taylor, G. K., et al. (2003) Web and database software for identification of intact proteins using "top down" mass spectrometry. Anal Chem 75, 4081–4086

17 Perkins, D. N., et al. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567

18 Wilm, M., and Mann, M. (1996) Analytical Properties of the Nanoelectrospray Ion Source. Anal Chem 68, 1–8

19 Wilm, M., et al. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nanoelectrospray mass spectrometry. Nature 379, 466–469

20 Valaskovic, G. A., Kelleher, N. L., and McLafferty, F. W. (1996) Attomole protein characterization by capillary electrophoresis-mass spectrometry. Science 273, 1199–1202

21 Belov, M. E., et al. (2000) Zeptomole-sensitivity ESI – FTICR mass spectrometry of proteins. Anal Chem 72, 2271–2279

22 Zubarev, R. A., et al. (1995) Approaches and Limits for Accurate Mass Characterization of Large Biomolecules. Analytical Chemistry 67, 3793–3798

23 Marshall, A. G., and Hendrickson, C. L. (2001) Charge reduction lowers mass resolving power for isotopically resolved ESI FTICR mass spectra. Rapid Commun Mass Spectrom 15, 232–235

24 Bruce, J. E., et al. (2000) Obtaining more accurate FTICR mass measurements without internal standards using multiply charged ions. J Am Soc Mass Spectrom 11, 416–421

25 Lee, S. W., et al. (2002) Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. Proc Natl Acad Sci USA 99, 5942–5947

26 Hogan, J. M., Pitteri, S. J., and McLuckey, S. A. (2003) Phosphorylation site identification via ion trap tandem mass spectrometry of whole protein and peptide ions: Bovine alpha-crystallin A chain. Anal Chem 75, 6509–6516

27 Reid, G. E., Stephenson, J. L., Jr., and McLuckey, S. A. (2002) Tandem mass spectrometry of ribonuclease A and B: N-linked glycosylation site analysis of whole protein ions. Anal Chem 74, 577–583

28 Sze, S. K., Ge, Y., Oh, H., and McLafferty, F. W. (2002) Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. Proc Natl Acad Sci USA 99, 1774–1779

29 Olsen, J. V., and Mann, M. (2004) Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. Proc Natl Acad Sci USA 101, 13417–13422

3

Top-down quantitation and characterization of SILAC-labeled proteins

Leonie F. Waanders, Stefan Hanke and Matthias Mann

J Am Soc Mass Spectrom 2007 18 (11), 2058–2064

Stable Isotope Labeling by Amino acids in Cell culture (SILAC) has become a popular labeling strategy for peptide quantitation in proteomics experiments. If the SILAC technology could be extended to intact proteins this would enable direct quantitation of their relative expression levels and of the degree of modification between different samples.

Here we show through modeling and experiments that SILAC is suitable for intact protein quantitation and top-down characterization. When SILAC-labeling lysine and/or arginine, peaks of light and heavy SILAC-doublets do not interfere with peaks of different charge states at least between 10 and 200 kDa. Unlike chemical methods, SILAC ensures complete incorporation – all amino acids are labeled. The isotopic enrichment of commercially available SILAC amino acids of nominally 95–98% shifts the mass difference between light and heavy state, but does not lead to appreciably broadened peaks.

We expressed labeled and unlabeled Grb2, a 28 kDa signaling protein, and show that the two forms can be quantified with an average standard deviation of 6%. We performed on-line top-down sequencing of both forms in a hybrid linear ion trap orbitrap instrument. The quantized mass offset between fragments provided information about the number of labeled residues in the fragments, thereby simplifying protein identification and characterization.

Introduction

The use of mass spectrometry to analyze biological samples has evolved tremendously over the last decade, and mass spectrometry-based proteomics, in particular, has become an indispensable part of modern life sciences[1]. In recent years, the need for quantitative as opposed to qualitative proteomics experiments has become apparent and a large number of different approaches have been developed, mainly based on incorporation of a 'light' and 'heavy' stable isotope tag[2]. In chemical approaches, such as in the original ICAT and in the iTRAQ method, the label is reacted with a functional group of an amino acid. In metabolic approaches, the label is instead incorporated by living cells through protein turnover. Our laboratory has previously described a metabolic labeling method called Stable Isotope Labeling by Amino acids in Cell culture (SILAC,[3]). In SILAC, an essential amino acid has been substituted by its stable isotope counterpart in the medium in which the cells grow, and this 'heavy' amino acid is hence incorporated into all expressed proteins. SILAC allows very accurate peptide quantitation in an automated, high throughput experimental setup.

In contrast to this 'bottom-up' approach, 'top-down' proteomics seeks to characterize intact proteins. Although not widely used in biological research yet, top-down proteomics has unique potential because it can characterize the complete primary structure of the proteins, including modifications that may be missed in the bottom-up approach. Until now, most top-down studies have focused on mass measurement of the protein, its

identification in databases using MS or MS/MS data or on measuring protein modifications. Hardware and software developments, such as improved resolution and sensitivity, better fragmentation techniques and more automated software, have recently made top-down identification and characterization much faster[4-7]. These developments should make quantitative rather than qualitative top-down proteomics more feasible. So far, however, very few studies have combined the top-down approach with protein quantitation. Gordon et al. demonstrated that relative molecular ion intensities can be used for intact protein quantitation[8]. Kelleher and co-workers investigated the use of ^{15}N labeling of yeast proteins for intact protein quantitation and determined 50 protein ratios[9]. In the same paper, the authors also chemically labeled yeast proteins with acrylamide and iodoacetamide. They concluded that the use of stable isotopes is preferred, since it prevents chromatographic shifts during LC-MS separation and make quantitation easier and more accurate. Furthermore, a general problem of chemical labeling strategies is the fact that it is not 100% complete in terms of incorporation. Due to steric hindrance in the intact proteins, not all amino acids react evenly well with the isotope labeled reagent, and therefore the degree of labeling is difficult to control.

Although quantitative approaches are rare in the top-down proteomics field, stable isotopes have been used for intact protein analysis for a number of years. In 2000, Smith and co-workers used deuteriated leucine to improve the identification of E. coli proteins as the mass offset in the full scan indicated the number of leucines present in the protein[10]. They also substituted other essential amino acids like Ile, Phe, Arg, His and Lys to determine their number in the protein[11]. 'Heavy' and 'light' labeled proteins were well separated as long as the labeled amino acid was present at least three times per protein.

We have recently investigated the LTQ-Orbitrap, a hybrid linear ion trap orbitrap mass spectrometer, for top-down proteomics[12]. We found that the good sensitivity and excellent mass accuracy and resolution are sufficient for fast and reliable protein identification and characterization. We now extend this work by investigating whether the SILAC technology is also applicable to intact protein quantitation and characterization. We show feasibility by modeling SILAC-labeling for proteins up to 220 kDa, and by quantifying a 28 kDa signaling protein that was expressed in medium with normal and with heavy arginine and lysine residues.

Materials and methods

Modeling

The theoretical applicability of SILAC for protein quantitation was tested by modeling the isotopic and charge state clusters of proteins of different molecular weight. Based on the amino acid frequencies as determined by McCaldon and Argos[13], we calculated the mass of a hypothetical sequence with 100 residues (~11 kDa) and extrapolated this to 55 and 220 kDa. For the modeling we used the Isotopica web application, developed by the groups of Fernandez-de-Cossio and Takao[14]. We first simulated SILAC-labeling with $^{13}C_6{}^{15}N_4$-arginine and $^{13}C_6{}^{15}N_2$-lysine, inducing a mass increment of 10.008 and 8.014 Da respectively, assuming that 100% of the substituted carbon and nitrogen atoms were 'heavy'.

The influence of incomplete labeling and isotope enrichment were tested by calculating the frequency of different protein forms in an arginine-labeled 55 kDa protein with approximately average amino acid distribution. To test the effects of imperfect isotope enrichment of the SILAC amino acids, we

varied the relative abundance of the substituted ^{13}C and ^{15}N in the labeled amino acid from 100 percent to a lower value such as 98 or 95%.

In order to simulate incomplete amino acid incorporation, we calculated the relative abundance of protein forms when the chance of incorporation of light labels was p=0.02 or p=0.05, assuming a binominal distribution. For every protein form, the isotope distribution of the 34th charge state was predicted with Isotopica (resolution 60,000) and protein forms were weighted according to their probabilistic frequencies.

In vitro expression & purification of Grb2

The gene for growth factor receptor-bound protein 2 (Grb2) was purchased from RZPD (clone RZPDo834A0934D). It was transferred into pDEST17 via Gateway-cloning for T7 RNA polymerase dependent expression as an N-terminally His_6-tagged protein. All chemicals and enzymes were purchased from Sigma or Roche Applied Science at the highest purity.

Recombinant N-terminally His_6-tagged Grb2 was expressed using a cell-free system prepared as described in ref. [15] with slight modifications. E. coli S30 lysate was prepared from BL21(DE3)RIL cells (Novagen). T7 RNA polymerase was also expressed in BL21(DE3) RIL containing the vector pAR1219 described in ref. [16], but without enzyme purification. Instead, we prepared another lysate from this IPTG-induced culture and added 60 µL of this lysate with 400 µL the standard lysate to 1 mL of reaction volume. The concentration of the reaction components were adjusted to 57 mM HEPES-KOH buffer (pH 8.2), 2 mM DTT, 1.2 mM ATP, 0.85 mM each of CTP, GTP and UTP, 100 mM creatine phosphate, 130 µg/mL creatine kinase, 2.0% PEG 8000, 0.64 mM 3',5'-cyclic AMP, 34 µM L(-)-5-formyl-5,6,7,8-tetrahydrofolic acid, 175 µg/mL E. coli total tRNA, 90 mM potassium glutamate, 80 mM ammonium acetate, 12 mM magnesium acetate, 2.0 mM of each of the 20 amino acids and 6.7 µg/mL of plasmid DNA. The reaction mixture was incubated for 2 h while shaking at 600 rpm at 30 °C.

For the heavy SILAC labeled Grb2, the normal arginine and lysine were replaced with their heavy-isotope counterparts $^{13}C_6{}^{15}N_4$-arginine and $^{13}C_6{}^{15}N_2$-lysine, also at a concentration of 2 mM. The reaction mixture contained the target protein mostly as insoluble precipitate which was solubilized with 6 M guanidinium chloride and subsequently purified employing Ni^{2+}-affinity chromatography according to the manufacturer's protocol for purification of denatured protein using Ni-NTA sepharose (Qiagen) and spin columns (MoBiTec). The purified protein was dialyzed against distilled water to remove imidazole and guanidinium which again lead to precipitation of the protein.

Sample preparation & mass spectrometry

Heavy and light labeled forms of Grb2 were aliquoted separately and stored at -80 prior to use. Before mass spectrometric analysis, the proteins were washed on RP-C18 StageTips[17], eluted and mixed in the desired ratio.

All experiments were done on a LTQ-Orbitrap (Thermo Fisher Scientific), coupled to a nanoLC system (Agilent). Online protein separation was performed by use of 75 µm x 150 mm IntegraFrit columns (New Objective) packed in-house with 5 µm RP-C18 beads (Reprosil-Pur Aq, 200Å pore size, Dr. Maisch). The column was connected to a short nanospray needle and spraying voltage was kept low (<2 kV) to prevent oxidation[18]. During loading and washing the flow was set to 500 nL/min; whereas during the actual gradient the flow rate was 250 nL/min. Buffers compositions were 0.5% acetic acid in mQ distilled water (buffer A) and 0.5% acetic

acid in acetonitrile (buffer B). The proteins eluted in a thirty-minute gradient from 40% to 90% of buffer B. CID fragmentation was performed in the LTQ, but MS/MS ions were detected in the orbitrap mass analyzer.

Data analysis

The MS scans were deconvoluted with Xtract-software (Thermo Fisher Scientific) and matched with the expected sequence. The light form of Grb2 was used to determine the sequence of the linker between the His-tag and the protein (see Table 1). CID fragments observed in light or in both light and heavy MS/MS spectra were searched in ProSight PTM, but without success. We therefore suspected they were internal fragments and searched them in Mascot (Matrix Science) after increasing light fragment masses by one water molecule (18.0152 Da). A peptide mass fingerprint search was performed against the human IPI database (version 3.19) and a small database with the predicted His-tagged Grb2 sequence with 10 ppm maximum mass deviation and without enzyme specificity.

Results and discussion

Simulation of SILAC quantitation

We first wanted to investigate whether SILAC would be applicable for proteins of all sizes and whether any overlap between charge states and labeled states could occur. To this end, we used Isotopica, a software package freely available on the internet and developed by Fernandez-de-Cossio and Takao[14], to model the effect of SILAC labeling. Isotopica can predict the isotopic cluster of specific charge states and can also calculate the mono-isotopic and average mass of all charge states (up to 50) for both heavy and light labeled proteins. As shown in Figure 3.1, panels a-c, SILAC labeling with heavy arginine and lysines results in mass offsets that are clearly distinguishable in a spectrum. The thickness of the lines is a proxy for the full width of the isotopic cluster. Even for a 220 kDa protein (Figure 3.1c) the two lines are separate, indicating that at least theoretically the isotopic clusters of the light and heavy protein forms do not overlap with consecutive charge states.

Because the mass offset induced by SILAC is in the range of a few hundred Dalton - when both arginine and lysine are labeled - post-translational modifications with a mass less than this would not overlap with either the heavy or the light form of the protein. For example, several phosphorylation sites on a 50 kDa protein would not overlap with the heavy version, which is approximately 500 Da higher in mass.

The simulation also allowed us to visualize the effect of incomplete isotope enrichment. In principle, SILAC amino acids should be labeled with ^{13}C or ^{15}N to 100 percent at the substituted sites but in practice commercial sources guarantee isotope enrichment between 95 and 98 percent. As described in the Experimental Section we modeled this by varying isotope abundances in the heavy form of an average 55 kDa arginine-labeled protein. Interestingly, incomplete isotope enrichment does not broaden the peaks but it does shift the mass of the heavy form to a lower value (Figure 3.1d). It follows from this that accurate mass measurement to determine the molecular weight of the intact protein should be performed on the light and not on the heavy form of the SILAC protein pair.

Next, we modeled the effect of incomplete labeling, that is, we modeled the case that not every light amino acid was replaced by its heavy counterpart. This could happen during SILAC cell adaptation, while light amino acids are still present. As Figure 3.1e shows, even incomplete labeling at the two percent level (on average no more than one amino acid per

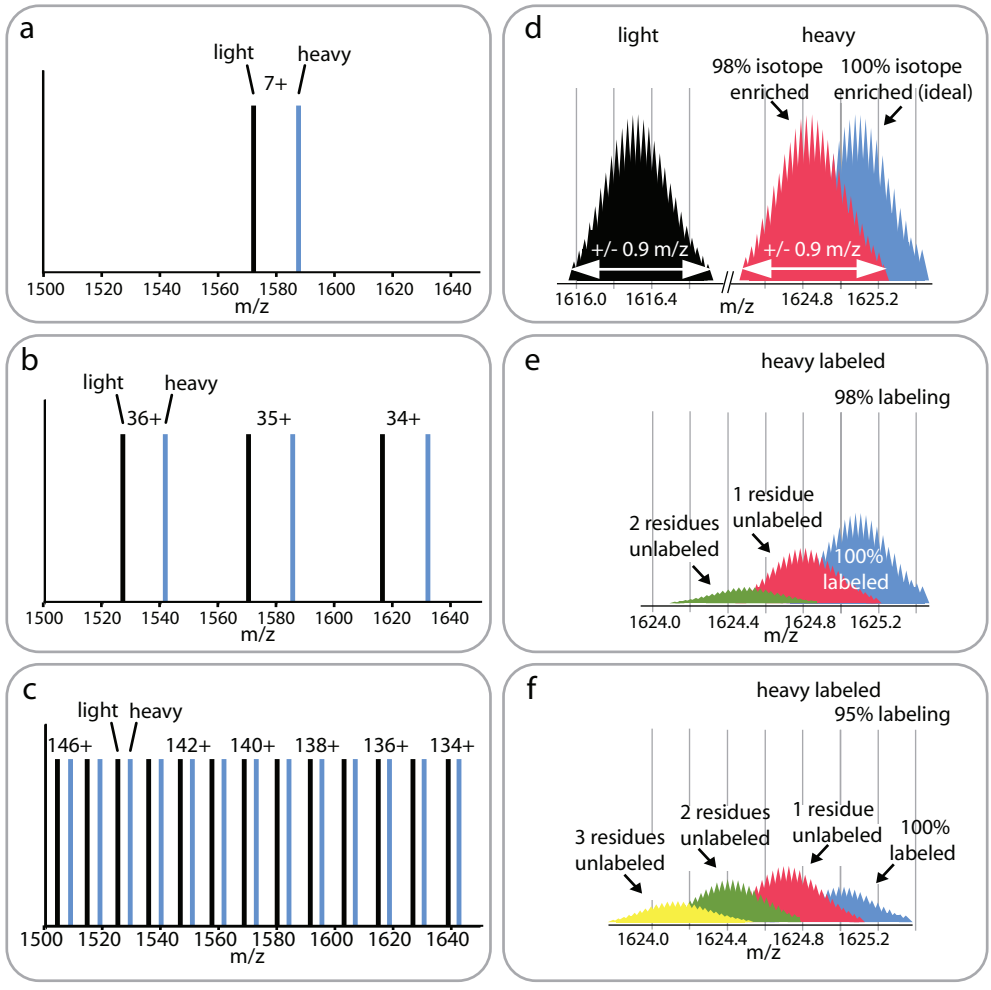


Figure 3.1 Simulation of heavy and light SILAC labeled proteins of various molecular weights with an average amino acid distribution. (a–c) SILAC-labeling with $^{13}C_6{}^{15}N_4$-Arg and $^{13}7_6{}^{15}N_2$-Lys increases the protein mass by approximately one percent, resulting in well separable isotope clusters, independent of protein size. The isotopic clusters of light and heavy labeled proteins of 11, 55 and 200 kDa in the m/z range 1500 to 1650 are depicted in panels a–c, respectively. The thickness of the lines indicates the full width of the isotope cluster, which for all size proteins remains below 1 Thomson. (d–f) The effects of incomplete isotope enrichment and incomplete mass labeling are modeled, based on the 34+ peak of a theoretical protein of 55 kDa with average amino acid composition, labeled with heavy arginine and measured with 60,000 resolving power. (d) A 98 percent isotope enrichment of ^{13}C and ^{15}N in the heavy amino acids results in a shift of the total heavy isotopic cluster from 100% enrichment. The width and the intensity of the isotopic cluster remain unchanged. (e) In contrast, if only 98% of the arginines and lysines are labeled with heavy amino acids, the result is a significant spread of the signal. (f) With 95% labeling the signal is even more spread and the total intensity is reduced to 33% of the original signal.

protein in our example), causes the signal to split into at least three states and this reduces the overall intensity and signal to noise (S/N) dramatically. At the five percent level, severe splitting into at least four states occurs, and the heavy peak is broadened at least three fold. Therefore, complete labeling is a precondition for successful quantitative top-down analysis. Fortunately, it is not difficult in SILAC experiments to achieve complete labeling; the only requirement is to grow the cells for a sufficient number of cell doublings.

Note, however, that this broadening is very likely to occur in chemical labeling strategies because it is very difficult to achieve close to 100% labeling on the desired amino acids while preventing any labeling of untargeted sites.

Expression & MS of SILAC-labeled Grb2

In order to experimentally test the feasibility and accuracy of quantitation by SILAC for an intact protein we expressed HIS-tagged Grb2, a signaling protein with a calculated monoisotopic mass of 27789.758 Da *in vitro* in normal media and in media with heavy arginine and lysine. Light and heavy protein forms were measured separately and mixed in approximately 1:1 or 2:1 ratio. The mixture was analyzed by online HPLC MS on the LTQ-Orbitrap (See Figure 3.2).

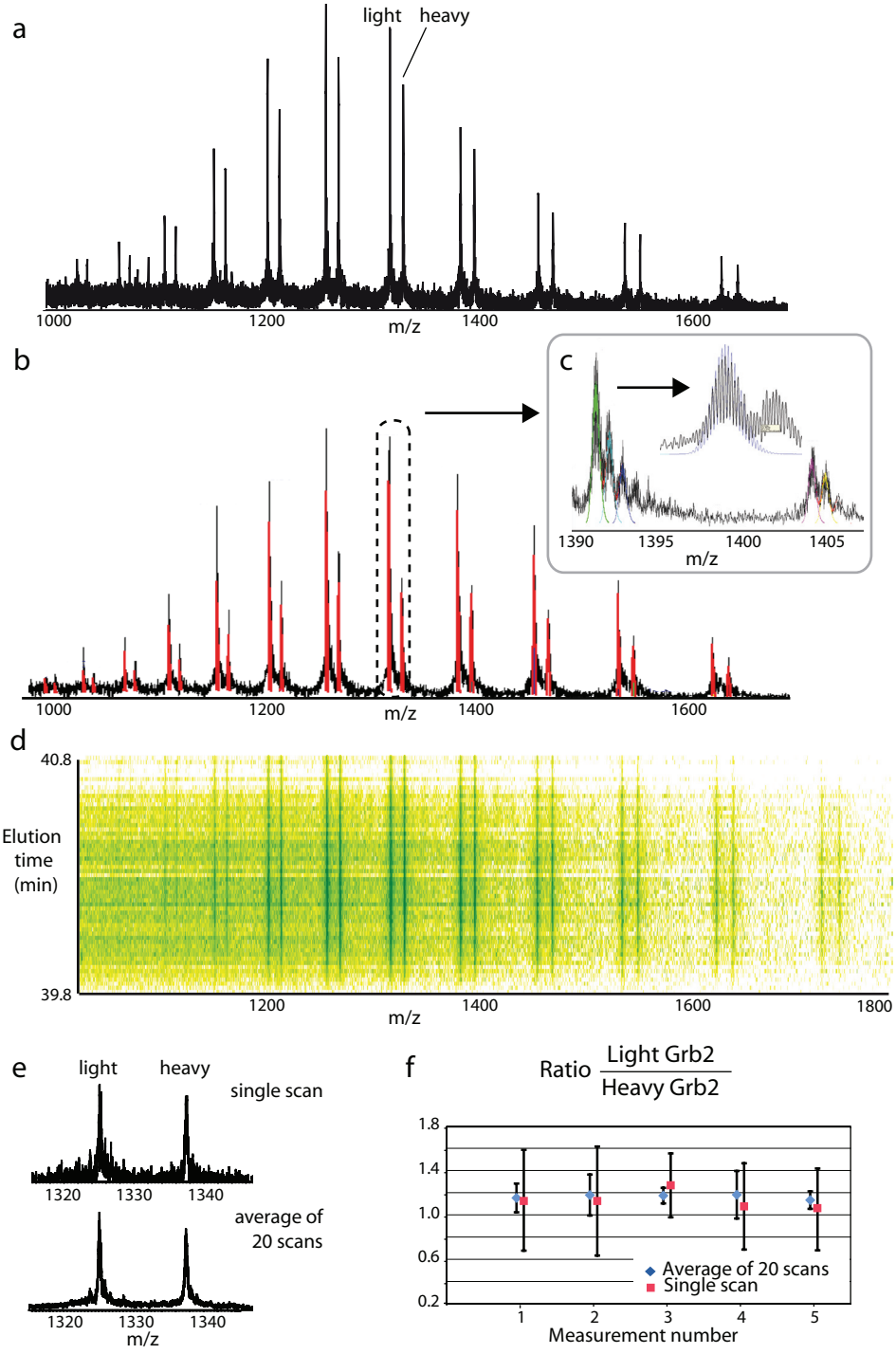
The resolving power of the orbitrap was sufficient to observe the isotopic clusters. By measuring the heavy labeled protein only, we obtained a single population and conclude that we achieved 100% label incorporation (data not shown).

Influence of isotope enrichment

Above, we modeled the influence of incomplete isotope enrichment of the SILAC amino acid. The SILAC amino acids used in our experiment were specified at 98% for ^{13}C. For ^{15}N they were specified as 95% for Lys and 98% for Arg. Thus about 2% of the labeled carbon and nitrogen atoms in their stable isotope labeled amino acids should be light (^{12}C and ^{14}N). However, when we used this value to predict the mass offset between the heavy and light labeled Grb2 protein forms, we could not explain an additional mass difference of 3 Da for the heavy labeled Grb2 form in comparison to the calculated one of 264.3 Dalton. Since the sequence of the protein is known and confirmed by the light labeled protein, this mass difference had to be caused by the heavy amino acids.

When overlaying the experimentally obtained spectra with the prediction made by Isotopica (Figure 3.2c), we discovered that the matching became very accurate when isotope enrichment was 99%. Thus, in this case, the isotope enrichment of the SILAC amino acids was almost complete and significantly higher than specified. Furthermore, our measurements confirmed the simulation in Figure 3.1d. The rare

Figure 3.2 (right page) Experimental spectra of mixed Grb2 forms, showing the experimental feasibility of SILAC. (a) Single spectrum of mixed light and heavy Grb2 forms, showing multiple charge states of both forms with isotopic clusters that do not overlap. (b) An experimentally obtained spectrum was overlaid with the Isotopica prediction (red), calculated using 99% heavy isotope enrichment. (c) The inserts show that the simulation is accurate down to the isotopic resolution, indicating higher isotope enrichment of Arg and Lys than specified by manufacturers. (d) Contour plot of the eluting SILAC labeled protein pair. The graph is color coded with light green corresponding to low and dark green to high intensity. The ratio between heavy and light Grb2 is constant over the different charge states and allows accurate quantitation based on relative intensities. (e) Averaging of spectra leads to significant noise reduction and improves quantitation accuracy. (f) Intensity ratios between light and heavy labeled Grb2, based on the eight most intense charge states of single or averaged spectra (20 scans) for 5 different LC-MS-runs were determined with their 95% confidence interval error bars.


a
light
heavy
1000
1200
m/z
1400
1600
b
c
1390
1395
1400
1405
m/z
1000
1200
m/z
1400
1600
d
40.8
Elution
time
(min)
39.8
1200
1400
1600
1800
m/z
e
light
heavy
single scan
1320
1330
1340
average of
20 scans
1320
1330
1340
m/z
f
Ratio
Light Grb2
Heavy Grb2
1.8
1.4
1.0
0.6
0.2
Average of 20 scans
Single scan
1
2
3
4
5
Measurement number

presence of ^{12}C and ^{14}N instead of ^{13}C and ^{15}N in the heavy amino acids did not cause problems for the analysis or quantitation, since the signal intensity and the width of the isotopic distribution remained the same. Note that the percentage of light atoms in the heavy label needs to be determined only once as it will be the same for all proteins that are labeled with that specific batch of heavy amino acids.

Protein quantitation using SILAC

A contour plot of the SILAC labeled Grb2 forms showed that they eluted as a one minute peak with no discernable retention time shift between them (Figure 3.2a). Complete isotopic separation and co-elution should allow for very accurate quantitation. The spreading of the protein signal into multiple charge states reduces the S/N but improves quantitation, since all charge states should show the same ratio between heavy and light labeled forms. To determine the quantitation precision, the ratio between Grb2-light and Grb2-heavy was calculated for a single scan and for an averaged scan, composed of 20 separate scans over the elution profile. In both cases the eight most intense charge states were used for the calculation. We then repeated this measurement in five separate HPLC runs to demonstrate the reproducibility of the measurement. In the insert of Figure 3.2, the average ratio is displayed with 95% confidence interval error bars. When averaging 20 scans, the mean standard deviation was 6%, whereas ratios determined from single scans had a standard deviation of typically 18%.

The ratio calculation can be further improved by using the complete elution profile of the proteins (see Figure 3.2b). Furthermore, as Ong et al.[19] and Du et al.[9] have noted, results become more accurate if they are corrected for noise, especially in case of low S/N peaks.

Improved assignment of top-down fragments by SILAC

Besides the accurate quantitation, SILAC can also be used for identification and characterization purposes. The information in fragmentation spectra is valuable in assigning modifications and in improving the reliability of protein identification. Modifications and truncations are more easily detected and more likely correctly assigned. Due to the speed and sensitivity of the orbitrap, such MS/MS spectra are readily obtainable on small proteins[12].

To demonstrate this in a SILAC experiment, we performed data-dependent CID fragmentation in the online format, collecting MS and MS/MS scans in the orbitrap – without microscanning to keep the duty cycle as fast as possible. The MS/MS scans were of relatively low intensity (Figure 3.3), but the resolution and mass accuracy of the fragments was sufficiently high that even low abundant ions could be clearly distinguished from the noise.

When comparing the MS/MS scans of heavy and light labeled Grb2, we observed very similar fragmentation that appeared very useful for assigning the fragments. The mass offset between the heavy and light fragments should exactly represent the number of lysines and arginines of every fragment. Table 1 lists the fragments and their mass deviations from the calculated values. Most of the fragments were internal and likely observed because the disulphide bridge was still present in the protein.

The absolute mass deviation of the fragments was very low, on average 3.2 ppm, and the number of lysines and arginines determined by the mass increment of the heavy labeled fragments matched exactly with the proposed identification. One potential false positive hit could be eliminated both because of its mass deviation of 14 ppm and also because of the fact that

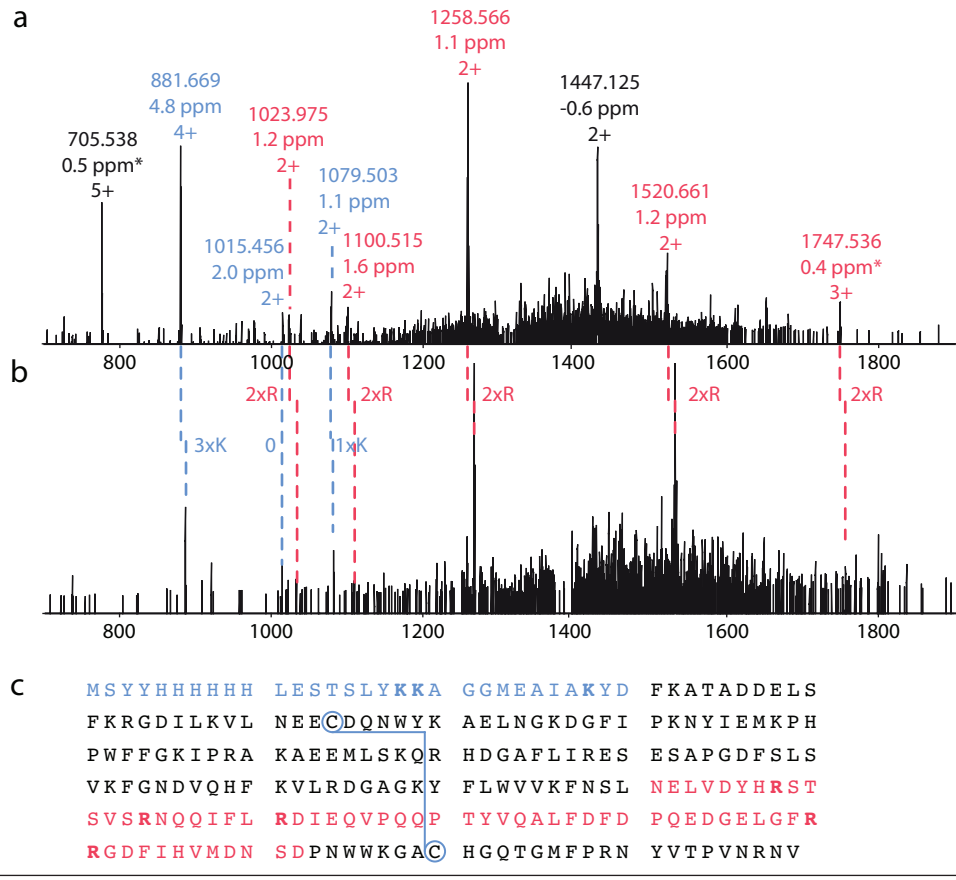


Figure 3.3 Identification of Grb2 is supported by direct comparison of the heavy and light CID fragmentation spectra. Mass differences between corresponding fragments observed in MS/MS spectra of the (a) light and (b) heavy Grb2 protein indicate the number of lysines and arginines per peptide and thereby simplify and confirm the peptide and protein identification. Assigned fragments have an average absolute mass accuracy of 3.1 +/- 2.3 ppm. N-terminal fragments are shown in blue and internal fragments in red. Asterisk indicates that ^{13}C peaks were used for mass accuracy determination. (c) In total the fragments covered 40% of the total protein sequence.

its counterpart in the spectrum of the heavy form of Grb2 did not indicate the correct number of lysines. In total the single scan CID fragmentation depicted in the Figure 3.3 resulted in fragments covering 40% of the protein sequence. It is clear that the SILAC information at the fragment level would have been very valuable to assign any protein modification, had it been present.

Conclusions and perspectives

Here we have investigated the applicability of the SILAC technology to top-down proteomics. Through theoretical modeling, we found that heavy and light SILAC-doublets do not interfere with each other even for very large proteins. The incomplete isotope enrichment of commercial SILAC amino acids

Table 3.1 Fragments of Grb2 identified by Mascot.

Precursor mass (m/z)		Charge	Δmass	Nr of heavy residues	Mass Accuracy (ppm)	Sequence
Light	Heavy					
954.7748		3			5.49	MSYYHHHHHHLESTSLYKKAGGME
978.7882		3			3.90	MSYYHHHHHHLESTSLYKKAGGMEA
1039.831	1045.504	3	16.03	2 Lys	(1.88) [1]	MSYYHHHHHHLESTSLYKKAGGMEAIA
881.6691	887.6808	4	24.04	3 Lys	4.79	MSYYHHHHHHLESTSLYKKAGGMEAIAKYD
705.5377		5			(0.50) [1]	MSYYHHHHHHLESTSLYKKAGGMEAIAKYD
1175.223		3			4.81	MSYYHHHHHHLESTSLYKKAGGMEAIAKYD
1015.456	1015.455	2	0.01	0	2.00	SYYHHHHHHLESTSLY
1079.503	1083.511	2	8.02	1 Lys	1.10	SYYHHHHHHLESTSLYK
1577.204		2			1.67	LFDFDPQEDGELGFRRGDFIHVMDNSD
1520.661	1530.668	2	20.02	2 Arg	1.20	FDFDPQEDGELGFRRGDFIHVMDNSD
1447.125		2			−0.57	DFDPQEDGELGFRRGDFIHVMDNSD
1100.515	1110.52	2	20.04	2 Arg	1.59	PQEDGELGFRRGDFIHVMD
1258.566	1268.573	2	20.03	2 Arg	1.07	QEDGELGFRRGDFIHVMDNSDP
1023.975	1033.983	2	20.03	2 Arg	1.18	GELGFRRGDFIHVMDNS
1747.536	1757.54	3	30.01	3 Arg	(0.42) [1]	FNSLNELVDYHRSTSVSRNQQIFLRDIE-QVPQQPTYVQALFDFD

[1] **^{13}C peaks were used for mass accuracy determination**

does not cause peak broadening but does shift the mass to lower values. This effect can be modeled very precisely. Incomplete labeling, on the other hand, would lead to distribution of the signal into several peaks, substantially broadening them and decreasing the signal to noise. Fortunately, incomplete labeling can easily be avoided in metabolic labeling experiments.

We have shown that a 28 kDa signaling protein, Grb2, can be readily quantified in the labeled versus the unlabeled form. The quantitation in a one to one mixture had a typical standard deviation of six percent and is mainly limited by the lower signal to noise in protein measurements compared to peptide measurements. Grb2 was fragmented and analyzed in an LTQ-Orbitrap instrument using single scan data from online experiments. The high mass accuracy combined with the quantized mass offsets significantly improves fragment identification when comparing tandem mass spectra of light and heavy SILAC labeled protein. The unambiguous information about the number of labeled residues per fragment or protein is a clear advantage of the SILAC technology over ^{15}N labeling.

In our experiment we chose to use two amino acids, arginine and lysine that were isotopically labeled, leading to a one percent mass offset between light and heavy form. Often, it may be more convenient to label with a single amino acid, for example, lysine. This would allow direct 'counting' of the number of lysines in the protein and

between fragments in the 'heavy' and 'light' tandem spectrum. Double amino acids may be valuable when analyzing proteins with potentially a high number of modifications.

Extension of the SILAC technology to intact protein analysis should allow direct quantitation of endogenous and unprocessed proteins. However, the most interesting application could be in the direct quantitation of multiple and combinatorial modifications of regulatory proteins as a function of cellular state. This goal will require SILAC quantitation of fragments isolating these modifications. It may also require non-ergodic fragmentation techniques such as ECD or ETD as well as several stages of fragmentation.

Acknowledgements

We thank Dr Takao and Dr. Fernandez-de-Cossio for their assistance with the Isotopica software. This work was partly funded by 'Interaction Proteome' a 6th Framework grant by the European Union research directorate and by the Max Planck Society for the Advancement of Science.

References

1 Aebersold, R., Mann, M. (2007) Mass spectrometry-based proteomics. Nature 422, 198-207

2 Ong, S. E., Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. Nat ure Chem Biol 1, 252-262

3 Ong, S. E., et al. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1, 376-386

4 Kelleher, N. L. (2004) Top-down proteomics. Anal Chem 76, 197A-203A

5 Bogdanov, B., Smith, R. D. (2005) Proteomics by FT-ICR mass spectrometry: top-down and bottom-up. Mass Spectrom Rev 24, 168-200

6 Whitelegge, J., et al. (2006) Top-down mass spectrometry of integral membrane proteins. Expert Rev Proteomics 3, 585-596

7 Zandborg, L., et al. (2007) ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. Nucleic Acids Res 35(2), W701-W706

8 Gordon, E. F., et al. (1999) Hydropathic influences on the quantification of equine heart cytochrome c using relative ion abundance measurements by electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. J Mass Spectrom 34, 1055-1062

9 Du, Y., et al. (2006) Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. Anal Chem 78, 686-694

10 Veenstra, T. D., et al. (2000) Proteome analysis using selective incorporation of isotopically labeled amino acids. J Am Soc Mass Spectrom 11, 78-82

11 Martinovic, S., et al. (2002) Selective incorporation of isotopically labeled amino acids for identification of intact proteins on a proteome-wide level. J Mass Spectrom 37, 99-107

12 Macek, B., Waanders, L. F., et al. (2006) Top-down Protein Sequencing and MS3 on a Hybrid Linear Quadrupole Ion Trap-Orbitrap Mass Spectrometer. Mol Cell Proteomics 5, 949-958

13 McCaldon, P. and Argos, P. (1988) Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. Proteins 4, 99-122

14 Fernandez-de-Cossio, J., et al. (2004) Automated interpretation of mass spectra of complex mixtures by matching of isotope peak distributions. Rapid Commun Mass Spectrom 18, 2465-2472

15 Kigawa, T., et al. (2004) Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. J Struct Funct Genomics 5, 63-68

16 Davanloo, P., et al. (1984) Cloning and ex-

pression of the gene for bacteriophage T7 RNA polymerase. Proc Natl Acad Sci USA 81, 2035–2039

17 Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nano-electrospray, and LC/MS sample pretreatment in proteomics. Anal Chem 75, 663–670

18 Morand, K., Talbo, G., and Mann, M. (1993) Oxidation of peptides during electrospray ionization. Rapid Commun Mass Spectrom 7, 738–743

19 Ong, S. E., Kratchmarova, I., and Mann, M. (2003) Properties of ^{13}C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). J Proteome Res 2, 173–181

Nanoelectrospray peptide mapping revisited:

Composite survey spectra allowing high dynamic range protein characterization without LC-MS on an orbitrap mass spectrometer

4

Aiping Lu, Leonie F. Waanders, Reinaldo Almeida, Guoqing Li, Mark H. Allen, Jürgen Cox, Jesper V. Olsen, Tiziana Bonaldi, Matthias Mann

Int J Mass Spec 2007 (268), 158–167

Fast and in-depth mass spectrometric (MS) determination of the primary structure of proteins, including posttranslational modifications, remains a challenging task. Proteins are usually digested to tryptic peptides that are measured either by MALDI peptide mapping or by liquid chromatography on-line coupled to tandem MS (LC-MS/MS).

Here we instead analyze peptides by a chip implementation of nanoelectrospray (TriVersa Nanomate) coupled to a linear ion trap-orbitrap hybrid instrument. The C-trap connecting the linear ion trap and orbitrap is filled repeatedly in different m/z ranges with up to a million charges. Each range is analyzed in the orbitrap repeatedly and separately, creating a survey spectrum composed of hundreds of single spectra. The composite spectrum is inherently normalized for different m/z ranges due to their different fill times and retains information on the variability of mass measurement and intensity.

Nanoelectrospray offers analysis durations of more than 30 minutes per microliter of peptide mixture, sufficient for in-depth peptide characterization by high resolution C-trap fragmentation in addition to high sensitivity ion trap fragment analysis. We obtain over thousand fold dynamic range and subfemtomole sensitivity. Automated analysis of digested BSA resulted in sequence coverage above 80% in low femtomole amounts. We also demonstrate identification of seven modified peptides for a histone H3 sample. Static spray allows relative quantitation of the same peptide with different modifications.

Chip-based nanoelectrospray on an orbitrap instrument thus allows very high confidence protein identification and modification mapping and is an alternative to MALDI peptide mapping and LC-MS/MS.

Introduction

During the last few years, efforts in mass spectrometry-based proteomics have concentrated on the qualitative and quantitative analysis of complex protein mixtures[1]. However, most biological mechanisms involve protein modifications, which are not easily or comprehensively picked up in these large-scale experiments[2,3]. In contrast to the few peptides required for identification by MS, the analysis of Post-Translational Modifications (PTMs) in principle requires peptides covering every part of the protein (100% sequence coverage). Furthermore, some modifications may be sub-stoichiometric, even in the purified protein of interest, requiring the analysis of several peptides covering the same sequence stretch.

MALDI Time-Of-Flight (TOF) or MALDI-TOF/TOF is a popular method to identify gel-separated proteins. MALDI sample preparation has been optimized and is rapid and convenient[4]. MALDI-TOF/TOF has been increasingly automated and now allows large number of gel spots to be identified, i.e. in combination with 2D gel electrophoresis. Nevertheless, the trend towards mixture

analysis and quantitative proteomics have made LC-MS/MS 'shotgun' methods increasingly popular[5-7]. In particular, the quality of MS/MS data in LC-MS/MS often makes protein identifications much more specific than with the MALDI method[8]. Further advantages of LC-MS/MS are its high sensitivity as peptides are concentrated into very small peak volumes and the extra information contained in the chromatographic retention time of each peptide. Disadvantages are the dynamic nature of LC-MS/MS, which makes it difficult to do repeat measurements of the same peak as well as to apply several fragmentation techniques during the elution time of typically less than 30 seconds.

In theory, nanoelectrospray[9,10] which is static and allows directed measurements, offers a compromise allowing both ready identification of proteins without LC separation while still offering extremely high accuracy protein identification and mapping of post-translational modifications. The original 'manual' nanoelectrospray has now largely fallen out of favor, mainly because of its low throughput. However, recently nanoelectrospray has been revived in a chip based form, commercially in the form of the Advion TriVersa Nanomate. Here we investigate the combination of this automated nanoelectrospray with a powerful new mass spectrometer, the hybrid linear ion trap - orbitrap[11].

Materials and methods

Sample preparation of protein standards

Unless otherwise specified, chemicals were from Sigma Aldrich. Bovine serum albumin (BSA, 2 mg/mL, Bio-Rad) was diluted to a concentration of 4 pmol/µL with 6 M urea/2 M thiourea, incubated in 1 mM dithiothreitol (final concentration) for 45 min at 56°C for protein reduction and subsequently in 5.5 mM iodoacetamide (final concentration) at room temperature in the dark for 30 min for alkylation. The solution was digested with 1:50 w/w protein amount of endoproteinase Lys-C (Wako) for 4 h at room temperature, then diluted 4x with 50 mM NH_4CO_3 and digested further with 1:50 w/w protein amount of trypsin (Promega) overnight at 37°C. The digestion was stopped by adding 1 percent (v/v) of absolute trifluoric acid. BSA peptides were desalted and stored on RP-C18 StageTip columns[12] and eluted right before mass spectrometric analysis with 50% methanol/0.5% formic acid.

Sample preparation of Histone H3

Complete™ proteases inhibitors (tablet, Roche) were added to all buffers below and the solutions were cooled to 4°C before use. Semi-confluent HeLa cells were collected and resuspended in Buffer-N (15 mM Hepes-KOH pH 7.6, 60 mM KCl, 15 mM NaCl, 0.5 mM EGTA, 10% Sucrose). Lysis was performed by adding 0.2% NP40 and rotating the cell suspension for 10 min at 4 °C. Cell lysates were carefully poured on 20 ml sucrose cushions (20% sucrose in Buffer-N). Nuclear pellets were fractionated upon centrifugation (4000 rpm, 15 min, 4 °C) and washed in PBS. Core histones, together with linker histones and high mobility group (HMG) proteins were then extracted by adding a half volume of ice-cold HCl (0.8 M) overnight with continuous rotation at 4 °C. The sample was centrifuged for 10 min at 12,000 g, and histones and the other acid-soluble proteins remained in the supernatant. Residual histones were re-extracted for 3-4 hours in 0.4 M ice cold HCl, the supernatants derived from the two extractions were pooled and dialyzed against 100 mM ice cold acetic acid. The dialyzed sample was aliquoted, lyophilized, and evaluated for purity and concentration by resuspension in H_2O and by performing SDS-PAGE (18%). About 100 µg histone sample was resuspended in 100

µL 0.1% TFA, 2% acetonitrile and directly loaded onto a reverse phase HPLC column (Jupiter C18, 4.60 mm x 250 mm, 5 µm, 300 Å) (Phenomenex) connected to an Aekta LC-system (Amersham). Individual histones were separated by applying a gradient from 20% to 80% acetonitrile in 0.1% TFA.

The total amount of histone H3 was estimated by SDS-PAGE. A fraction containing 1.5 µg of histone H3 was dried down and redissolved in a buffer composed of 100 mM Tris-HCl, 10 mM $CaCl_2$, pH 7.6 for overnight Arg-C (1:50, w/w) digestion at 37 °C. One half of the peptide solution was desalted and stored using RP-C18 StageTip columns, while the other half was desalted and stored using Strong Cation Exchange (SCX) StageTip columns. Peptides on the RP-C18 column were eluted by 50 µL 80% acetonitrile/0.5% acetic acid, and the peptides on the SCX column were eluted by 50 µL 5% ammonium hydroxide/30% methanol. Both eluates were combined, dried down, and redissolved in 50% methanol/0.5% formic acid (1 pmol/µL or 15 ng/µL) for nanoelectrospray.

Mass spectrometric analysis

All experiments were performed using a linear ion trap - orbitrap hybrid mass spectrometer (LTQ-Orbitrap, Thermo Fisher Scientific) with a TriVersa Nanomate (Advion Biosciences) as ion source. A positive voltage (1.5 kV) is applied on the chip while the mass spectrometer sample orifice remains at 0 kV. The electrostatic field between the chip and the orifice drives the positive ions towards the mass spectrometer. The flow rate is dependent on the chip diameter. When not mentioned otherwise, all results were acquired with the low flow rate chip (ID 2.5µm) from Advion, providing a flow rate of 20 nL/min. At this flow rate, one microliter of sample provided stable static electrospray longer than 30 min, just like in 'classical' nanoelectrospray.

Every sample, consisting of 1 µL of solution was sprayed twice and MS spectra were acquired either by full range acquisition (full scan) or multiple overlapping segmented range acquisition (Selected Ion Monitoring, or SIM scans). For the BSA sample, four segmented SIM mass ranges (300-500, 450-650, 600-800, 750-1350) were recorded. For the histone H3 sample, the four SIM segments were chosen as 300-550, 500-650, 600-750, and 700-900 m/z. MS/MS fragmentation was performed by data-dependent selection of the five most intense peaks in the segmented mass range. Dynamic exclusion was set to 150 seconds, longer than the acquisition time per two overlapping segments.

Data analysis

The Mascot engine was used for mass spectrometry data identification (Matrix Science). BSA peaks were searched in IPI_Human_v3.13 to which the BSA sequence had been added, using 5 ppm precursor mass tolerance, 0.5 Da fragment mass tolerance (maximum mass deviation or MMD[13]), carbamidomethylation (C) as fixed modification, and oxidation (M), N-acetylation, deamidation (NQ), pyro-glutamate (N-term QC) as variable modifications. Up to three missed cleavages were allowed and every fully tryptic, unique peptide ('bold red' in the Mascot report) without a second protein match was accepted as a hit.

Histone H3 peaks were searched in a histone database (276 non-redundant sequences, including different histone proteins/variants, keratins and the proteases used), using 5 ppm MMD for precursor ions, 0.5 Da MMD for ion-trap fragmentation, and 0.01 Da mass tolerance for C-trap fragmentation (minimum possible in Mascot), and seven variable modifications, including methionine oxidation, N-terminal acetylation, mono-, and dimethylation of lysines and arginines

and lysine trimethylation and acetylation. Up to two missed cleavages were allowed and peptides with a score higher than that corresponding to a significance value of p=0.05 were accepted.

Results and discussion

Automated nanoelectrospray coupled to the LTQ-Orbitrap

'Classical' nanoelectrospray requires handling of fragile pulled needles, which is both time consuming and a skill demanding considerable dexterity. In contrast, the TriVersa achieves the same low flow rates and thereby sensitivity using a micro-machined chip that is operated completely automatically. Here we describe operation of the automated nanoelectrospray combined with a high accuracy mass spectrometer, the LTQ-Orbitrap. The TriVersa automatically takes a tip, aspirates the sample, and transfers it to the nozzle of the chip, located in front of the mass spectrometer (Figure 4.1). As can be seen in the Figure 4.1, the LTQ-Orbitrap contains a C-trap, which functions as a container for ions transferred from the ion trap and waiting to be ejected into the high resolution analyzer - the orbitrap. Importantly, the instrument allows any ion populations isolated in the ion trap to be accumulated in the C-trap for final high resolution analysis in the orbitrap. This high resolution scan in the orbitrap takes 0.5 to 1 s, depending on the resolution chosen.

While the instrument is extremely sensitive, its duty cycle is limited by the fact that the C-trap only accommodates 10^6 ions, which is often achieved with ion accumulation for just a few milliseconds. Secondly, the dynamic range is also limited by dominant ions (typically in the low to middle m/z range), which can make up a large fraction of the total ion population. We reasoned that

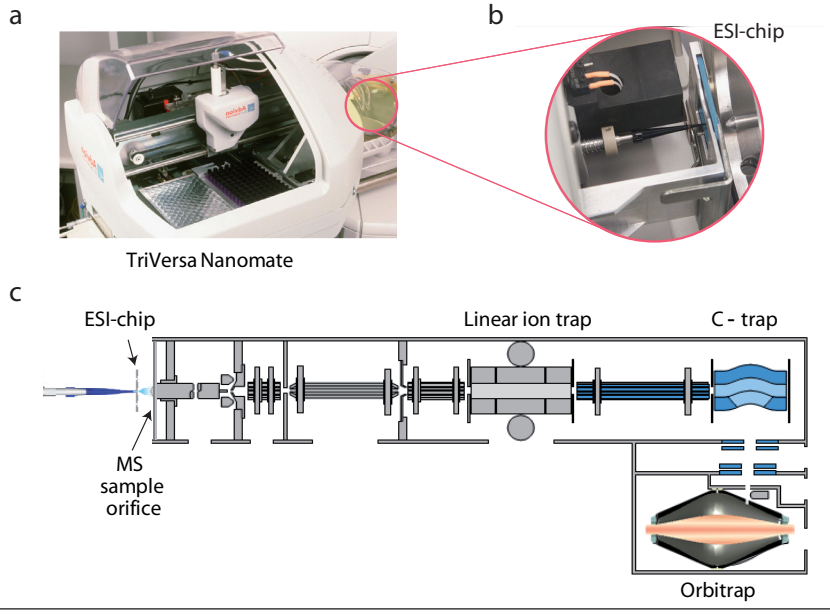


Figure 4.1 Schematics of the TriVersa Nanomate coupled to the LTQ Orbitrap. Samples were applied by coated tips to the nozzle of the electrospray chip in front of the TriVersa instrument. The low flow rate chip (ID 2.5 µm) provided a stable flow rate of 20 nL/min, which is in the 'true' nanoelectrospray range.

the combination of nanoelectrospray and LTQ-Orbitrap should allow us to ameliorate both problems. Instead of acquiring a single full scan spectrum, we decided to acquire a large number of spectra by filling up the C-trap to capacity for each of a number of segmented mass ranges. This should lead to a 'normalized' mass spectrum consisting of a 'matrix' of individual spectra for several mass segments and averaged over many scans. This 'composite' spectrum should have a much larger dynamic range and peptide mass measurement accuracy than a single full scan spectrum or averaged full scan spectra. Furthermore, the long spray time allows directed and iterative peptide fragmentation experiments. Peptides can be identified by peptide mass fingerprinting (PMF), ion trap fragmentation with read out in the ion trap or in the orbitrap, fragmentation in the C-trap or any combination of these. We therefore sought to devise efficient MS/MS schemes to characterize the maximum number of peptide peaks.

Acquisition methods for the composite spectrum and MS/MS acquisition

We found that a three step procedure, encompassing peptide mapping, data dependent sequencing and directed sequencing of 'missing' peaks, was optimal for protein characterization (Figure 4.2). In the first step the full mass range is divided by SIM scans into multiple overlapping segments (several hundred m/z units wide), which were acquired in the orbitrap. The segmented mass ranges, shown in the Figure 4.2a, were chosen so that the accumulation time for every segment would be similar. Each mass segment window was measured many times to gain sensitivity and precision. If several minutes are allocated to acquisition of the composite survey spectrum, then each mass segment is typically acquired more than hundred times.

First identification is based on PMF analysis. Because of the high mass accuracy of the orbitrap, particularly when including an internal mass standard in each spectrum (see below), dominant proteins in the sample are readily identified at this stage. In the second step data dependent fragmentation is performed in each segmented range. Again SIM survey scans are recorded for each m/z range but now they are followed by ion-trap MS/MS spectra of the five most intense peaks. For each m/z range, the SIM-MS/MS cycle is repeated for a time adapted to the number of MS/MS candidates (Figure 4.2b).

Peptide identification is performed in the MS/MS ion search mode and peaks identified by PMF are confirmed by the MS/MS ion search. Since MS/MS spectra contain more information than the peptide mass alone, many peaks that cannot be identified only by the precursor mass are identified at this stage. This analysis still leaves some peptide peaks unfragmented – mainly because of their low signal, which may mean that they do not appear in every scan. These peaks are then targeted by a so-called 'inclusion list' in the third part of the measurement sequence. Figure 4.2c presents an overview of the three step sequence.

Since the acquisition methods for composite full spectra and SIM-MS/MS take only four and five minutes respectively, one microliter of sample sprays long enough to record both steps three times. A second microliter is used for step three in which we specifically target peaks not fragmented yet. Several microscans are applied for both MS (SIM) and MS/MS acquisition to boost sensitivity and data quality.

Subfemtomole sensitivity

Having established an efficient protocol for comprehensive protein characterization, we wanted to assess its sensitivity on a model protein. Using the strategy as described in

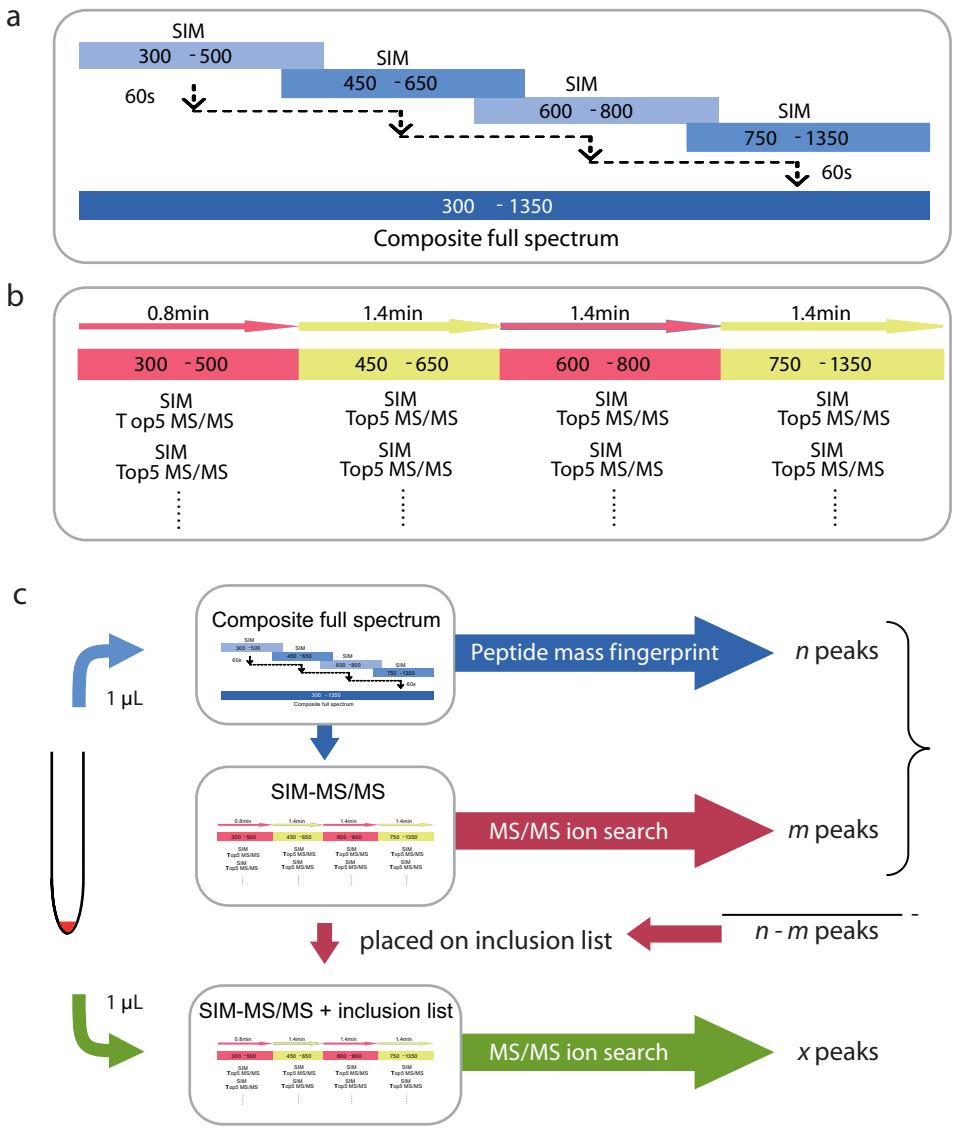


Figure 4.2 Schematic description of the acquisition strategy as performed for protein characterization of tryptic digested BSA. (a) Selected Ion Monitoring (SIM) scans of multiple mass segments were repeatedly analyzed in the orbitrap mass spectrometer and these multiple SIM acquisitions were combined in one 'composite' spectrum. (b) After each SIM scan, the five most intense ions were data-dependently selected for MS/MS fragmentation in either ion-trap or orbitrap. (c) Overview of the complete protein characterization method, including the acquisition of the composite spectrum and the data- dependent SIM-MS/MS mode. These two methods are followed by a third strategy that aims to identify peptides that were not sequenced yet. In this directed SIM-MS/MS mode all precursor masses that were not fragmented so far are placed on an inclusion list. The entire experiment can be carried out with only two microliters of very diluted sample solution.

Figure 4.2c, we obtained a sequence coverage of more than 80% for BSA. The missing peptides were generally very short and some of them did appear under different digestion conditions as 'missed cleavage' peptides. We found that the BSA concentration could be diluted to 25 fmol/µL without losing protein sequence coverage (data not shown). Illustrating the excellent sensitivity of the set up described in this paper, more than 60% of the BSA sequence was still identified when the protein was diluted to 500 amol/µL (see Table 4.1). As shown in the table, the inclusion list SIM-MS/MS method turned out to be particularly advantageous for lower protein concentrations.

Extremely high mass precision in the composite full spectra

The orbitrap detection is based on inherently very precise frequency measurement and is, in our experience, much less affected by space charge than a Fourier-transform ion cyclotron resonance (FT-ICR) instrument. When combined with 'lock mass' injection in every spectrum, the orbitrap is capable of achieving low to sub-ppm mass accuracy[14]. We reasoned that, by measuring each SIM mass range multiple times, the mass values of peaks extracted from the composite full spectra should become even more precise, since the standard deviation of a population is inversely dependent on the square root of the number of measurements. In our case,

Table 4.1 BSA sequence as identified by the three-step method introduced in this paper. In total 2 µL of trypsin digested BSA sample was used for every concentration. One microliter BSA was used to generate the composite SIM spectra and to carry out the data-dependent SIM-MS/MS method. Both methods were performed three times to assess their reproducibility. The second microliter was used specifically to characterize the peptides only identified by PMF but not sequenced yet. When all identification methods were combined, a sequence coverage of 66% was reached even for 500 attomole/µL.

BSA concentration	Acquisition methods	Sequence identified			
25 fmol/µL	Composite SIM	87%	88		88%
		79%			
		81%			
	SIM- MS/MS	63%	75%	77%	
		73%			
		72%			
	Inclusion list MS/MS	1.4%			
5 fmol/µL	Composite SIM	73%	82%		83%
		77%			
		78%			
	SIM- MS/MS	54%	65%	75%	
		51%			
		60%			
	Inclusion list MS/MS	10.1%			
500 amol/µL	Composite SIM	45%	60%		66%
		42%			
		49%			
	SIM- MS/MS	24%	35%	46%	
		20%			
		24%			
	Inclusion list MS/MS	10.5%			

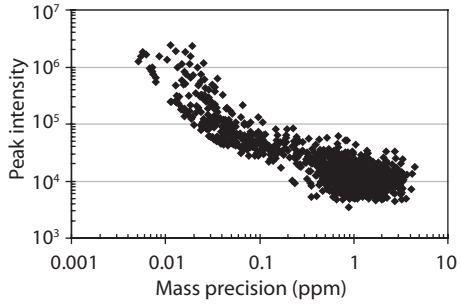


Figure 4.3 Mass precision correlated to ion intensity. Plot of the intensity of different ions extracted from the composite spectrum versus their precision, from thousands of scans depicted on a double logarithmic scale. Note that for most peaks we achieve sub ppm mass precision.

by repeated measurement of the same spectrum for one hundred times, the standard deviation of the mass value should be decreased by a factor of ten.

In Figure 4.3, the peak intensity is plotted versus peak precision on a double logarithmic scale, and the data show that more intense ions yield better precision within the same measuring time. As can be seen in the Figure 4.3, a large percentage of the peptides have a precision well below one ppm, with some peptides even reaching 100 ppb or less. Thus we conclude that the TriVersa-LTQ-Orbitrap combination is capable of extremely high mass accuracy, comparable or superior to any other platform currently used in proteomics.

More than 6000 fold dynamic range in the composite full spectra

Summing up multiple spectra filters out noise but boosts low intense ions that can not be distinguished from background peaks in single spectra. We demonstrate further improvement of the dynamic range by collecting segmented m/z ranges instead of one full spectrum. Often a few, very intense ions comprise 90% of the total ion number of a spectrum, whereas in the segmented mass range strategy, ranges with low intense signals are accumulated for a longer time in order to reach the same specified injection target value.

As depicted in Figure 4.4a, the composite spectrum is much more feature rich in the higher mass range compared to the spectrum consisting of averaged full mass range spectra with the same total acquisition time. Both the composite spectra and the summed-up full spectra were acquired in the same experiment. As is apparent from the Figure 4.4, the S/N was much better (see arrows). With the 25 fmol/µL tryptic digested BSA sample, we assigned BSA peaks with a signal intensity difference of up to 6700 (Figure 4.4b). While the dynamic range of the orbitrap is specified at 10^4, this value applies only to a simple two component mixture. In our experience, dynamic range in complex peptide mixture analysis is around 10^3 in LC-MS experiments, so the composite spectra exhibit a comparable or superior dynamic range for complex samples to online experiments.

Sequence coverage comparable to LC-MS

We compared the sequence coverage obtained after 5 min of SIM-MS/MS acquisition with a conventional LC-MS run, both times using 50 fmol of BSA and the same parameter settings on the LTQ-Orbitrap. For this experiment, a higher flow rate chip (ID 5 µm) was used for nanoelectrospray, resulting in a flow rate of 200 nL/min. The LC-MS run took in total 53 minutes, of which the actual gradient lasted for 28 minutes. Both methods were performed twice. The sequence coverage obtained by these two methods was very comparable, 83.7% for nanoelectrospray SIM-MS/MS and 78.5% for LC-MS/MS. Detailed identification information for each tryptic peptide in the BSA sequence is shown in Figure 4.5. Most of the peaks were identified by both methods,

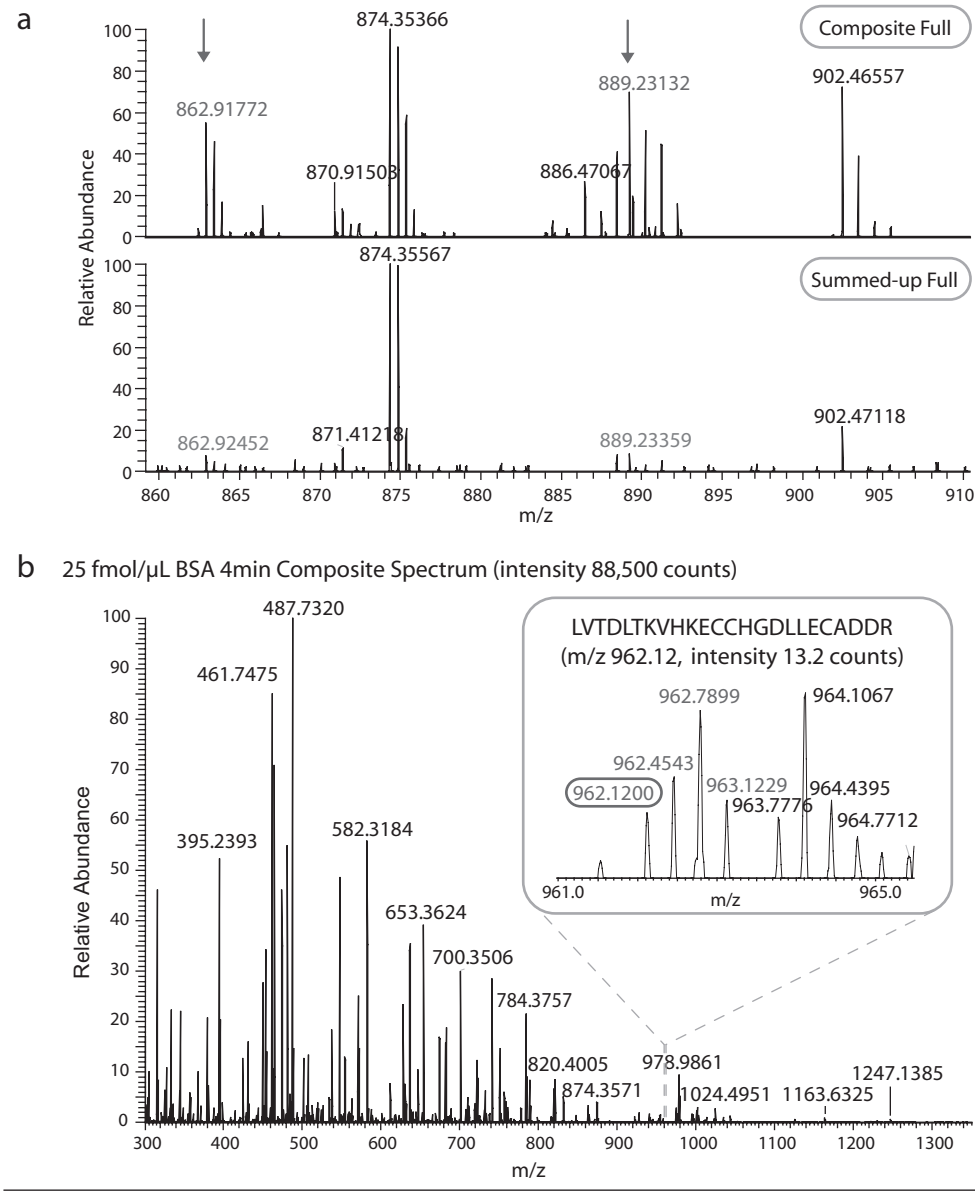


Figure 4.4 Advantages of composite spectra. (a) In the composite spectrum ions of low intensity were boosted in comparison to full spectra summed-up for the same time span. 500 attomole/µL BSA sample was sprayed, and the composite spectra and the summed-up full spectra were acquired directly after one another. Whereas for intense ions there is no large visible difference, the S/N ratio for low intense ions increases dramatically. (b) A dynamic range of over six thousand fold was obtained in the composite spectrum. 25 fmol/µL BSA resulted in a composite spectrum, in which the most intense peptide with m/z of 487.73 (2+, DLGEEHFK) had an intensity of $8.85 \cdot 10^4$, while the peptide at m/z 962.12 3+, LVTDLTKVHKECCHGDLLECADDR) was observed with an intensity of 13.2 (inset), resulting in a dynamic range of 6700.

Figure 4.5 Comparison of the BSA sequence coverage obtained in SIM-MS/MS versus LC-MS/MS mode. In both experiments 50 fmol of BSA solution were used and both measurements were performed with the same instrument settings. Peptides in blue are identified in both methods (421 residues, 72.2%), red peptides only in SIM-MS/MS (67 residues, 11.5%) and green peptides only in LC-MS mode (37 residues, 6.3%). Black peptides were not identified (59 residues, 9.9%).

but some low intensity peaks were only sequenced in LC-MS. This may be due to the concentration effect of chromatography, where each ion elutes in a very short time span in contrast to the long but 'diluted' duration in nanoelectrospray. On the other hand, there were a few peptides that co-eluted with others and disappeared before having a chance to be fragmented in LC-MS but those were sequenced in the SIM-MS/MS run.

Characterization of Histone H3 post-translational modifications

Histones are the protein constituents of nucleosomes around which DNA is wound in eukaryotic cells. Histone tails on the nucleosome are subject to enzyme-mediated posttranslational modifications (PTMs) of selected amino acids, such as lysine acetylation, lysine and arginine methylation, serine phosphorylation and attachment of small ubiquitin peptides[15,16]. These modifications, singly or in combination, are thought to generate an epigenetic code that specifies different patterns of gene expression and silencing[17].

Characterization of post-translational modifications on bulk histones by employment of mass spectrometry approaches has proven to be very successful as recently reviewed by Hunt and co-workers[18]. Here we investigate the suitability of the nanoelectrospray-orbitrap combination to characterize modifications on histone H3 purified from human HeLa cells, separated from other histone molecules by RP-chromatography and in-solution digested with Arg-C protease. In order to distinguish between several modifications present on such molecules, many of them only differing in single methyl or acetyl groups, we also employed high resolution read out of MS/MS spectra in the orbitrap. Fragmentation by higher energy injection into the C-trap[19] leads to triple-quadrupole like behavior and preservation of the full mass range in the MS/MS spectra. Fragmentation spectra were acquired at a resolution of 30,000 and the mass accuracy was in the low ppm range for these fragmentation spectra.

More than 500 peaks were extracted from the composite full spectra. Spectra recorded in SIM-MS/MS mode were searched in a histone database and 46% of the sequence of

histone H3 was identified (based on peptides with a score higher than $p<0.05$). The heavily modified N-terminal sequence was completely covered and seven differently modified peptides were detected (Table 4.2). In six of them the modified residues were unambiguously determined. In particular, the high mass accuracy of the orbitrap allowed easy distinction between trimethylation and acetylation, both of which are important histone modifications that have the same nominal mass. In the seventh peptide, which was the both trimethylated and acetylated peptide KSTGGKAPR, the modified sites could not unambiguously be assigned to the sequence by the MS/MS spectrum acquired in the LTQ, as the mass difference between these modifications (0.0364 Da) is far less than what the LTQ can distinguish. Therefore, a second microliter of the sample was sprayed and C-trap fragmentation was performed combined with recording of the MS/MS spectra in the orbitrap. With the resulting high mass accuracy in the fragmentation spectra, both types of modifications were confirmed and trimethylation was assigned to K9 and acetylation to K14.

The reported seven modified peptides are relatively abundant in the bulk preparation histone sample, and have already been reported by either top-down[20] or LC based bottom-up[18] mass spectrometric method. Our approach provides an alternative way to characterize histone by bottom up mass spectrometric analysis without online LC separation. Preliminary work furthermore indicates that modifications on short peptides that escape detection by LC-MS/MS can be detected by automated nanoelectrospray (data not shown).

Relative quantitation by deconvoluted peak intensity

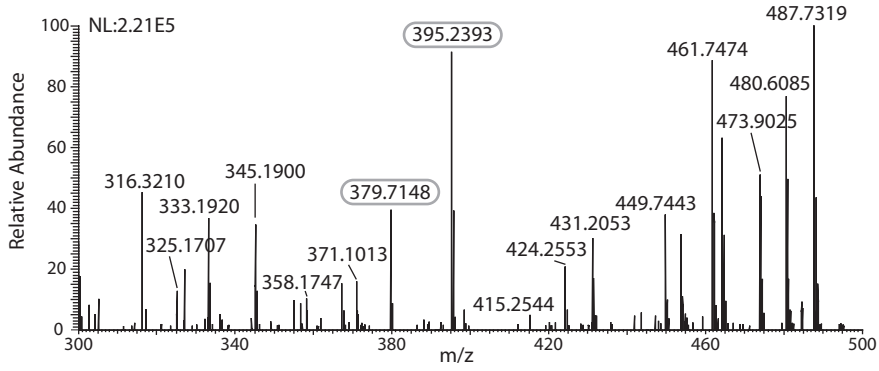
As mentioned above, quantification of peptides and proteins is becoming more and more important. An advantage of the acquisition of a large number of spectra is the increasing precision, not only of the mass value but also of the intensity ratio between ions. This is illustrated by the ratio determination between two BSA peptides as well as between a methylated and non-methylated histone peptide. Figure 4.6 illustrates the relative quantitation of the BSA peptides with the m/z 395.239 (2+) and 379.715 (2+). As shown in Figure 4.6a, both peaks are present in the spectra of m/z segment 300 to 500. Figure 4.6b reveals that the ratio of the relative intensities varied per spectrum between 1.7 and 3.5. However, with increasing number of accumulated spectra quantitation becomes more and more precise. As shown in Figure 4.6c, the 99% confidence interval for quantitation decreases from 13% after accumulating 10 scans to 0.9% after accumulation of 1500 scans (15 minutes acquisition).

In the case of histone H3 we investigated quantitation of the normal peptide against a slightly modified form. The advantage of quantitation by nanoelectrospray is that both peptides are present in the same scan, whereas in LC-MS/MS experiments the

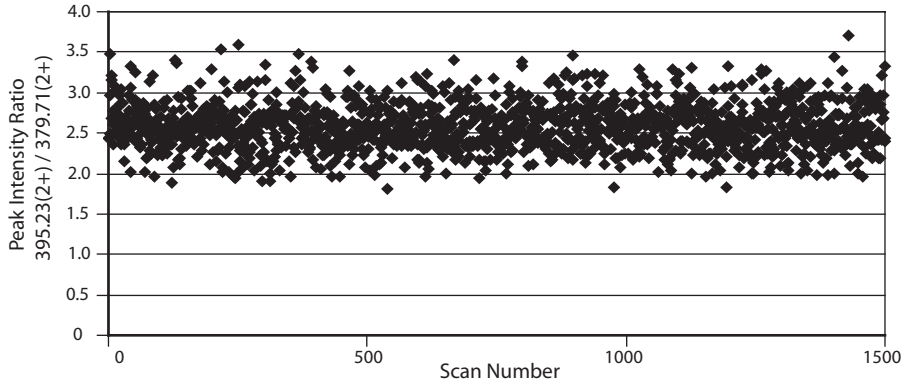
Table 4.2 Histone H3 modifications (combinations) identified by the SIM-MS/MS method.

Modified amino acids	Sequence
K4 Monomethylation	TKQTAR
K9 Dimethylation + K14 Acetylation	KSTGGKAPR
K9 Trimethylation + K14 Acetylation	
K23 Acetylation	KQLATKAAR
K18 Acetylation + K23 Acetylation	
K79 Monomethylation	EIAQDFKTDLR
K79 Dimethylation	

a



b



c

Number of scans	Average of peak intensity ratios	Quantitation ratio with 99% confidence interval
10	2.748	2.748 ± 0.354
50	2.623	2.623 ± 0.124
100	2.570	2.570 ± 0.086
200	2.546	2.546 ± 0.058
300	2.524	2.524 ± 0.051
400	2.526	2.526 ± 0.047
500	2.513	2.513 ± 0.039
1000	2.511	2.511 ± 0.027
1500	2.513	2.513 ± 0.022

Figure 4.6 Highly accurate relative quantitation by the ratio of ion intensities within one segmented spectrum. Segmented SIM scans from m/z 300 to 500 were acquired for 15 min with 25 fmol/µL tryptic BSA. (a) A single SIM scan including the ions with m/z 395.239 (2+) and 379.715 (2+) of which the ratios of intensities were quantified. (b) Per single scan the ratio varies from 1.5 to 3.5. (c). With an increasing number of accumulated spectra, the precision of the quantitation ratio increases.

Table 4.3 Ratio of peptide pairs with identical sequence but different modification state.

Identified Peptide Pairs	Ratio with 99% confidence interval
TKQTAR / $TK_{(Monomethyl)}QTARK$	11.791 ± 0.916
$K_{(Dimethyl)}STGGK_{(Acetyl)}APR$ / $K_{(Trimethyl)}STGGK_{(Acetyl)}APR$	9.849 ± 0.826
$KQLATK_{(Acetyl)}AAR$ / $KQLATK_{(Diacetyl)}AAR$	14.651 ± 1.307
EIAQDFKTDLR / $EIAQDFK_{(Monomethyl)}TDLR$	11.391 ± 0.868
$EIAQDFK_{(Monomethyl)}TDLR$ / $EIAQDFK_{(Dimethyl)}TDLR$	8.206 ± 0.990

methylated peptide elutes typically slightly later. In order to avoid inaccuracy due to transmission 'edge effects' in the SIM windows, we chose to quantify based on the full spectrum. The ratio of the identified histone H3 peptide pairs are listed in Table 4.3. As is apparent from the table, every (additional) modification was subjected to a small fraction (7–12%) of the peptides molecules. Note that this value gives a general idea of the absolute stoichiometry of this methylation site but that it needs to be corrected for the different ionization efficiencies of the modified vs. the unmodified peptides[21].

Conclusions

In this paper, we have endeavored to revive nanoelectrospray, an 'old' protein mapping method using no chromatographic peptide separation, as an alternative to MALDI peptide mapping. Using the advantages of a stable spray in combination with the LTQ-Orbitrap mass spectrometer, we have introduced the concept of 'composite spectra', which are spectra composed of a high number of segmented SIM scans. These composite spectra allow very high sensitivity, accuracy and dynamic range due to optimized C-trap fill times for each mass segment. In the automated format of the TriVersa Nanomate, nano-electrospray measurements are robust, user-friendly and easily amenable for different protein samples while using very low amounts of spraying solution. Since the instrument can readily switch between MS and MS/MS, and between fragmentation in the linear ion trap and the C-trap, it allows for a large number of complementary methods and is very flexible and user-friendly.

The LTQ-Orbitrap mass spectrometer is a dual instrument with two independent detection systems (LTQ and orbitrap), which can be operated simultaneously, thus the ideal combination would be to fragment high intense ions in the orbitrap while simultaneously performing ion-trap fragmentation of low intensity ions. This would increase the duty cycle and analysis speed. However, this requires direct access to the LTQ-Orbitrap acquisition software, which we are currently lacking. At the other extreme of acquisition sophistication is the so called 'ion-mapping' technique. In this method, the whole mass range is scanned step by step by SIM-scans of for example 6 Dalton windows with and without applying collision energy to fragment the ions in this small window. Especially for complex mixtures this method could in principle be very valuable in a nanoelectrospray setup since the dynamic range is expected to be further increased by these small segments. However, when we tried this method we found that it allocates too much time to 'empty' regions and is thus overall less efficient than the method described here.

In order to further improve protein characterization, we plan to access to the LTQ-Orbitrap acquisition software directly and perform genuine 'real time' data acquisition. Fragmentation (MS/MS or MS^3)

will focus on the peaks recognized in the survey scan but not identified in the search with expected variable modifications. This will allow identification of new peptide sequences, variant alleles or unexpected modifications. Of course, digestion with multiple enzymes is also an obvious next step for even more for more in-depth characterization of modified peptides.

Acknowledgments

This work was partially funded by 'Interaction Proteome' and 'HEROIC', two 6th Framework grants by the European Union research directorate.

References

1 Aebersold, R., Mann, M. (2003) Mass spectrometry-based proteomics. Nature 422, 198-207

2 Mann, M., Jensen, O. N. (2003) Proteomic analysis of post-translational modifications. Nat Biotechnol 21, 255-261

3 Jensen, O. N. (2006) Interpreting the protein language using proteomics. Nat Rev Mol Cell Biol 7 , 391-403

4 Kussmann, M. and Roepstorff, P. (2000) Sample preparation techniques for peptides and proteins analyzed by MALDI-MS. Methods Mol Biol 146, 405-424

5 Yates, J., et al. (1997) Direct Analysis of Protein Mixtures by Tandem Mass Spectrometry. J Protein Chem 16, 495-497

6 Peng, J. and Gygi, S. P., (2001) Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC-MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome J Mass Spectrom 36, 1083-1091

7 Ong, S. E. and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. Nature Chem Biology 1, 252-262

8 Steen, H. and Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. Nat Rev Mol Cell Biol 5(9), 699-711

9. Wilm, M. and Mann, M. (1996) Analytical properties of the nanoelectrospray ion source. Anal Chem 68, 1-8

10 Wilm, M., et al. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry Nature 379, 466-469

11 Makarov, A., et al. (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. Anal Chem 78 , 2113-2120

12 Rappsilber, J., Ishihama, Y. and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem 75, 663-670

13 Zubarev, R. and Mann, M. (2007) On the proper use of mass accuracy in proteomics. Mol Cell Proteomics 6, 377-381

14 Olsen, J.V., et al. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics 4 , 2010-2021

15 Spotswood, H. T. and Turner, B. M. (2002) An increasingly complex code. J Clin Invest 110, 577-582

16 Kouzarides, T. (2007) Chromatin modifications and their function. Cell 128, 693-705

17 Jenuwein, T. and Allis, C. D. (2001) Translating the histone code. Science 293, 1074-1080

18 Garcia, B. A., et al. (2007) Organismal Differences in Post-translational Modifications in Histones H3 and H4. J Biol Chem 282(10), 7641-7655

19 Olsen, J. V., et al. (2007) Higher-energy C-trap dissociation for peptide modification analysis. Nature Methods 4, 709 - 712

20 Thomas, C. E. , Kelleher, N. L. and Mizzen, C. A. (2006) Mass spectrometric characterization of human histone H3: a bird's eye view. J Proteome Res 5, 240-247

21 Steen, H., et al. (2006) Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. Mol Cell Proteomics 5, 172-181

5

A novel chromatographic method allows online reanalysis of the proteome

Leonie F. Waanders*, Reinaldo Almeida*, Simon Prosser, Jürgen Cox, Daniel Eikel, Mark H. Allen, Gary A. Schultz, Matthias Mann

Mol Cell Proteomics 2008 (8), 1452-1459

** Shared first co-authorship*

Liquid chromatography combined with electrospray ionization is widely used for direct analysis of polar and labile molecules by mass spectrometry (LC–MS). The online coupling in LC–MS is a major strength but also causes a principal limitation – that each eluting analyte has to be analyzed immediately and is not available for detailed interrogation after the LC–MS run.

Here we develop a new chromatographic strategy, which removes this limitation. After column separation the flow is split, one portion is analyzed directly and the other is diverted to a capture capillary. After the direct LC–MS run, the flow is switched and the portion stored in the capillary is analyzed ('replay run'). We describe a set up consisting of an analytical column, a splitting valve and a focusing column, which performs at full sensitivity and undiminished chromatographic resolution.

We demonstrate three principal advantages of this system; nearly continuous MS utilization, duplicate analysis without requirement for additional sample and targeting of important but undersampled features in the replay run.

Introduction

The development of electrospray mass spectrometry has allowed analysis directly from the liquid phase[1]. This feature of electrospray makes it eminently suitable for the online coupling of separation and ionization before MS analysis. In nanoscale liquid chromatography–mass spectrometry (LC–MS) analyte species are concentrated into very small volumes, increasing sensitivity. Furthermore the excellent separation capacity of high performance chromatographic systems is multiplied with the high resolution of modern mass spectrometers, resulting in exceptional combined separation power[2]. Both small molecules and peptides and proteins are routinely analyzed by this powerful technology. However, compared to offline methods such as nanoelectrospray[3] or MALDI[4], online coupling also has some inherent drawbacks. The short time during elution of a peak requires a fast and automatic decision on which peak to sequence. In complex mixtures, many peptides co-elute and some may not be sequenced at all[5]. Peptides of special interest, for example, those with regulatory post-translational modifications, should be characterized in depth, but the fact that they are important may only become evident after the analysis.

Some of these drawbacks can in principle be addressed by slowing down the flow ('peak parking', [6]), fraction collection or by repeat injection. However, none of these methods is ideal from an analytical standpoint. Peak parking is of limited utility for complex mixtures as the flow may have to be stopped every few seconds and the run would be extended to impractical lengths. Fraction collection is useful in many instances but at very low flow rates it is less practical because of the low volume of fractions. In nanoflow LC peptides typically elute in about

50 nL (based on an elution time of 15 s using 200 nL/min flow rate). To handle these fractions one would need approximately 1 µL of sample, but adding buffer reduces the concentration 20-fold causing a dramatic loss of signal intensity. Repeat injections multiply total required analysis time and may not be optimal if sample is limited, since re-injection consumes twice as much sample.

We wished to develop a novel concept in LC-MS that would allow targeted measurement of analyte mixtures without compromising sensitivity or chromatographic performance while requiring little or no additional time. Since electrospray is a concentration dependent process and therefore maintains full signal at decreased flow rates, we and others previously developed splitting systems in which the column effluent was directed to MS analysis as well as to a fraction collector to enable the reanalysis of chromatographic fractions[7,8]. However, at very low flow rates, fraction collection became increasingly difficult which prompted us to explore alternative ways of storing the chromatographically separated sample.

Here we describe a novel setup in which we collect part of a column effluent in a long capture capillary, which we reanalyze after the direct run. In this paper we describe the new concept and evaluate the system in terms of chromatographic and mass spectrometric performance. Furthermore we show the applicability for a complex proteomics sample and demonstrate a useful application: targeting important peptides that were not characterized in sufficient detail in the direct run. A 'head to head' and exhaustive comparison with other possible LC-MS setups or MALDI methods is not the subject of this paper.

Material and methods

RePlay Setup

We constructed the 'RePlay system' consisting of a 6-port splitting valve, a flow sensor, a long capillary serving as a capture capillary and short capillaries to adjust the split ratio (see Figure 5.1). The splitting valve was specially constructed for accurate flow ratios and extremely low dead volumes at nL/min flow rates (Advion Biosystems). LC was performed on a nano-HPLC 1200 (Agilent) with a 10 cm long 75 µm ID IntegraFrit™ ProteoPepII analytical column (5 µm RP-C18 resin, New Objective), coupled to the RePlay valve in which the flow was split. The gradients were essentially as described previously[12], with peptides eluting from 13% to 60% of solvent B (0.5% acetic acid in 80% acetonitrile). One part of the effluent was directed to a 7 cm in-house pulled 75 µm ID fused silica emitter packed with ReproSil-Pur C18-AQ 3 µm resin (Dr. Maisch GmbH), termed 'focusing column' for direct online LC-MS analysis on a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific). Meanwhile, the second part of the effluent was collected in the fused silica capture capillary of typically 13 m length and 30 µm inner diameter (ID). The capture line (Composite Metal Services Ltd) was capable of storing around 10 µL, appropriate for a 100 min gradient at an effective flow rate of 100 nL/min. For the 30 min gradient a 3 m capillary was used. This stored gradient was then directed to the same 7 cm pulled column by valve-switching triggered from the MS acquisition software.

Split adjusters, 10 µm ID capillaries (Composite Metal Services Ltd), were cut to the length resulting in the desired split ratio. These split adjusters needed to be fine-tuned whenever the focusing column was replaced. We recommend to carefully tighten – not to overtighten – these capillaries since we found that small glass particles easily

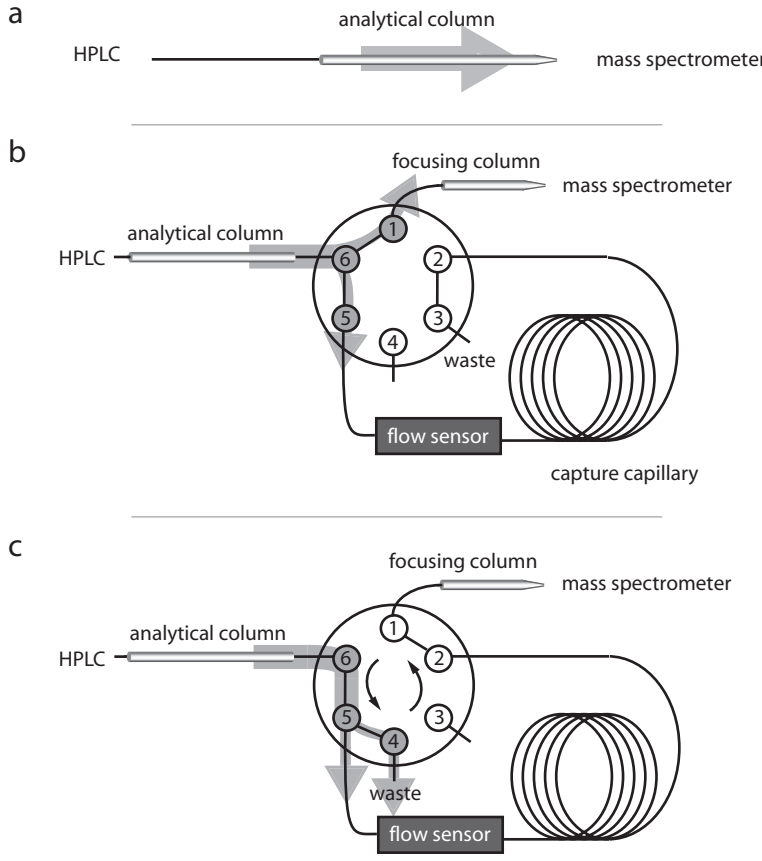


Figure 5.1 Schematics of the conventional LC–MS and the RePlay setup. (a) Conventional LC–MS setup using a column with integrated emitter placed in front of the mass spectrometer. The light grey arrow indicates the direction of the flow. (b) In the 'direct run' of the RePlay setup the six-port splitting valve is positioned such that part of the effluent of the first nanoLC column ('analytical column') flows to the mass spectrometer via a short second column ('focusing column'). The other part is stored in a long capture capillary, which has the volume appropriate to hold the complete gradient. (c) When the valve is switched the stored gradient is directed to the mass spectrometer ('replay run'). Capture capillaries of different lengths serve as split adjusters – at port 3 in the direct run and port 4 in the replay run – and control the split ratio, which is read out by the flow sensor. The split adjuster at port 4 can be replaced by a plug, but a split allows higher flow rates and reduces the time used for washing and loading of the analytical column.

clog the capillaries with a small ID (used as split adjusters). For calculation of the theoretical peak broadening we used the 'capillary flow calculator' option of the 'Molecular Weight Calculator', software freely available on the internet (http://www.alchemistmatt.com). This program calculates the width of the peak as a consequence of diffusion when inserting the initial peak width, the length and ID of the capillary and the flow rate.

Liver proteome analysis

Frozen mouse liver tissue was homogenized and in-solution digested by trypsin as previously described[9], desalted and concentrated on in-house prepared StageTips[10] and eluted for LC-MS analysis. With a split ratio of 1:1, we ran a 100 min gradient (100 nL/min to MS) and reanalyzed the sample in another 100 minutes (100 nL/min). We also split 3:1 (direct:replay), used a 100 min direct gradient (200 nL/min to MS) and reanalyzed the sample in 35 min (200 nL/min). During the replay run the LC loaded the next sample on the analytical column. The run was analyzed by in-house developed MaxQuant software (version 1.07.5) essentially as described in ref. [11]. The data was searched using Mascot (version 2.1.04, Matrix Science Ltd) against the mouse IPI database (version 3.37) supplemented with frequently observed contaminants and concatenated with reversed copies of all sequences (2 x 51,467 entries). Enzyme specificity was set to trypsin, allowing for cleavage N-terminal to proline and between aspartic acid and proline[12]. Carbamidomethylcysteine was set as fixed and N- acetylation and methionine oxidation were set as variable modifications. The maximum allowed mass deviation (MMD, [13]) was set to 5 ppm for monoisotopic precursor ions and 0.5 Da for MS/MS peaks. Maximally three missed cleavages were allowed. The false positive rate at the peptide level and false discovery rate at the protein level were set to 1 % and the required minimum peptide length was 6 amino acids. If the identified peptide sequence set of one protein was equal to or contained another protein's peptide set, these two proteins were grouped together by MaxQuant and they were not counted as independent protein hits. Proteins were considered identified when at least two peptides were identified (of which one uniquely assignable to the respective sequence).

Determination of SUMO-2 interaction partners

HIS_6-SUMO-2 conjugates were purified as described previously[14]. The samples were subsequently digested by trypsin and concentrated on in-house made StageTips[10]. After measuring the direct run with a 'top 10' method employing fragmentation of the 10 most abundant ions in the ion trap, a peak list was created with MaxQuant software (version 1.07.5) using default parameters. This list was screened for precursors, that – when fragmented – showed ions with m/z corresponding to SUMO-2 b-ions in the range of b_5 to b_{18}. Of the precursors that had more than three fragment matches, and with a total mass larger than a free SUMO-2 peptide, the highest expected isotope peak was added to an inclusion list. In the replay run only these ions were sequenced by use of higher collision dissociation (HCD) fragmentation[15]. We deselected charge state, precursor mass screening and dynamic exclusion, and fragmented all ions matching an inclusion list item with a 30 ppm tolerance and with a minimum intensity of 30,000. These high-resolution fragmentation spectra were validated manually.

Results and discussion

RePlay setup

In a first iteration of our systems we used a storage capillary instead of the fraction collector and we reversed the flow for reanalysis of the stored gradient. This arrangement inverted the order of the elution and resulted in good resolution of late eluting peaks (those stored in the capillary for a short time) but not of early eluting peaks (data not shown). To enable First In First Out analysis, we combined the functions of flow splitting and redirection of the flow into a single device, a low-flow, low dead

volume splitting valve (Figure 5.1). Finally, in order to prevent possible deterioration of chromatographic performance in the second run due to diffusion in the capture capillary, we introduced a second 'focusing' column prior to the electrospray emitter. We termed this system 'RePlay'. The effluent from the analytical column was split in the appropriate, fixed ratio for direct analysis in the mass spectrometer and for storage in a capture capillary (Figure 5.1a). At the end of the gradient, the valve was switched and the stored gradient separation was eluted through the same focusing column as in direct mode (Figure 5.1b). We tested the performance of the RePlay system (Advion BioSystems) with a bovine serum albumin (BSA) standard (Sigma-Aldrich). As shown in Figure 5.2, peak intensity and chromatographic resolution was very similar between direct and RePlay analysis with no deterioration over the gradient (Figure 5.2c). The fact that the signal in electrospray is concentration dependent rather than dependent on total analyte amount, readily explains this feature of the RePlay setup.

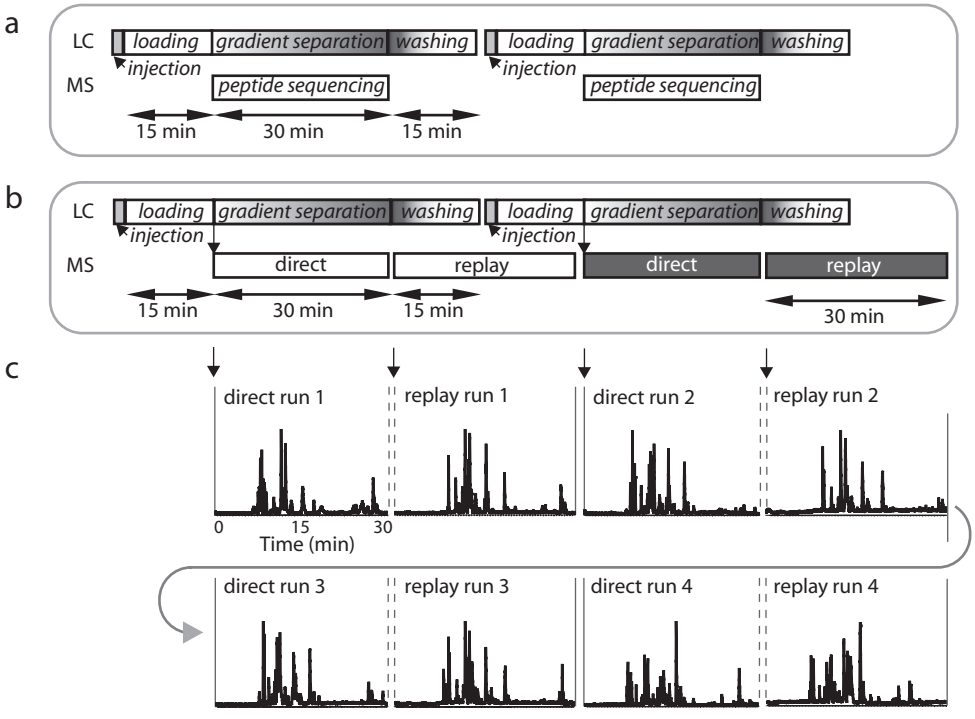


Figure 5.2 Schematics of a conventional LC-MS and RePlay setup and base peak chromatograms of four consecutively acquired replay runs with 30 minute gradients. (a) In the conventional LC-MS setup the mass spectrometer sequences effectively only 50% of the time when using short gradients. (b) In the RePlay setup the second analysis can be carried out while the LC system is washing and loading the analytical column. In this way no extra time is needed. (c) BSA chromatograms of four replay runs measured in a sequence. Chromatographic performance of direct and RePlay were very similar.

Chromatographic performance

We tested different split ratios of the total flow (typically 200 nL/min) between direct and replay run from 1:4 to 3:1 and found that the chromatographic performance was excellent with splits in this range. For the experiments reported in this paper, we chose a split ratio of 1:1 or 3:1 to keep the volume of the capture capillary small and RePlay analysis time short. To test for potential detrimental effects of diffusion, we varied the inner diameter (ID) of the capillary between 20 and 40 µm and stored the gradient in the capture capillary for up to 30 minutes. Theoretically, using a flow rate of 100 nL/min, a peak of 20 s measured at the base would diffuse into 33 seconds in a 20 µm ID capillary, whereas the same peak would broaden to 45 s in a 30 µm ID tubing and to 57 s in a capillary of 40 µm ID. All data presented in this paper were acquired using a 30 µm ID capture capillary, and for this capillary no significant influence on chromatographic performance was observed. This indicates that the diffusion can still be reversed by the focusing column under these conditions. However, when we did not employ the focusing column we indeed observed appreciable peak broadening.

Figure 5.2a, b and Figure 5.3a illustrate one of the advantages of the RePlay system compared to the routine workflow in our laboratory. During the replay run, the analytical column is re-equilibrated and loaded with a new sample. Thus, the mass spectrometer is continuously sequencing peptides and the loading and washing steps do not subtract from the duty cycle of MS utilization. As shown in Figure 5.2c, four runs were analyzed in 8 x 30 minutes with essentially 100% duty cycle, whereas without RePlay, the duty cycle is about 50%. For 100 minute gradients we accomplished a duty cycle of 90% by directing one-quarter of the flow to the capture capillary and pushing the gradient out three times faster (see Figure 5.3). However, to ensure continued high chromatographic resolution we found it advantageous to include a few minutes' wash and equilibration of the focusing column in all protocols.

RePlay analysis of a complex proteomic samples

Next, we characterized the performance of the RePlay system for complex peptide mixtures typical of proteomics experiments. We loaded slightly less than 1 µg of mouse liver tryptic digest, corresponding to less than 20,000 hepatocyte cells, onto the RePlay system. Analysis was performed with a 100 minute gradient by a standard 'top 5' method on the LTQ-Orbitrap, as routinely used in our laboratory[16]. The base peak chromatogram in Figure 5.4 shows that the peptide pattern is largely indistinguishable between direct and replay run. Peaks eluted on average within 15 s in the direct and 16 s in the replay run, (with standard deviations of 5 s and 7 s, respectively; see Figure 5.4b). This is as good or better performance than our standard one-column setup for proteomics experiments. At a false discovery rate of one percent, 6,535 fully tryptic peptides were identified in the direct run, and 5,936 in the replay run, documenting the high reproducibility between the runs. A total of 8,383 unique peptides were identified in the combined analysis. Even though this experiment was meant to investigate reproducibility, the replay run nevertheless added 1,848 peptides (28 %) not identified in the direct run. On the protein level, the RePlay system identified 1,093 proteins with two unique peptides and at a false discovery rate of one percent, demonstrating that it is well suited to complex proteomic samples. This performance will likely further improve if the mass spectrometer is programmed to sequence only peptides in the replay run that have not been sequenced in the direct run.

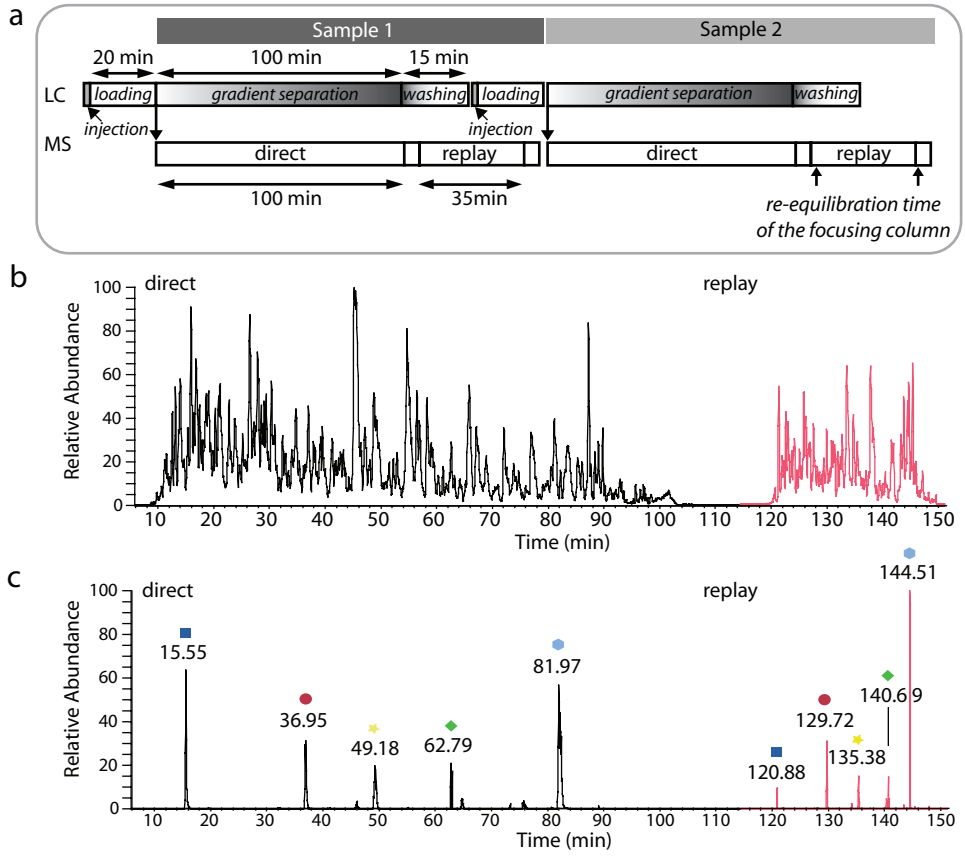


Figure 5.3 Timing schedule and base peak chromatogram of a setup using a 3:1 split and a fast replay run demonstrating excellent chromatographic performance in both runs without using additional time. (a) Timing schedule of two automated replay runs with the nanoLC and MS operating asynchronously. After sample loading and before starting the gradient the LC system starts the mass spectrometer. After finishing the gradient, the MS measures the replay run while the LC system loads the next sample. (b) Liver digests were run with a direct gradient of effectively 100 min and 33 minutes duration for the replay run. In nearly 90% of the measurement time, the mass spectrometer sequenced peptide ions. As shown in (c), the elution pattern and intensity of direct and replay run are very similar.

To demonstrate the ability of reanalyzing a complex mixture without adding to total analysis time, we split one quarter of the flow into the capture capillary (Figure 5.3). The total flow was 270 nL/min during direct analysis (200 nL/min to the focusing column) that lasted for 100 min. During the replay run the flow to the focusing column was also 200 nL/min, 'squeezing' the reanalysis into 35 min. Figure 5.3b and c show that this regime preserved peak intensities and slightly sharpened chromatographic peaks. Note that total analysis time was 150 min, very close to the 140 min standard cycle used in our laboratory – in which peptides elute for about 100 min.

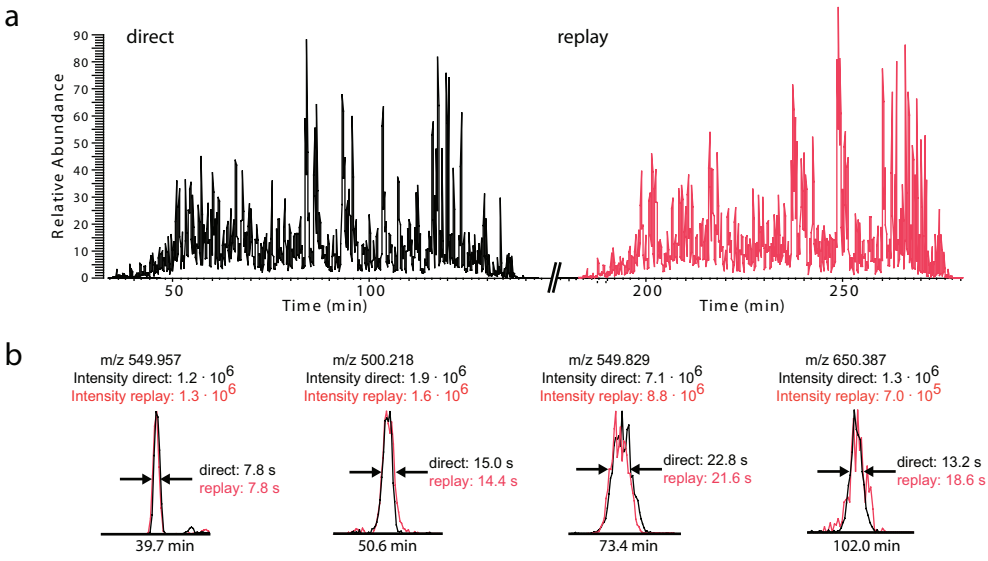


Figure 5.4 Chromatographic performance in complex proteome analysis with the RePlay system. (a) Typical base peak chromatogram of a RePlay analysis consisting of 100 minute direct (black) and 100 minute replay run (red) indicating the similarity of the chromatography performance. One µg of tryptic digest of liver homogenate was measured with the RePlay system using a 1:1 split (100 nL/min direct: 100 nL/min RePlay). (b) Selected ion chromatograms (0.01 m/z units wide) of four peptide peaks selected over the length of the gradient showing essentially identical signal intensities and peak widths at half height in direct and replay runs.

Targeting of SUMO-2 substrates in the replay run

Online reanalysis also allows targeting important peptides that were missed or unidentified in the direct analysis. We demonstrate this principle by targeted re-analysis of peptides that are extremely difficult to identify by tandem mass spectrometry. In our work investigating the conjugation of substrates with ubiquitin family proteins, we looked for specific proteins that are in-vivo sumoylated by SUMO-2[14]. Unlike ubiquitin, SUMO-2 leaves a large 32 or 34 amino acid peptide on the substrate peptide after tryptic digest, which makes it highly charged. The complex fragmentation spectra of the cross-linked peptides are notoriously difficult to identify with low resolution data[17]. For confident identification of the sumoylated peptide sequence, confirmation of high resolution and high accuracy data is required. To identify these conjugated SUMO-2 peptides in a streamlined fashion we purified HIS_6-SUMO-2 conjugates from HeLa lysate and measured them in the standard way in the direct run. Low resolution fragmentation spectra from the linear ion trap which indicated possible SUMO-2 conjugated peptides were then targeted in the replay run. In this run we only targeted these potential substrate peptides, fragmented them by Higher energy Collision Dissociation (HCD, ref. [15]) and analyzed them with high resolution and mass accuracy in the orbitrap instead of in the ion trap. Note that this analysis is slower and less sensitive and therefore could not have been performed efficiently in the direct run. Figure 5.5 shows the setup of the

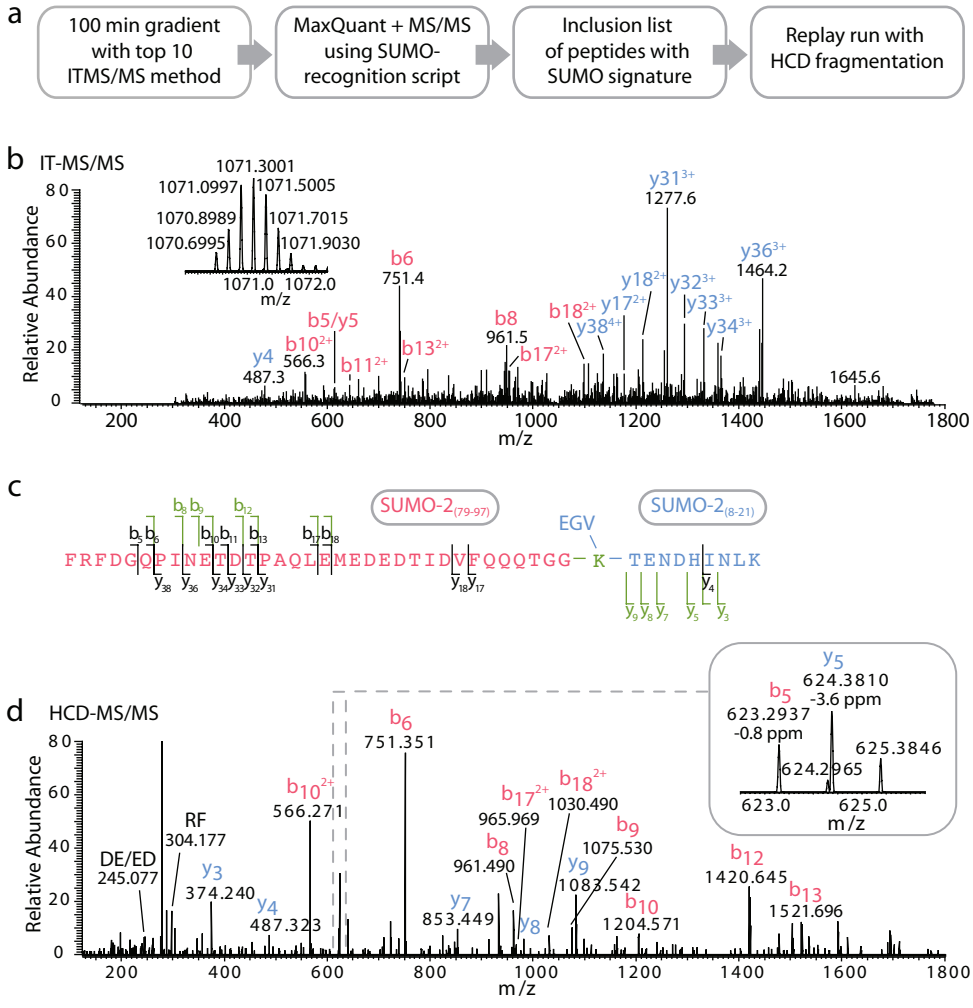


Figure 5.5 Targeted experiment to identify SUMO-2 substrates with high confidence. (a) After a 100 minute gradient with an MS method sequencing the 10 most abundant peaks in the ion trap by CID, a peak list of MS2 fragments was generated. This peak list was checked for peptides that contained multiple b-ions of the SUMO-2 C-terminal peptide and the corresponding peptide precursor masses were inserted in an inclusion list. These peptides were then specifically targeted by HCD and detected in the orbitrap in order to obtain high resolution and high mass accuracy data. (b) Annotated ion trap fragmentation spectrum of a SUMO-2 SUMO-2 conjugate, the orbitrap precursor of which is shown as an inset. Due to low resolution, charge states cannot be assigned and masses are inaccurate. (c) Sequence of SUMO-2 SUMO-2 peptide, indicating fragments observed in the ion trap (black and half-height) and in the orbitrap (green and at full height). (d) HCD-fragmentation spectrum of the same ion, with orbitrap resolution set to 7,500. Charge state can easily be assigned and the average absolute mass accuracy was in the low ppm range. Inset shows excellent quality of the fragmentation spectrum. Given the identification from spectrum (d), the quintuply-charged precursor ion in (b) matches within 1.35 ppm.

experiment as well as results for a SUMO-2 conjugated peptide. In the replay run, but not in the direct run, charge states and fragment masses are determined with very high accuracy (see inset of Figure 5.5d). Interestingly, the ion trap CID did not fragment the crucial substrate part of the cross-linked peptide (Figure 5.5b), therefore re-analysis with a different fragmentation method (HCD) was especially beneficial in this case (Figure 5.5d). The RePlay analysis produced an essentially complete run of y-ions of the substrate peptide.

This experiment was performed with a 3:1 split ratio where the RePlay analysis was kept entirely within the time usually needed for washing and loading, but that would suffice for characterizing a handful of interesting peptides. In this way, the mass spectrometer could – instead of standing idle – acquire targeted and very high quality data 'for free'.

Conclusions

In summary, we have demonstrated that the RePlay system enables very efficient use of the mass spectrometer, which is an important advantage given the high demand on sequencing time in large-scale proteomics laboratories and the costs of high resolution mass spectrometers. We also demonstrated a 'targeted analysis' in which the replay run enabled the collection of crucial, complementary data, thereby dramatically enhancing the information obtained in the direct run.

The great value of the RePlay system lies within the combination of several conceptual features that are otherwise only available individually; in separate set ups. Table 5.1 illustrates the features of RePlay compared to several other possible formats for LC-MS/MS experiments. Compared to a standard LC-MS setup, RePlay allows a second analysis using no or little extra analysis time. A duplicate injection in a standard set up would also allow investigating a sample to greater depth, but in that case it comes at the expense of more sample usage and doubled analysis time. Compared to a setup with two parallel columns and LC systems, RePlay achieves nearly the same time utilization of the mass spectrometer, without requiring an additional pump or LC system. Similarly, in the case of performing dual injections with dual pre-column setup one would require twice the sample amount and an extra (or

Table 5.1 Comparison of conceptual parameters for different LCMS setups.

	Conventional LC-MS/MS	Duplicate injection (conv. setup)	RePlay	Pre-column setup	Setup with 2 parallel columns
Sample utilization	+	-	++	+	+
Depth of analysis	+	++	++	+	+
MS utilization	-	-	++	+	++
LCMS analysis time	+	-	++	++	++
Setup simplicity	+	+	+/-	+/-	+/-
LC requirement	+	+	+	+	-

N.B. The conventional nanoflow LC-MS/MS setup is taken as a reference. The most favourable settings are highlighted in grey.

more complex) LC system. Not included in the table are 'fraction collection' and 'peak parking', which are impractical for high throughput proteomics for reasons described in the introduction of this chapter.

The only analytical cost of the RePlay system is a higher complexity of the setup. The split needs to be fine tuned whenever pressure changes after the split occur. However, we found the system to be robust, it did not lead to any loss of signal, and the chromatographic performance was better than or equivalent to our standard setup.

Most of the measurement time in standard LC-MS/MS is currently spent on obtaining information that later on turns out to be uninteresting. For example, one may be interested in only the peak pairs with unequal ratios in isotope based quantitative proteomics or one may only be interested in modified peptides. We envision that the RePlay system will be mainly used to analyze such interesting features in a focused way. This was demonstrated here with the example of the SUMO cross-linked peptides. However, dedicated acquisition software should enable a multitude of interesting features accessible to detailed proteomic analysis. For this purpose, the system will probably be configured to perform the replay analysis in a short time, during washing and loading of the analytical column.

However, beyond these applications, the system opens up a number of attractive possibilities, including ultrasensitive nanoscale proteomics. In this application, the flow would be split asymmetrically, allowing a very long analysis time in the replay run. Importantly, the LC system would still work at normal nanoflow rates (~200-500 nL/min), but the effective flow to the MS can be reduced a factor 5-10, without wasting any sample. This enables two analyses with ultrahigh sensitivity of a single injection and may be very beneficial for samples where the amount is limiting, for example in studies where cells or tissue samples are obtained by laser capture microdissection. In conclusion, we predict that the RePlay system will become a powerful and universal addition to the LC-MS tool chest.

Proteome data

Tables with lists of identified proteins and peptides are deposited at MCP online.

Acknowledgements

Ivan Matic and Alfred C. O. Vertegaal provided the SUMO-2 conjugates and help with SUMO-2 analysis and Yong Zhang provided software assistance for targeting SUMO-2 substrates. We thank Peter Bandilla for technical assistance.

References

1 Fenn, J. B., et al. (1989) Electrospray ionization for mass spectrometry of large biomolecules. Science 246, 64-71

2. Hernandez-Borges, J., et al. (2007) Recent applications in nanoliquid chromatography. 30, 1589-1610

3. Wilm, M., and Mann, M. (1996) Analytical properties of the nanoelectrospray ion source. Anal Chem 68, 1-8

4. Karas, M., and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem 60, 2299-2301

5. Kuster, B., Schirle, M., Mallick, P., and Aebersold, R. (2005) Scoring proteomes with proteotypic peptide probes. Nat Rev Mol Cell Biol 6, 577-583

6. Davis, M. T., et al. (1995) A microscale electrospray interface for on-line, capillary liquid chromatography/tandem mass spectrometry of complex peptide mixtures. Anal Chem 67, 4549-4556

7. Staack, R. F., Varesio, E., and Hopfgartner, G. (2005) The combination of LC–MS/MS and chip-based infusion for improved screening and characterization of drug metabolites. Rapid Commun Mass Spectrom 19, 618–626
8 Corso, T. N., et al. (2006) Ultralow-volume fraction collection from NanoLC columns for mass spectrometric analysis of protein phosphorylation and glycosylation. Anal Chem 78, 2209–2219
9 Shi, R., et al. (2007) Analysis of the mouse liver proteome using advanced mass spectrometry. J Proteome Res 6, 2963–2972
10 Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem 75, 663–670
11 Graumann, J., et al. (2007) SILAC-labeling and proteome quantitation of mouse embryonic stem cells to a depth of 5111 proteins. Mol Cell Proteomics, 7(4), 672–683
12 Olsen, J. V., Ong, S. E., and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol Cell Proteomics 3, 608–614
13 Zubarev, R., and Mann, M. (2007) On the proper use of mass accuracy in proteomics. Mol Cell Proteomics 6, 377–381
14 Matic, I., et al. (2008) In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. Mol Cell Proteomics 7, 132–144
15 Olsen, J. V., et al. (2007) Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods 4, 709–712
16 Olsen, J. V., et al. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics 4, 2010–2021
17 Pedrioli, P. G., et al. (2006) Automated identification of SUMOylation sites using mass spectrometry and SUMmOn pattern recognition software. Nat Methods 3, 533–539

6

Quantitative proteomic analysis of single pancreatic islets

Leonie F. Waanders, Karolina Chwalek, Mara Monetti, Chanchal Kumar,
Eckhard Lammert and Matthias Mann

Submitted manuscript

Technological developments make mass spectrometry (MS)-based proteomics a central pillar of biochemical research. MS has been very successful in cell culture systems, where sample amounts are not limiting. To extend its capabilities to extremely small, physiologically distinct cell types isolated from tissue, we developed a high sensitivity chromatographic system that measures nanogram protein mixtures for eigth hours with very high resolution. This technology is based on splitting gradient effluents into a capture capillary and provides an inherent technical replicate.

In a single analysis this allowed us to characterize kidney glomeruli isolated by laser capture microdissection to a depth of more than 2400 proteins. From pooled pancreatic islets of Langerhans, another type of 'mini-organ', we obtained an in-depth proteome of 6873 proteins, many of them involved in diabetes. We quantitatively compared the proteome of single islets, containing 2000-4000 cells, treated with high or low glucose levels, covering most of the characteristic functions of beta-cells.

Our ultrasensitive analysis recapitulated known hyperglycemic changes but we also find novel components upregulated such as the mitochondrial stress regulator Park7. Direct proteomic analysis of functionally distinct cellular structures opens up new perspectives in physiology and pathology.

Introduction

Mass spectrometry is inherently an extremely sensitive analysis technique: single proteins have been analyzed at the attomole level many years ago[1,2]. Similarly, it is now possible to identify thousands of proteins in single experiments and a first comprehensive proteome identification and quantitation has been reported[3]. However, the complexity of the proteomic mixtures typically necessitates micrograms or milligrams of starting material – protein amounts that can readily be extracted from cell or tissue homogenates[4,5]. The analysis of small tissue substructures with specialized functions is medically important but has remained much more difficult. Ideally these structures are separated from the surrounding tissue to enrich proteins specific to these areas. Although such selection is possible by laser-capture microdissection (LCM), successful use of this technology has so far overwhelmingly been in the areas of RNA and DNA assays. In the few cases where LCM was combined with MS analysis, many thousands cells were collected (taking presumably several hours or days of selection), to obtain the protein amounts typical for MS[6]. Only a few investigators have attempted to measure smaller cell numbers. Luider et al. reported 1000 identified proteins from 3000 cells and a similar number from LCM-selected breast cancer tissues, but they did not sequence peptides by mass spectrometry, relying instead on mass measurements only[7,8].

Main obstacles in analyzing small protein amounts are the considerable sample losses

during preparation and the limited 'dynamic range' of liquid chromatography – tandem mass spectrometry (LC–MS/MS) setups used in proteomics. Sample preparation efficiency has improved by the advent of polished sample tubes (e.g. Eppendorf Protein LoBind) and MS-friendly detergents[9,10], thereby allowing shorter extraction and digestion protocols and reduced unspecific surface binding.

The second important obstacle, the dynamic range, refers to the number of different ion species that can be observed simultaneously. With highly complex samples, as are typical for proteomic experiments, low intensity ions are likely to be masked by more abundant species. In order to also detect these low abundance proteins, extensive fractionation is commonly applied. But this is not possible with very low starting amounts as sample losses are too high.

Recently, we described the 'RePlay chromatography system' to allow direct reanalysis of every sample without losing signal intensity[11]. In RePlay, post-column effluent is split with one part directed to online LC–MS/MS analysis and the other one stored for another round of analysis after the first one is finished. We envisioned that this setup should also be extremely valuable for analysis of probes where starting amount is limiting, and therefore optimized the system for high sensitivity 'single shot' analysis. We used optimized sample preparation methods and the RePlay system in combination with the sensitive LTQ-Orbitrap mass spectrometer. Due to the split flow of the RePlay system flow rates could be reduced five-fold compared to normal operation while maintaining normal backpressures and a continuous stable flow and spray. In addition we doubled the gradient time to increase the sequencing events per sample and applied the unique RePlay ability to re-measure the sample without losing signal intensity. Thus we inherently obtain a technical replicate, increasing quantitation accuracy, even from the smallest proteomic samples.

Here we demonstrate that our novel high-sensitivity RePlay workflow enables the analysis of small 'mini-organs': we examined mouse glomeruli selected by LCM and single islets of Langerhans, handpicked after purification. In both cases the starting material was less than 400 ng of proteins – and still resulted in more than 2000 protein identifications. Besides high confident identifications, we demonstrate quantitation of protein expression levels in single islets of Langerhans. These islets were stimulated in high or low glucose prior to LC–MS/MS and we identified many known and some novel responders, like antioxidants, proteins involved in glucose catabolism or vesicle secretion processes.

Materials and methods

LCM workflow

Cryosections of mouse kidney were cut (12–16 µm thick), placed on glass sections and stained with methylene blue. Per sample, about 50 glomeruli were selected (tissue areas of approx. 500,000 μm^2) and dissected by a MicroBeam laser (Carl Zeiss MicroImaging).

Pooled islets & total pancreas homogenate

Pancreatic islets were isolated from 12–16 weeks old C57Bl6/J male mice. Islet preparation was done as described in ref. [12]. Prior to removal of the pancreas, Liberase I solution was injected into the pancreatic duct (~2 ml/mouse). The pancreas was incubated at 37°C for 23 min after which the digestion was stopped by adding media supplemented with 10% foetal bovine serum, followed by 3 min centrifugation at 1000 g. Low glucose DMEM was added to the pellet, mixed vigorously, and filtered though a 400 µm diameter wire mesh. Remaining islets were

washed once. The islets were spun down and resuspended in histopaque 1077 (Sigma). Low glucose DMEM was added carefully, prior to centrifugation for 50 min at 1100 g at 10°C without acceleration or braking. The islet layer was collected and washed twice with DMEM to remove histopague. Total pancreas tissue was obtained by dissecting directly after pancreas excision.

Single Islet preparation

Pancreatic islets were isolated from C57Bl6/J male mice, 8-10 weeks old as described above. Purified islets were recovered overnight in 11 mM glucose CMRL medium containing 15% heat inactivated foetal bovine serum, 100 units/mL penicillin, 100 µg/mL streptomycin, 1 mM sodium bicarbonate, 0.15% sodium bicarbonate and 0.001%beta-mercaptoethanol.Subsequently, the cells were incubated for 24h with either 16.7 mM (high) or 5.6 mM (low) glucose in Krebs-Ringer-Hepes buffer (25mM HEPES pH7.4, 115 mM NaCl, 24 mM $NaHCO_3$, 5 mM KCl, 2.5 mM $CaCl_2$, 1 mM $MgCl_2$, 0.1% BSA). Single islets were handpicked.

Sample digestion

LCM and single islet samples were dissolved in 0.1 % PPS Silent Surfactant (Protein Discovery) in 50 mM ammonium bicarbonate with 1 mM dithiothreitol, heated to 95°C for 3 min and sonicated for 2 min. Subsequently proteins were alkylated in 1 mM chloroace-tamide in ammonium bicarbonate for 30 min and digested overnight with porcine trypsin (Promega) in an enzyme-to-protein ratio of 1:50. After acidification by trifluoroacetic acid, peptide mixtures were concentrated and desalted on StageTips[13] and eluted directly before LC-MS measurement. To prevent protein losses during sample preparation LoBind Eppendorf Tubes and Axygen Maxymum Recovery tips were used and sample contact was minimized. For pooled islet and pancreas proteome the peptide samples were additionally separated into 12 fractions by peptide isoelectric focusing as described[14].

RePlay Setup

The RePlay system was based on the one described in[9], but developed for high sensitivity: the capture capillary was 19 m x 30 µm inner diameter (ID); the focusing column, a pulled capillary of 22 cm x 40 um ID was packed with 3 µm ReproSil-Pur C18-AQ resin (Dr. Maisch GmbH)[15]. For the LCM and single islet measurements we used a split ratio of 1:1 (MS : capture capillary), an effective flow rate of 50 nL to the MS in both direct and replay run, and 240 min gradients. For the pooled islet and pancreas proteome gradients of 2h were used with 150 nL/min effective flow rates. The flow sensor was placed in the waste line to prevent potential dilution effects.

The gradient generated by a Proxeon easyLC (Proxeon) was coupled, via the RePlay system and an electrospray interface, to an LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific). MS and MS/MS were acquired data-dependently of the ten most intense ions per survey spectrum, using MS/MS target values of 5000, with a minimal intensity of 1000 ions and a maximum fill time of 150 ms. For survey spectra one million ions were accumulated with maximum fill times of 500 ms.

Data analysis

MS/MS data were analyzed by in-house developed MaxQuant software (version 1.0.12.27), essentially as described[10]. The data was searched using Mascot (version 2.1.04, Matrix Science Ltd) against the mouse IPI database (version 3.37) supplemented with frequently observed contaminants and concatenated with reversed copies of all sequences (2 x 51,467 entries). Enzyme

specificity was set to trypsin, allowing for cleavage N-terminal to proline and between aspartic acid and proline and a maximum of two missed cleavages. Carbamidomethylcysteine was set as fixed and N-acetylation and methionine oxidation were set as variable modifications. The initial maximum allowed mass deviation was set to 7 ppm for monoisotopic precursor ions and 0.5 Da for MS/MS peaks. The required minimum peptide length was six amino acids. If the identified peptide sequence set of one protein was equal to or contained another protein's peptide set, these two proteins were grouped together and the proteins were not counted as independent hits. Proteins were considered identified when at least two peptides were identified of which at least one uniquely assignable to the respective sequence. The false discovery rate (FDR) at the peptide level and protein level were set to 1% for the glomeruli, pooled islet and pancreas proteome. Based on the additional identification confidence provided by the pooled islet dataset, we accepted single unique peptide identifications with 5% FDR on peptide and 2.5% on protein level. For quantitation proteins were additionally filtered, and were required to be detected in 4 out of 8 most intense LC-MS/MS runs with at least 2 unique peptides. Parametric ANOVA analysis was performed with $p<0.01$ significance cut-off.

Results and discussion

There are two challenges in the measurement of very small in vivo cell populations by liquid chromatography – tandem mass spectrometry (LC-MS/MS): efficient recovery of the proteins and high analytical sensitivity. To meet the first challenge, we implemented a workflow with a minimal number of handling steps starting from LCM (Figure 6.1a) or mini-organ purification (Figure 6.1b) and resulting in a trypsin digested proteome. LC-MS/MS sensitivity was improved in three ways (Figure 6.1c): (1) By reducing the effective flow rate to 50 nL/min and employing smaller diameter columns; (2) By separating the peptide mixture with very long gradients (4h) to increase peptide sequencing events by the high-resolution LTQ-Orbitrap mass spectrometer; (3) By splitting the flow after chromatographic peptide separation, thereby providing a technical replicate from a single sample loading without loss of signal (RePlay[11]).

Glomeruli

First, we tested this high-sensitivity workflow on mouse glomeruli, about 50 of which we isolated by LCM from histological slices. Glomeruli are the filtering units in the kidney and consist of only a few hundred cells in mice. Automated analysis proved robust and the chromatographic setup retained excellent average peak width (<22s) in direct and replay runs (Figure 6.1d). Overall signal was the same in both runs and individual peptide intensities, which ranged over more than four orders of magnitude, correlated highly between them ($R=0.92$; Supplementary Figure 6.1a). Data processing with MaxQuant[16] resulted in the identification of 2406 distinct proteins from a single injection. Virtually all known glomeruli proteins are contained in this data set. Repeat measurements resulted in 2402 and 2377 proteins identified and the three data sets shared 86% of protein identifications (Supplementary Figure 6.1b). For comparison, our recent proteomic analysis of 20 fractions of mouse liver tissues using thousand-fold more material resulted in a similar number of identifications[17].

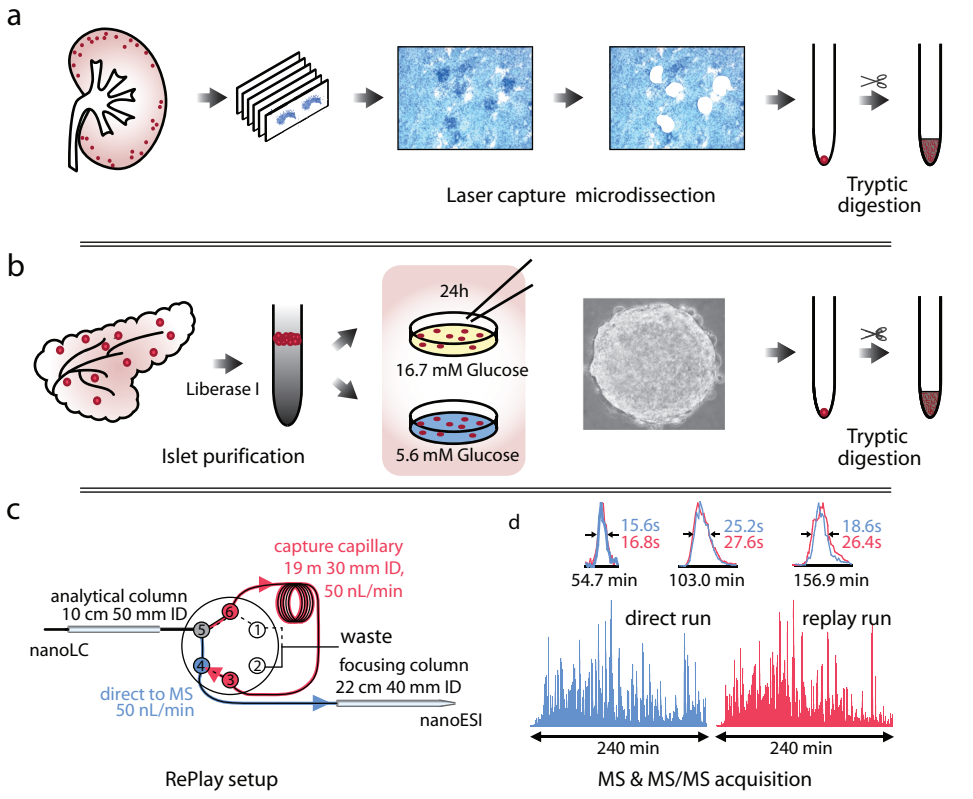


Figure 6.1 Schematic workflow for high-sensitive LC-MS analysis with direct replicate measurement. (a) Workflow for laser capture microdissection of mouse kidney glomeruli followed by protein extraction and digestion. (b) Purification of islets of Langerhans from mouse pancreas by Liberase I treatment and gradient centrifugation. Islets were stimulated with high or low glucose, individually collected, and proteins solubilized and digested by trypsin. (c) High sensitivity nanoLC-MS setup with the RePlay to obtain two measurements per injection. (d) Total ion chromatogram in direct and replay analysis are very similar. Insets show identical resolution and intensity for three typical peptide peaks.

In-depth proteome of pancreatic islets

Islets of Langerhans are another type of mini-organ essential for human health. In mice, 2000 islets are dispersed throughout the pancreas of which the insulin producing β-cell is the most predominant[18]. To obtain an in-depth islet proteome we pooled purified islets and analyzed them after peptide fractionation. Triplicate analysis employing RePlay and material from six mice resulted in the identification of 7014 proteins with at least two peptides and a FDR of 1%. The largest previous islet proteome is less than half the size and almost completely contained in our proteome (94%)[19]. This list consists of (i) islet specific proteins, (ii) proteins shared between islets and the rest of the pancreas, and (iii) proteins from the rest of the pancreas that contaminate the islet preparation. To distinguish these classes we measured

an in-depth proteome of complete pancreas homogenate, of which the islets constitute less than 2%, and quantified proteins against the islet proteome by label-free quantification in MaxQuant[20]. Elastase, trypsin, and other digestive enzymes produced by the exocrine pancreas were more than 30-fold more abundant in the total pancreas proteome compared to purified islets. After filtering out proteins with a similar behavior, we retain a high quality islet proteome of 6873 proteins. Our data indicate that islets, which are highly specialized tissue structures, use of at least a third of the genes in the genome. Moreover, several proteins were identified as the pancreas specific isoform of the gene and our data provides isoform specific tissue expression information for many more.

The known endocrine hormones produced by this mini-organ, such as insulin 1 and 2, glucagon and secretogranins are among the most highly expressed proteins in the proteome judged by added peptide intensity[3]. To contrast islet specific expression to expression in whole pancreas we categorized islet proteins by their enrichment factors (Figure 6.2a). More than half of the islet proteome is within a four-fold abundance range compared to whole pancreas, and these proteins are predominantly associated with house-hold functions. The pancreatic islet is an intricately assembled structure connecting the endocrine cells to the circulation and neural control. Concordantly, proteins connected to cell communication and cell adhesion make up much of the specific islet proteome. In the 1133 proteins enriched more than four-fold in islets, membrane and extracellular proteins comprise a large part (Figure 6.2b). This class of proteins contains various G-protein coupled receptors, cAMP regulators, receptor tyrosine kinases and ion channels - potential and established targets for diabetes drugs, as well as autoantigens in type I diabetes (Figure 6.2b-c, Supplementary Figure 6.1c). Interestingly, 86 of 396 identified kinases are highly enriched in the islets. Our dataset also contains many unannotated proteins, especially among those that were close to the limit of detection and therefore not quantified and categorized (Figure 6.2a).

Single islet analysis

High fat and high glucose conditions are reported to negatively affect β-cell function, a process that is central to the development of type 2 diabetes[21]. Dynamics of β-cells are frequently studied in cell culture; however, it is physiologically more relevant to investigate these cells in their native cellular context[22]. We wondered if our high sensitivity workflow was capable of analyzing single pancreatic islets. To this end, we determined single, isolated islet proteomes after 24h at 16.7 mM and 5.6 mM glucose, representing hyperglycemic and basal conditions, respectively. One day of elevated glucose concentration significantly affects metabolic functions at the mRNA level[23], but does not yet evoke irreversible changes[24].

First we assessed the proteome coverage in single islet experiments. Similar to the case of glomeruli we achieved very good correlation on the peptide level between direct and replay runs (Figure 6.3a). Protein quantitation integrates several peptide measurements and therefore the quantitative reproducibility is even higher on the protein level (R=0.98; Figure 6.3b). We sequenced 15,193 different peptides originating from 2013 unique proteins (2.5% FDR). Per biological replicate we identified on average 84% of the total number of proteins. 97% of the single islet proteins (1949) were also present in the large islet proteome determined above (Supplementary Figure 6.1d). The single islet proteome contained all the islet specific hormones as well as known

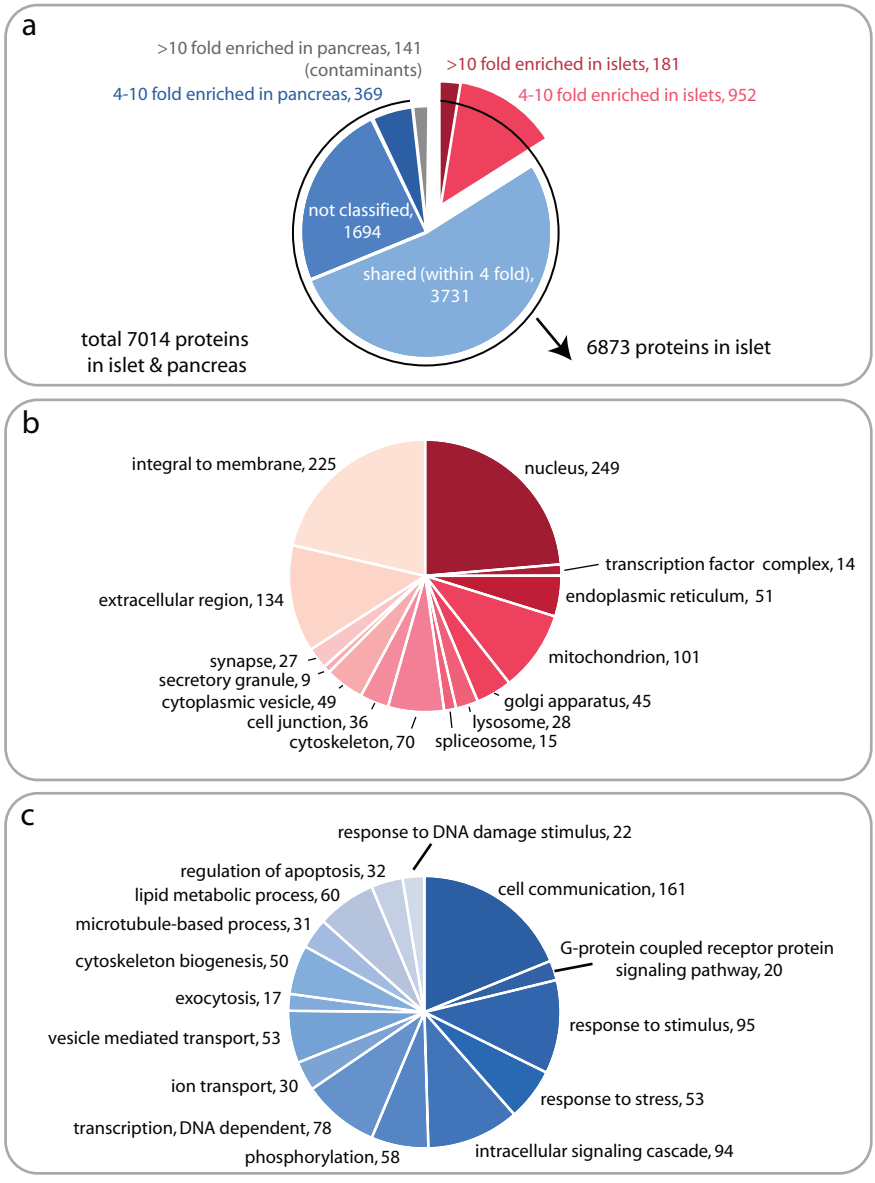


Figure 6.2 The proteome of pooled islets of Langerhans. (a) Classification of purified islet proteome versus pancreatic proteome, based on protein ratios between the two samples. After subtracting pancreas contaminants (141), a high confidence list of 6873 islet proteins was obtained. Note that for proteins with less than 3 peptides and close to the limit of detection, accurate ratios and hence classification were not assigned ('not classified'). Proteins enriched more than four-fold (total 1133) were categorized by (b) GO cellular compartmentalization and (c) GO biological process. Numbers indicate proteins per category in the enriched dataset.

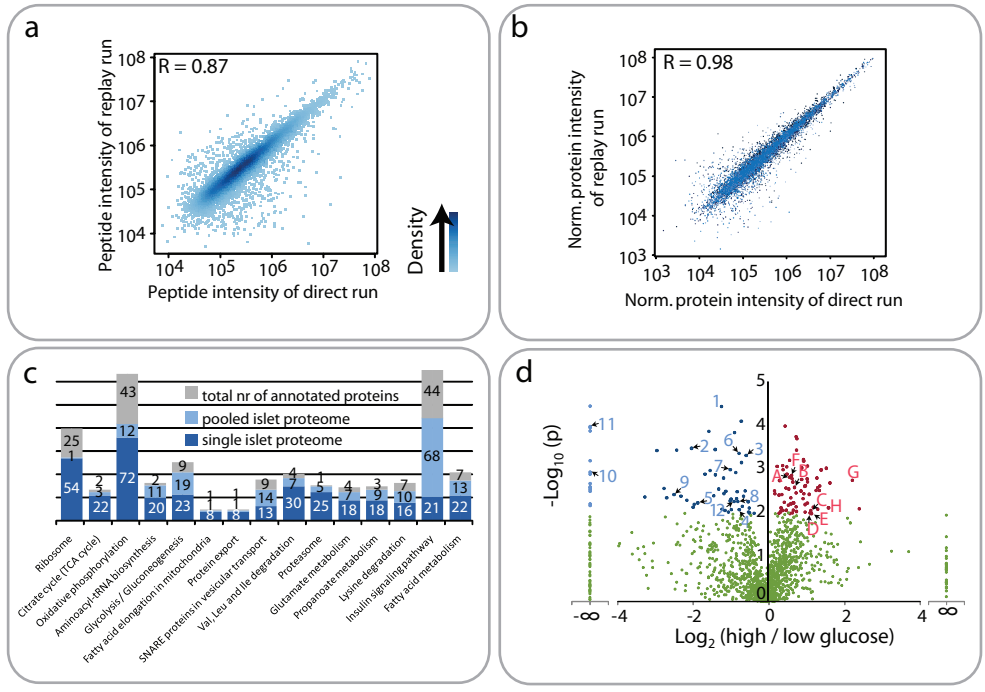


Figure 6.3 Quantitative analysis of single islets by MS-based proteomics. (a) Correlation of peptide intensities between direct and replay runs of one single islet measurement. (b) Comparison of six direct and replay runs indicates very high quantitative reproducibility at the protein level. (c) Annotation of single islet and pooled-islet proteome to KEGG pathways important for β-cell function. Numbers indicate identified proteins within annotated group. The single islet proteome covers most essential β-cell processes like glycolysis, TCA cycle and oxidative phosphorylation. (d) Volcano plot showing p-values (-log10) versus protein ratio of high / low glucose treated islets (log2) of all 1482 proteins fulfilling strict quantitation criteria. (Red, 77 upregulated proteins; blue, 65 downregulated proteins; green, not significantly changed upon 24h glucose stimulation; ANOVA with p<0.01). Gene names of proteins discussed in the text: A, ANT1; B, Ero1lα; C, Prdx3; D, Park7; E, Gsn; F, Vil1; G, Actn1; H, Itgb1; 1, Pcsk1n; 2, Pdyn; 3, ATP2a2; 4, ATP2a3; 5, Ins2; 6, Sytl4; 7, Vamp2; 8, Vamp8; 9, Scamp3; 10, Rab3b; 11, Capza1; 12, Capza2.

β-cell proteins like glucose transporter 2, glucokinase, and prohormone convertase 1 and 2. Furthermore, we covered the majority of proteins involved in classical β-cell functions (Figure 6.3c, Supplementary Figure 6.1e-g); for example, between 40 and 85% of the annotated proteins (irrespective of tissue) in glycolysis, TCA cycle and oxidative phosphorylation, ribosomes, and secretory granules (Fig 4).

Protein expression changes upon hyperglycemia

Next we determined changes of the proteome after 24h glucose stimulation. Quantitative comparison was based on proteins identified with ≥ 2 unique peptides and detected in ≥ 4 out of 8 runs (1487 quantifiable proteins). With these stringent criteria and ANOVA p<0.01, there were 77 up- and 65 downregulated proteins. Some of

the most significantly affected proteins are highlighted in the volcano plot in Figure 6.3d. Many regulated proteins have previously been associated with hyperglycemia in animal models for diabetes or in patients (Supplementary Table 4). Concordant with a recently published microarray study[23] we find a general upregulation of glycolysis, the TCA cycle and ATP translocation (Figure 6.4). In contrast to elevated insulin transcription, we observed reduced insulin protein levels, presumably due to depletion of insulin from internal stores as a consequence of continuous secretion. This assumption is indirectly supported by the finding of decreased levels of two hormone processing inhibitors prodynorphin (Pdyn) and ProSAAS (Pcsk1n)[25].

Increased glucose catabolism induces oxidative stress, a process detrimental to β-cells, and we find molecular evidence for this, in the form of antioxidant proteins, supporting and expanding previous observations (e.g. ERO1l-α and Prdx3, Figure 6.3 and 6.4)[26]. Prdx3, upregulated 2.4 fold, protects β-cells from mitochondrial hydrogen peroxide stress and mice overexpressing Prdx3 are more glucose tolerant upon high-fat diet feeding and better protected against hyperglycemia[27]. Interestingly, DJ-1 (Park7) and Mn^{2+}-superoxide dismutase (Sod2) were upregulated 2 and 1.7 fold ($p<0.05$), respectively. DJ-1 is a mitochondrial oxidative stress and apoptotic regulator implicated in Parkinson's disease[28,29] but has not yet been associated with β-cell responses in hyperglycemia. It induces mitochondrial antioxidant proteins like Sod2[30]. To explore DJ-1's connection to β-cell dysfunction, we quantified its expression in mouse islets after high-fat diet and found the protein upregulated 1.7 fold (Supplementary Figure 6.1h) – similar to the short term hyperglycemic stimulation – suggesting that it is involved in β-cell regulation under multiple stress conditions.

In contrast to the increased protein levels in glucose catabolism and induced stress mechanisms described above, proteins associated with vesicle secretion are reduced (Figure 6.4). In the case of vesicle-associated membrane protein (VAMP)-2, synaptotagmin-like protein 4 (Sytl4) and Rab3b, who interact upon Ca2+ increase, this has been observed before[31,32]. Decreased expression of Sytl4 reduces the number of docked vesicles to the plasma membrane, which is suggested to effect insulin secretion[24,33]. Decreased mRNA and protein expression of VAMP2 and various other secretory proteins have also been associated with impaired protein secretion in diabetic patients[34] and animal models for diabetes[35]. Furthermore, we observed depletion of all three members of the sarcoplasmic/endoplasmic reticulum calcium ATPase (SERCA) complex, which has also been correlated with insulin secretion deficiencies in diabetic animal models[36]. In contrast, mRNA levels of SERCA increase after 18h of glucose stimulation[23], possibly pointing to a compensatory mechanism for acute protein depletion.

Vesicle access to the plasma membrane is thought to be hindered by dense cortical F-actin and depolymerization of F-actin upon glucose stimulation potentiates insulin secretion[37]. Supporting this view, we observed altered levels of proteins involved in remodeling the cortical cytoskeleton network – the depolymerizing enzymes Villin 1 and Gelsolin are up and inhibitory enzymes Capza 1 and 2 are downregulated. Gelsolin's role in this process has already been demonstrated[38]. Moreover, α-actinin-1 and 4 were upregulated. Actinin-1 associates with cell adhesion molecules like integrin-β1, which was increased three fold and plays an important role in β-cell extracellular matrix interaction[39]. β1-integrin is an essential factor for β-cell proliferation

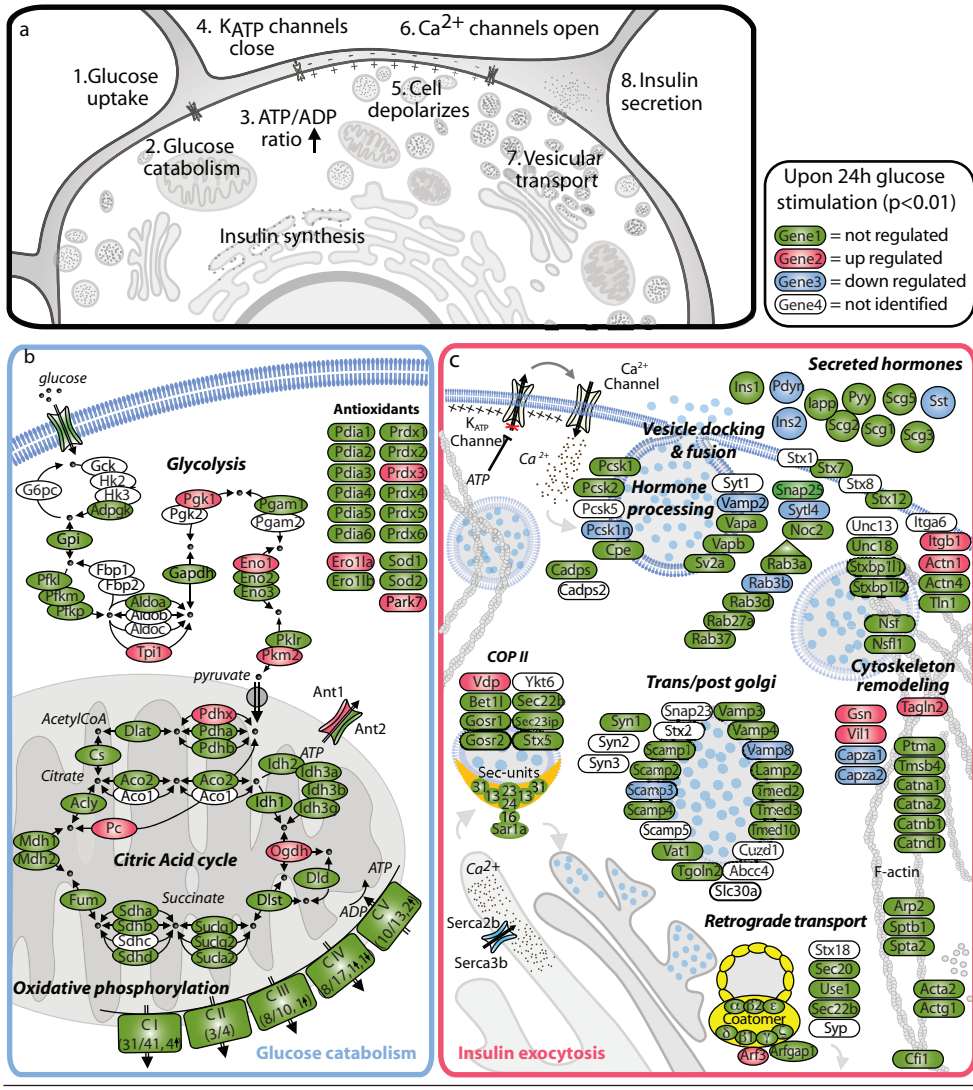


Figure 6.4 Main differentially expressed proteins upon 24h glucose stimulation in the context of physiological processes related to glucose stimulated insulin secretion. (a) Cellular events leading to insulin secretion after normal glucose stimulation. Main affected processes are presented in further detail, showing proteins (gene name) with the following color-coding: green = identified, red = upregulated, blue = downregulated and white = not identified. (b) Glucose uptake, glycolysis, TCA cycle, oxidative phosphorylation and antioxidants. (c) Vesicle transport and cytoskeleton remodeling.

and communicates with laminins in the basement membrane[12]. Consistently, there is a direct relation between its expression level and the rate of insulin secretion upon glucose stimulation[40]. Thus, our single islet analysis directly supports an important role of cytoskeletal remodeling in high glucose conditions.

Conclusions

In conclusion, by optimizing the proteomic workflow, using reduced flow rates, direct replicate measurements and extremely accurate and highly sensitive mass spectrometry, in vivo structures comprising a few thousand cells can now be quantitatively analyzed. Our study highlights a beneficial aspect of this directed analysis: The functionally relevant proteome turns out to be expressed relatively highly in these specialized cells and is – by avoiding the dilution effect of whole tissue analysis – more readily accessible to proteomic measurements. Already, the method described here could be used to evaluate single islets developed from in vitro differentiated stem cells or disease control of a few islets taken from human biopsies. The flow rate and column diameter can be significantly lowered[41] and MS sensitivity and scan speed increased[42]. Thus a few hundred cells are analyzable or alternatively depth of coverage can be increased significantly. This allows in depth proteomic analysis of virtually all features visible by standard histological methods.

Proteome data

Raw data as well as supplementary tables containing identified proteins and peptides are deposited at Tranche (proteomecommons.org) and are available on request.

Acknowledgements

Tami Geiger, Ana Velic and Gabriele Friedemann assisted with glomeruli collection and data analysis. Irena Constantinova provided the single islet picture in Figure 6.1. We thank Tobias Walther, Maxi Hilger, Nina Hubner, and Tami Geiger for critical reading of the manuscript. This project was partially supported by PROSPECTS, a 7th framework EU grant. K. C. and E. L. were supported by the German Research Council (DFG LA1216)..

References

1 Valaskovic G. A., Kelleher N. L. & McLafferty F. W. (1996) Attomole protein characterization by capillary electrophoresis-mass spectrometry. Science 273(5279), 1199-202

2 Aebersold R. & Mann M. (2003) Mass spectrometry-based proteomics. Nature 422, 198-207

3 de Godoy L. M., et al. (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 455, 1251-1254

4 Yates J. R., et al. (2005) Proteomics of organelles and large cellular structures. Nat Rev Mol Cell Biol 6, 702-714

5 Mann M. & Kelleher N. L. (2008) Precision proteomics: The case for high resolution and high mass accuracy. Proc Natl Acad Sci U S A 105, 18132-18138

6 Espina V., Heiby M., Pierobon M., & Liotta L. A. (2007) Laser capture microdissection technology Exp rev mol diagn 7, 647-657

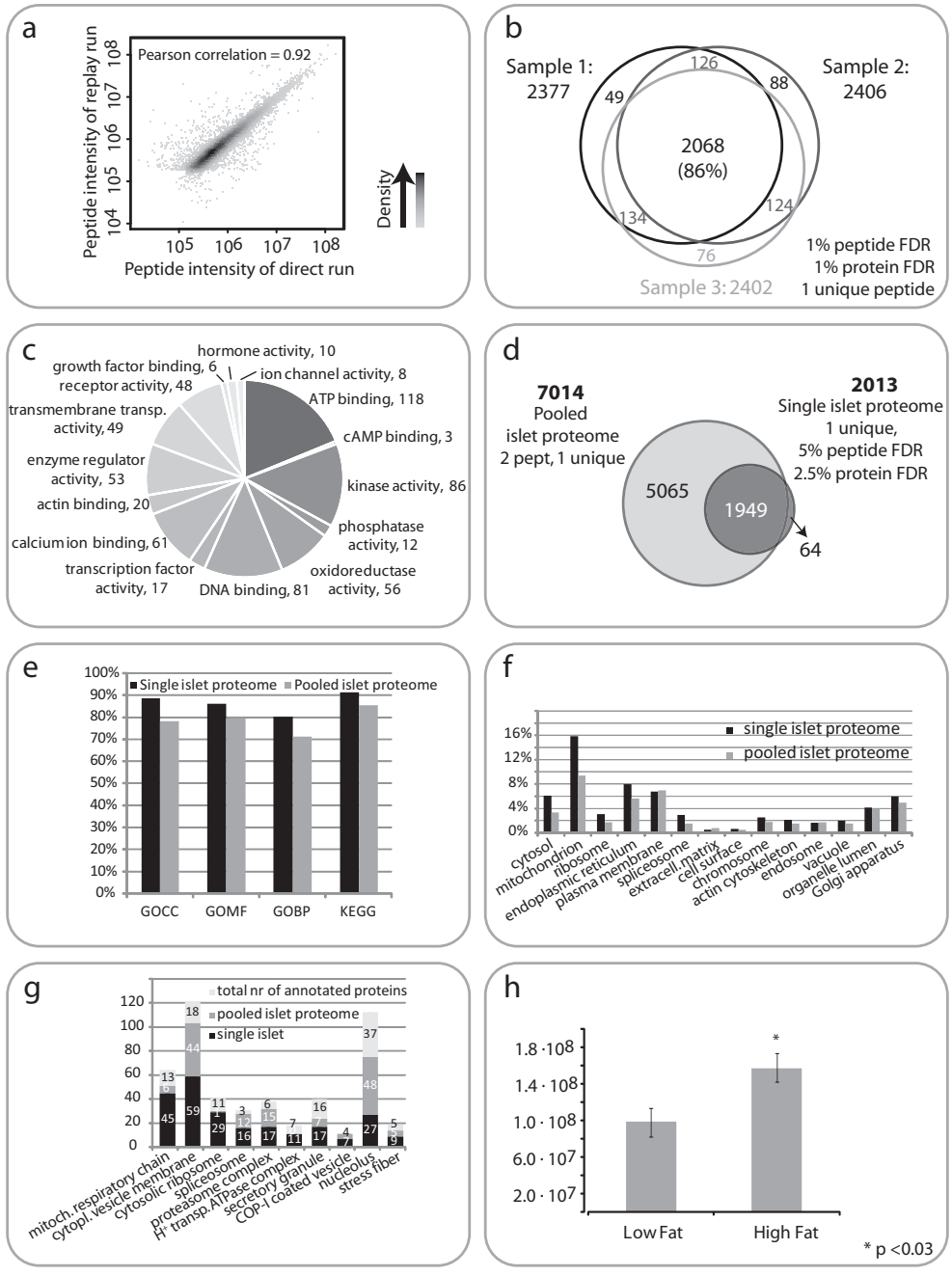
7 Umar A., et al. (2007) NanoLC-FT-ICR MS improves proteome coverage attainable for ~3000 laser microdissected breast carcinoma cells. Proteomics 7, 323-329

8 Umar A., et al. (2009) Identification of a Putative Protein Profile Associated with Tamoxifen Therapy Resistance in Breast Cancer. Mol Cell Proteomics 8, 1278-1294

9 Norris J. L., Porter N. A., & Caprioli R. M. (2003) Methods for direct biomolecule identification by matrix-assisted laser desorption ionization (maldi) mass spectrometry. Anal Chem 75, 6642-6647

10 Yu Y. Q., et al. (2005) A rapid sample preparation method for mass spectrometric characterization of N-linked glycans. Rapid Commun Mass Spectrom 19, 2331-2336

11 Waanders L. F., et al. (2008) A Novel Chro-



Supplementary Figure 6.1 (a) Peptide intensity correlation between direct and replay runs (technical replicates) of a glomeruli sample collected by LCM. (b) Venn diagram of protein identified in three individually collected glomeruli samples (biological replicates) (c) Annotation of pooled islet dataset by GO–molecular function. (d) Venn diagram of pooled islet proteome versus proteins identified in single

matographic Method Allows On-line Reanalysis of the Proteome. Mol Cell Proteomics 7, 1452-1459
12 Nikolova G., et al. (2006) The vascular basement membrane: a niche for insulin gene expression and Beta cell proliferation. Dev Cell 10, 397-405
13 Rappsilber J., Ishihama Y., & Mann M. (2003) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Analytical chemistry 75, 663-670
14 Graumann J., et al. (2008) Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. Mol Cell Proteomics 7, 672-683
15 Ishihama Y., et al. (2002) Microcolumns with self-assembled particle frits for proteomics. J Chromatogr A 979, 233-239
16 Cox J. & Mann M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 26, 1367-1372
17 Shi R., et al. (2007) Analysis of the mouse liver proteome using advanced mass spectrometry. J Proteome Res 6, 2963-2972
18 Kulkarni R. N. (2004) The islet β-cell. Int J Biochem Cell Biol 36, 365-371
19 Petyuk V. A., et al. (2008) Characterization of the Mouse Pancreatic Islet Proteome and Comparative Analysis with Other Mouse Tissues. J Proteome Res 7, 3114-3126
20 Cox J., Luber C. A., Nagaraj N., & Mann M. (2009) submitted and available upon request
21 Kahn SE (2000) The importance of the beta-cell in the pathogenesis of type 2 diabetes mellitus. Am J Med 108 Suppl 6a, 2S-8S
22 Kinard T. A., et al. (1999) Modulation of the Bursting Properties of Single Mouse Pancreatic -Cells by Artificial Conductances. Biophysical Journal 76, 1423-1435
23 Bensellam M., et al. (2009) Cluster analysis of rat pancreatic islet gene mRNA levels after culture in low-, intermediate-and high-glucose concentrations. Diabetologia 52, 463-476
24 Tsuboi T., et al. (2006) Sustained Exposure to High Glucose Concentrations Modifies Glucose Signaling and the Mechanics of Secretory Vesicle Fusion in Primary Rat Pancreatic β-Cells. Diabetes 55, 1057-1065
25 Fricker L. D., et al. (2000) Identification and Characterization of proSAAS, a Granin-Like Neuroendocrine Peptide Precursor that Inhibits Prohormone Processing. J Neurosci 20, 639-648
26 Dekkers D. H., et al. (2008) Identification by a differential proteomic approach of the induced stress and redox proteins by resveratrol in the normal and diabetic rat heart. J Cell Mol Med 12, 1677-1689
27 Chen L., et al. (2008) Reduction of mitochondrial H2O2 by overexpressing peroxiredoxin 3 improves glucose tolerance in mice. Aging Cell 7, 866-878
28 Hague S., et al. (2003) Early-onset Parkinson's disease caused by a compound heterozygous DJ-1 mutation. Ann Neurol 54, 271-274
29 Yokota T., et al. (2003) Down regulation of DJ-1 enhances cell death by oxidative stress, ER stress, and proteasome inhibition. Biochem Biophys Res Commun 312, 1342-1348
30 Zhong N. & Xu J. (2008) Synergistic activation of the human MnSOD promoter by DJ-1 and PGC-1alpha, regulation by SUMOylation and oxidation. Hum Mol Genet 17, 3357-3367
31 Abderrahmani A., et al. (2006) ICER induced

islets, showing 97% overlap. (e) Percentage of identified proteins in single and pooled islet proteome annotated for gene ontology and KEGG. (f) Comparison of proteins of different cellular compartments identified in single versus pooled islet proteome. (g) Annotation of proteins in single and pooled islet dataset to gene ontology identifiers (cellular compartmentalization) involved in β-cell function. (h) Upregulation of Park-7 (DJ-1) upon high fat diet. Average of 2 (high fat) and 3 (low fat) experiments, per sample 2 mice pooled. Mean value ± SEM.

by hyperglycemia represses the expression of genes essential for insulin exocytosis. Embo J 25, 977–986

32 Dubois M., et al. (2007) Glucotoxicity inhibits late steps of insulin exocytosis. Endocrinology 148, 1605–1614

33 Torii S., et al. (2002) Granuphilin modulates the exocytosis of secretory granules through interaction with syntaxin 1a. Mol Cell Biol 22, 5518–5526

34 Ostenson C.G., et al. (2006) Impaired gene and protein expression of exocytotic soluble N–ethylmaleimide attachment protein receptor complex proteins in pancreatic islets of type 2 diabetic patients. Diabetes 55, 435–440

35 Zhang W., et al. (2002) Down–regulated expression of exocytotic proteins in pancreatic islets of diabetic GK rats. Biochemical and Biophysical Research Communications 291, 1038–1044

36 Varadi A., et al. (1996) Isoforms of endoplasmic reticulum Ca^{2+}–ATPase are differentially expressed in normal and diabetic islets of Langerhans. Biochem J 319(2), 521–527

37 Thurmond D. C., et al. (2003) Glucose–stimulated insulin secretion is coupled to the interaction of actin with the t–SNARE (target membrane soluble N–ethylmaleimide–sensitive factor attachment protein receptor protein) complex. Mol Endocrinol 17, 732–742

38 Tomas A,. et al. (2006) Regulation of pancreatic beta–cell insulin secretion by actin cytoskeleton remodelling: role of gelsolin and cooperation with the MAPK signalling pathway. J Cell Sci 119, 2156–2167

39 Honda K., et al. (1998) Actinin–4, a novel actin–bundling protein associated with cell motility and cancer invasion. J Cell Biol 140, 1383–1393

40 Bosco D., et al. (2000) Importance of cell–matrix interactions in rat islet beta–cell secretion in vitro: role of alpha6beta1 integrin. Diabetes 49, 233–243

41 Ficarro S. B., et al. (2009) Improved electrospray ionization efficiency compensates for diminished chromatographic resolution and enables proteomics analysis of tyrosine signaling in embryonic stem cells. Anal Chem 81(9), 3440–3447

42 Schwartz J. C, et al. (2008) Proc 56th ASMS Conf Mass Spectrom and Allied Topics, Denver, CO, WPAA 039

7

Summary and concluding remarks

Mass spectrometry (MS)-based proteomics is a rapidly maturing technology increasingly demonstrating its large potential in biological and medical research. For many biological questions involving identification and/or quantitation of proteins, well-established and high-throughput LC-MS methods can now be applied.

However, there are a number of applications for which MS potential is tremendous, but for which standard methods are inappropriate or deliver insufficient depth of analysis. In this thesis, we tackled three of these challenging subjects either by optimizing existing approaches or by establishing novel ones.

The first subject, covered in Chapter 2 and 3, deals with the analysis of intact proteins by MS, which is commonly referred to as top-down MS. Its goal is to characterize proteins without prior digestion in order to obtain information about (co-occurrence of) co- and post-translational modifications, splice isoforms and the relative stoichiometry of the resulting protein forms. Combinatorial modifications are known to influence protein signaling, but this information is often lost upon digestion. We investigated the capabilities of the LTQ-Orbitrap for top-down MS analysis, and found that the spectral quality is excellent for proteins of 10 to 25 kDa, with low nanomole sensitivity and full isotopic resolution. With mass accuracies in the low ppm range on both protein and peptide levels, target proteins were confidently identified in the databases and fragments could be assigned unambiguously. Interestingly, we observed several phosphorylated fragment ions in CID fragmentation spectra. On the peptide level, this labile modification often results in an identification hampered by neutral loss.

The observation that the stoichiometry of modified protein forms can be determined from the respective intensities in the survey scan motivated us to explore whether stable isotope labeling by amino acids in cell culture (SILAC) might work in combination with top-down MS. This metabolic labeling technique allows for direct spectral comparison of multiple samples and thereby increases quantitation accuracy. First we mathematically established that the isotopic clusters of the differentially labeled proteins would be separable at any charge state up to at least 146^{+}. Secondly we demonstrated, by analyzing the 28-kDa adapter protein Grb2, that we could determine the ratio between 'light' and 'heavy' labeled forms with 6 % accuracy. The labeling furthermore helped with fragment assignment, as quantized SILAC mass offsets reveal the number of labeled amino acids in the sequence.

Although we demonstrated the potential of the LTQ-Orbitrap and the applicability of SILAC for intact protein characterization, we encountered several problems that currently prevent the application of top-down MS to large scale proteomic projects. The main bottleneck in this regard is the low throughput. At present, there is no LC system that allows high-quality separation for a wide range of proteins and as a result most protein characterizations are done on (semi) purified samples. In addition, the LTQ-Orbitrap only allows accurate analysis

of small proteins. Proteins larger than 40-60 kDa cannot be properly resolved, thereby hindering accurate mass determination and identification. A further complication is the lack of appropriate software for data analysis. Only recently has software become available to identify proteins from top-down data in a somewhat high-throughput fashion. Altogether, despite continued progress being made in this field, it will most likely take several years before top-down MS can be applied side-by-side with bottom-up MS.

The second application, discussed in Chapter 4, describes the establishment of technology aimed at allowing fast and in-depth analysis of simple peptide mixtures. Typically, MALDI-TOF instruments are used to characterize purified proteins. That instrumentation however, delivers inferior spectrum quality and identification confidence compared to the LTQ-Orbitrap which is – at the same time – more versatile. To demonstrate that such analyses can be performed on the orbitrap very quickly and with low sample consumption, we developed a MS method that fully characterizes a protein and its modifications using only small amounts of sample in 1 or 2 µL of spraying liquid. With automated chip-based nanoelectrospray we collected many hundreds to thousands of surveys spectra of different m/z regions, and combined them into one composite spectrum. This resulted in average mass precision in the range of 400 ppb, and a dynamic range of 6700, enabling us to confidently identify 66% of the protein sequence from 500 attomole of albumin as well as various post-translationally modified peptides of histone H3. Since this method can be completed in approximately 5 minutes, it is appreciably faster than any LC-MS based approach and could considerably increase the sample throughput in MS facilities.

To make similar in-depth protein characterization attractive and applicable to proteomic studies, it needs to be compatible with online LC separation. We thus evaluated the possibility of collecting small fractions while analyzing the sample by online LC-MS. Unfortunately, we observed that due to the low flow rates and sample concentrations used in nanoLC, sample losses due to absorption to collection vessels were detrimental. The solution to this problem was to employ a fused silica capillary instead. This technology represents subject three of this thesis (Chapter 5 and 6). As described in Chapter 5, we developed a completely novel LC-MS set-up, called RePlay, enabling the collection of a column effluent inside a long capture capillary. This system, co-developed by and now commercially available from Advion BioSciences, allows fully automated direct reanalysis of the stored effluent after the normal LC-MS measurement has finished. Since the sample is stored after chromatographic separation, the signal intensity is not compromised and with a small focusing column in-line to the mass spectrometer, we obtained excellent chromatographic separation. Conceptually, information obtained in the first run can be utilized to specifically target important proteins or peptides in the 'replay' run. During this reanalysis the next sample can already be loaded on the first column, thereby rendering the LC-MS setup more time efficient.

In Chapter 6 we subsequently described a direct application of this novel system: analysis of low sample amounts, as is often needed when analyzing human tissue material. We optimized the sensitivity of the system by reducing the effective flow rate to 50 nL/min, and ran 4h long gradients. Using the RePlay system we obtained a 'built-in' technical replicate of every precious sample.

We demonstrated this ultrasensitive setup by analyzing single islets of Langerhans from mice pancreas, which are composed of approximately 2000 to 4000 hormone-producing cells. We confidently identified 2013 proteins and showed that with the obtained depth we covered at least 40% of the proteins in all processes related to glucose catabolism and insulin secretion. Furthermore, we quantified over 1400 proteins with multiple peptides and confirmed the regulation of functionally important proteins after 24h glucose stimulation. In addition we even found novel, potential hyperglycemia markers, including DJ-1. Besides doubling the number of proteins identified, we also are the first that obtained important and solid quantitative information from such low sample amounts.

In conclusion, using the novel ultrasensitive setup around the RePlay system, we are able to perform in-depth quantitative MS analysis using less than a microgram of starting material. Thus, by virtue of this methodology, proteomics now enters a variety of research fields that were until now – for reasons of limiting sample amounts – restricted to DNA and RNA analysis.

Although we present a technological development that holds great promise, it should be noted that the characterization of low abundant but biologically important proteins such as transcription factors, transmembrane receptors etc. from samples of limiting protein amounts will still require further developments, especially with respect to detection sensitivity. Inside the mass spectrometer, improvements in hardware and electronics may enable collection of spectra with fewer ions but with higher dynamic range. Faster advances however, seem achievable in optimization of the ionization process. We observed that at reduced flow rates, electrospray ionization was very sensitive to the applied voltages, and current default settings are not necessarily optimal. Further fine tuning of spraying needles in combination with the induced field strength may increase the ionization efficiency.

Also on the chromatographic side, the system may be further optimized to improve sensitivity. Since for small sample amounts surface interaction should be minimized, the entire peptide separation is ideally performed via online LC. In Chapter 6 we used 10 and 22 cm long columns with 4h gradients. Small chromatographic improvements are possible by packing longer or smaller diameter columns. Robust implementation of an orthogonal, second dimension LC (2D-LC) would be far from trivial but perhaps has even higher potential in this regard..

In conclusion, we have developed three different, advanced (LC-)MS methods, all with the aim to analyze proteomic samples to significant depth. All three methods are designed to be used with the latest and most sophisticated mass spectrometry instrumentation. Nevertheless, as explained above, in two out of three applications further improvements are desirable. We have brought these applications a considerable step forward, but they will present exciting challenges for years to come.

8

Nederlandse samenvatting en conclusies

De analyse van alle eiwitten in een cel of weefsel op basis van massaspectrometrie (MS), aangeduid met de term proteomics, is een snel ontwikkelende technologie die veelbelovend is voor biologisch en medisch onderzoek. Met behulp van MS kan zowel de identiteit als de hoeveelheid van een eiwit worden bepaald, ook wanneer er wordt gewerkt met complexe eiwitmengsels.

Voor veel kwalitatieve en kwantitatieve eiwitanalyses bestaan daarom inmiddels goed gedocumenteerde preparatietechnieken en geautomatiseerde meetmethoden. Er zijn echter ook een aantal onderzoeksgebieden waarvoor MS in principe zeer veel potentieel heeft, maar waarvoor de bestaande methoden niet toepasbaar of onvoldoende diepgaand zijn.

In dit proefschrift wordt beschreven hoe we voor een drietal MS toepassingen de bestaande methoden hebben geoptimaliseerd of nieuwe methoden hebben ontwikkeld. Hiermee wordt het groeiende scala aan MS-applicaties verder uitgebreid.

De eerste toepassing, besproken in Hoofdstuk 2 en 3, behandelt de MS-analyse van intacte eiwitten, dat ook wel 'top-down MS' wordt genoemd. In tegenstelling tot de standaard 'bottom-up' proteomics methode, waarbij de eiwitten voor de meting in stukken (peptiden) worden geknipt, is het doel in top-down MS om de eiwitten intact te meten. Dit kan aanvullende informatie opleveren over aanwezigheid van specifieke splice varianten en/of co- en post-translationele modificaties. Vooral wanneer meerdere modificaties op een enkel molecuul aanwezig zijn, is de kennis over een eventuele wisselwerking zeer waardevol. Helaas gaat deze informatie vaak verloren bij eiwitdigestie. Bovendien kan voor alle moleculaire varianten de relatieve stoichiometrie bepaald worden op basis van signaalintensiteit. We hebben de mogelijkheden van de LTQ-Orbitrap massaspectrometer voor top-down analyses onderzocht en stelden vast dat de kwaliteit van de orbitrap spectra zeer hoog is voor kleine eiwitten (10 tot 25 kDa). Slechts honderd femtomol (10^{-15}) eiwit verkregen we intense signalen en behaalden we volledige isotoopresolutie. Door de nauwkeurige massabepaling was het vervolgens mogelijk de eiwitten te identificeren in de database en konden we de fragmenten (als gevolg van geïnduceerde fragmentatie in de massaspectrometer) – eveneens zonder twijfel – toekennen. Interessant was, dat we bovendien verschillende gefosforyleerde fragmenten aantroffen. Op peptidenniveau is de fosfaatgroep vaak instabiel, resulterend in een neutraal massaverlies dat de identificatie van dit soort peptiden bemoeilijkt. Bij eiwitfragmenten was dit duidelijk minder aan de orde.

De observatie dat de verhouding van de gemodificeerde eiwitten vastgesteld kan worden aan hand van de respectievelijke signaalsterktes in MS spectra, motiveerde ons om te testen of eiwitten ook voor top-down analyse gelabeld kunnen worden met 'stabiele isotoop-labeling door aminozuren

in cell cultuur' (SILAC) (Hoofdstuk 3). Deze techniek maakt namelijk een directe spectrale vergelijking tussen meerdere monsters mogelijk en vergroot daardoor de nauwkeurigheid van de kwantificering. Allereerst hebben we door modellering aangetoond dat de isotoopclusters van verschillend gelabelde eiwitten niet overlappen (tot minimaal 146+ ladingen). Daarna demonstreerden we aan de hand van het 28-kDa grote eiwit Grb2 dat de verhouding tussen 'licht' en 'zwaar' gelabelde eiwitten inderdaad zeer precies kan worden vastgesteld (6% onnauwkeurigheid). Bovendien stelden we vast dat SILAC helpt bij de toekenning van fragmenten: uit het massaverschil, geïntroduceerd door de labeling, is het aantal gelabelde aminozuren af te leiden.

Ondanks dat we hebben aangetoond dat de LTQ-Orbitrap en SILAC zeer goed inzetbaar zijn voor de studie van intacte eiwitten, constateerden we ook diverse problemen die de toepassing van top-down analyse op dit moment beperken. Top-down MS is op dit moment nog ongeschikt voor grootschalige proteomics projecten, vanwege de lage 'throughput' en het ontbreken van chromatografische methoden die het brede scala aan eiwitten goed kunnen scheiden. Diverse eiwitten slaan neer of binden irreversibel aan de kolom. Als compromis worden voor top-down MS metingen daarom meestal (half)gezuiverde monsters gebruikt, maar dit brengt over het algemeen veel extra werk met zich mee. Daarnaast is de LTQ-Orbitrap niet goed in staat om eiwitten groter dan 40 of 60 kDa nauwkeurig te meten. Slecht gescheiden isotoopclusters beperken een nauwkeurige massabepaling en bemoeilijken daardoor de eiwitidentificatie. Ten slotte is er een gebrek aan goede 'top-down'-software. Pas sinds kort is er een programma beschikbaar dat intacte eiwitsignalen op relatief geautomatiseerde wijze kan analyseren. Al met al verwachten we dat het – ondanks continue vooruitgang in dit werkveld – nog zeker enkele jaren zal duren voordat top-down MS net zo veelvuldig als 'bottom-up MS' gebruikt zal worden.

De tweede toepassing in dit proefschrift beschrijft een nieuwe methode om simpele eiwitmengsels met 'bottom-up' methoden sneller en toch met meer diepgang te analyseren. In Hoofdstuk 4 tonen we aan dat de LTQ-Orbitrap ook voor dit type metingen zeer geschikt is: beter en tegelijkertijd veelzijdiger dan een MALDI-TOF massaspectrometer die doorgaans voor dit soort analyses wordt gebruikt. We ontwikkelden een methode die met 1 of 2 microliter oplossing eiwitten en eventuele modificaties volledig karakteriseert. Met zogenaamde chip-gebaseerde nanoelectrospray verzamelden we honderden of duizenden spectra, die we vervolgens samenvoegden tot één 'composiet-spectrum'. Deze aanpak leverde een zeer nauwkeurige massabepaling op (gemiddeld ongeveer 400 ppb) en een maximaal intensiteitverschil van 6700; voldoende om met grote zekerheid 66% van de eiwitsequentie van niet meer dan 500 attomol (10-18) albumine te bepalen. Op vergelijkbare wijze karakteriseerden we diverse post-translationele modificaties op histon H3. Dergelijke metingen kunnen worden uitgevoerd in slechts 5 minuten. Dit is aanzienlijk sneller dan conventionele analyses, waarbij de peptiden direct voor MS-bepaling met vloeistofchromatografie (LC) worden gescheiden. De hier beschreven methode is daarom met name interessant voor MS-faciliteiten, die (eenvoudige) monsters voor andere onderzoekslaboratoria meten, aangezien hiermee veel meer monsters per dag kunnen worden gemeten.

Diepgaande analyses zoals hierboven beschreven voor zuivere eiwitten zijn ook wenselijk voor complexe eiwitmengsels in typische proteomics projecten. In dit geval is echter nog altijd een scheiding op basis van vloeistofchromatografie noodzakelijk. Daarom testten we het opvangen van kleine fracties tijdens de normale 'LC-MS'-meting, die we vervolgens naderhand op een vergelijkbare wijze als in Hoofdstuk 4 zouden kunnen analyseren. Helaas constateerden we dat met de typische stromingssnelheden en geringe concentraties, de verliezen als gevolg van absorptie aan de wand van de opvangbuisjes te groot waren. De oplossing van dit probleem bleek het vervangen van deze buisjes door een dun glascapillair. Zoals beschreven in Hoofdstuk 5, ontwikkelden we een nieuwe set-up waarbij na de scheiding van peptiden op een LC kolom een deel van het eluaat wordt afgesplitst en opgevangen in een zeer lang capillair. Dit zogenaamde 'RePlay' systeem is in samenspraak met Advion BioSciences ontwikkeld en via dit bedrijf inmiddels verkrijgbaar. De bewaarde vloeistof kan na de directe LC-MS meting naar de massaspectrometer worden geleid, zodat elk monster direct nogmaals gemeten kan worden. In deze 'replay' meting kan informatie van de eerste analyse worden gebruikt, zodat de onderzoeker belangrijke eiwitten nog eens extra kan bestuderen. Bovendien kan tijdens deze tweede meting een volgend monster al op de eerste kolom geladen worden, en daardoor wordt de meettijd efficiënter benut.

In Hoofdstuk 6 beschrijven we tenslotte een directe toepassing van dit nieuwe systeem: de analyse van zeer geringe monsters. We verbeterden de gevoeligheid van het systeem door de stromingssnelheid terug te brengen naar 50 nL/min en door in plaats van 2 nu 4 uur lange gradiënten te meten. De RePlay leverde een direct duplo-meting op, waardoor we elk monster in acht uur analyseren. Deze aangepaste opstelling hebben we vervolgens gebruikt om afzonderlijke eilandjes van Langerhans te bestuderen. Deze 'mini-orgaantjes' bestaan uit slechts 2000 tot 4000 hormoonproducerende cellen en zijn essentieel voor een goede bloedsuikerspiegel. In totaal konden we in een enkel eilandje maar liefst 2013 verschillende eiwitten aantonen, meer dan dubbel zoveel dan dusver in de literatuur aangetoond (met vergelijkbaar kleine hoeveelheden). Het aantal eiwitten is ook dusdanig groot, dat we meer dan 40% van de eiwitten die een rol spelen bij het glucose metabolisme en insuline secretie, in onze dataset aantreffen. Van meer dan 1400 eiwitten konden we bovendien de hoeveelheid vergelijken tussen eilandjes die 24 uur blootgesteld waren aan een hoog dan wel laag glucose gehalte. Daarmee bevestigden we diverse bekende eiwitveranderingen en vonden ook nieuwe potentiële hyperglykemie markers, waaronder DJ-1. Kortom, de nieuwe ultrasensitieve opstelling op basis van het RePlay systeem maakt het mogelijk om een diepgaande kwantitatieve analyse uit te voeren uitgaande van slechts een paar honderd nanogram eiwitmateriaal. Door deze nieuwe methodologie wordt proteomics nu toepasbaar in tal van nieuwe onderzoeksvelden, die tot dusver alleen te bestuderen waren op DNA of RNA niveau.

Deze technologische ontwikkeling is weliswaar zeer veelbelovend, maar verdere innovatie blijft noodzakelijk om biologisch interessante eiwitten die van nature in zeer geringe mate voorkomen (bijvoorbeeld transcriptiefactoren en transmembraan-eiwitten) beter te detecteren. De belangrijkste innovatie zal een verbetering van de gevoeligheid (sensitiviteit) moeten zijn. Deze gevoeligheid kan worden vergroot door hardware en elektronica-verbeteringen in de massaspectrometer, maar waarschijnlijk

ook door optimalisatie van het ionisatie proces. We constateerden namelijk dat bij de lage chromatografiesnelheden kleine voltage-veranderingen een grote uitwerking hadden op de intensiteit van het signaal. De gebruikte instellingen waren niet noodzakelijk optimaal en een aanpassing van de veldsterkte kan de ionisatie efficiëntie daarom mogelijk verhogen. Ten slotte pleitten we ervoor dat de peptidenscheiding verder wordt geoptimaliseerd. Voor geringe hoeveelheden monster is het essentieel om het contactoppervlak zo klein mogelijk te houden, en dus is chromatografische scheiding direct voor de massaspectrometer ideaal. Hier gebruikten we één type kolom, maar de toevoeging van een extra (orthogonale) scheidingsdimensie (2D-LC) kan een verbeterde scheiding en daardoor hogere sensitiviteit opleveren. Helaas bleek dit tot nu toe allesbehalve triviaal.

Samengevat hebben we drie zeer uiteenlopende, geavanceerde LC-MS methoden ontwikkeld om eiwitmonsters diepgaand te analyseren. Alle drie methoden zijn ontworpen op basis van, en voor de – op dit moment – meest moderne en hoogwaardige massaspectrometrie apparatuur. Niettemin zijn voor twee van de drie toepassingen verder ontwikkelingen wenselijk. We hebben nieuwe toepassingen een aanzienlijke stap verder gebracht, maar desondanks blijven er nog diverse interessante uitdagingen over voor de nabije toekomst.

List of publications

Publications

1 Macek B.[*], **Waanders L. F.***, Olsen J. V., and Mann M. (2006) Top-down protein sequencing and MS3 on a hybrid LTQ-Orbitrap mass spectrometer, Mol Cell Proteomics, 5(5), 949-958 (impact factor: 9.6)

2 Lu A., **Waanders L. F.**, Almeida R., Li G., Allen M., Cox J., Olsen J. V., Bonaldi T., and Mann M. (2007) Nanoelectrospray peptide mapping revisited: Composite survey spectra allow high dynamic range protein characterization without LC-MS on an orbitrap mass spectrometer, Int J Mass Spectrom, 268 (2), 158-167 (impact factor 2006: 2.3)

3 **Waanders L. F.**, Hanke S., and Mann M. (2007) Top-down quantitation and characterization of SILAC-labeled proteins, J Am Soc Mass Spectrom 18(11), 2058-64 (impact factor: 3.3)

4 **Waanders L. F.**[*], Almeida R.*, Prosser S., Cox J., Eikel D., Allen M. H., Schultz G. A., and Mann M. (2008) A novel chromatographic method allows online reanalysis of the proteome; Mol Cell Proteomics, 7, 1452-1459 (impact factor 2006: 9.6) (* Shared first-authorship)

5 **Waanders L. F.**, Chwalek K., Monetti M., Kumar C., Lammert, E., and Mann M. Quantitative proteomic analysis of single pancreatic islets (submitted manuscript)

[*] Shared first-authorship

Oral presentations (at international conferences)

1 Waanders L. F., Macek B., Bonaldi T., Almeida R., Olsen J.V., and Mann M. *54th Conference of the American Society of Mass Spectrometry,* Seattle, USA, May – June 2006

2 Waanders L. F., Raijmakers R., Vree-Egberts W., Olsen J. V., Holmdahl H., Pruijn G., and Mann M. *16th Meeting of Methods in Protein Structure Analysis*, Lille, France, Sept 2006

3 Waanders L. F., Macek B., Almeida R., and Mann M., (Usermeeting) *40. Tagung der Deutsche Gesellschaft voor Massenspektrometrie*, Bremen, Germany March 2007

4 Waanders L. F., Mann M. *6th North American FT-ICR Conference*, Lake Tahoe, USA, April 2007

5 Waanders L. F., Nielsen M. L., and Mann M., (Workshop) *55th Conference of the American Society of Mass Spectrometry*, Indianapolis, USA, June 2007

6 Waanders L. F. (Internet Webinar) Advion BioSciences Inc, November 2007. Available online: *https://advion.on.intercall.com/confmgr/public_stored_docs.jsp?docType=recordings*

7 Waanders L. F., Almeida R., Allen M., and Mann M. *7th HUPO conference*, Amsterdam, NL, August 2008 Finalist of European Proteomics Association – Young Investigator finalist

Acknowledgements

Dankwoord

August 2005, exactly four years ago, was the time I arrived in Munich. Excited to dive into another international experience; to live in the heart of Europe; and thrilled to work in one of the best scientific environments in the world.

Now, four years later, I am writing the last chapter of my thesis - memorizing the great time I had in the 'Mann'-lab... A time with many cultural highlights, friendships, nature adventures, combined with a world-class working environment and many international opportunities. Honestly, a time never to forget!

Now is also the time to thank the people that enabled me to have this exciting experience.

First of all, I would like to thank you, Matthias. You have been a great supervisor! Always with a warm word and an open door, inspiring to deliver the best... I admire your vision and your energetic ability to create opportunities! I also value the responsibility and trust you gave me. I will never forget the first conference I visited, the ASMS 2006 in Seattle: You trusted me to present my work for over a thousand people!

Second of all I thank Ger and Walther. As my supervisors during my master thesis, you have opened many doors for me. In my first year in Bavaria you supported me financially and all these years you have been following me from the sideline. I am happy to graduate in Nijmegen, according to the Dutch ceremonies. Thanks for your kind support during all those years!

Next, a word of thanks to my colleagues. Or I'd better say friends, as we spent a lot of time (in and outside of the lab) together. In our leisure time, we have climbed mountains, drank liters of beer in the Bavarian beer halls and gardens, visited cinema's, restaurants, parks and musea. In Munich and surroundings and in other countries, while visiting conferences. There is too much to mention, so I will only report a few highlights...

Michiel, you were my Dutch buddy. Our 'German period' overlapped almost entirely and thus we explored Munich together. Throughout all the years you were always in for a party or excursion or ready to give (well-intended) advice! ;) Dankjewel dat je vandaag ook mijn paranimf wilt zijn!

Nina, you have become one of my best friends! We have shared a lot of hobbies & dreams in life and I hope we will continue to do so! Thanks for your great friendship!

Cuiping and Francesca, both friends since Denmark. I greatly enjoyed our bicycling trips and chats over the years. Now further apart, but I hope we will keep in touch. Cuiping, it was great to visit you in China! And although Italy was so close, visiting you in your home country, Francesca, is still on my wish-list. But we will come, you can count on that!

Mara, Tami, Maxi and Christian: my fellow 'kitchen-chefs'. I still regularly enjoy making & eating the delicious recipes I learned from you. I think back warmly to our nice lunchtimes and friendship.

Aiping, Cuiping, Chanchal, Ivan, Florian, Stefan & Christian. Being the first Mann-lab grad students, we tried to make our stay as efficient and successful as possible. We shared dreams and concerns... Sometimes we did it the hard way, but we made it!

Stefan, I will miss your great (karaoke) parties!

Tizi, Francesca, Lyris and all the other girls… Thanks for the friendship, and our ladies-nights – there could only be too few!

Johannes, I learned a lot from you. You are the most critical person I have ever met, but fortunately combined with good senses & humor. I appreciate your English lessons as well as your help with my thesis.

Boris and Ana. You were an example of a happy couple living apart for several years. I am happy it worked out for Doede and me, as it did for you. Thanks for your friendship & your hospitality during our nice stay in Cres (Croatia).

Jesper, Boris, Gaby, Sasha, Johannes, Nina, Chuna, Cuiping, Steve, Michael, Sofia, Milko, Peter, Korbi, Bianca, and Mara: office mates over the years. I appreciated your kind chats, advice & support!

Theresa and Tine, for your chats as well as your help in all administrative matters!

Over the years many people worked, visited and left the department. I would like to thank all of you for the nice time we had. For the support, discussions and all the fun!

In my work I have been collaborating with people from other groups or companies.

Most importantly, I had several shared projects with Advion Biosciences. Basically from the start I have worked with you, Reinaldo! And I have always enjoyed it. You taught me a lot of practicalities and the long working hours also created time to get to know each other personally. Thanks for your friendship!

I also want to thank Mark, Glyn, Amy, Tom, Colleen, Gary, Simon, Bob and other colleagues from Advion. It was a pleasure to work with you!

Trying to create new applications based on high-tech mass spectrometry equipment, I benefitted a lot from the technical knowledge from Volker, Georg, Steven, Alexander, Vlad, Bernard, and other colleagues in Thermo Fisher Scientific. I greatly appreciate your help and time!

Ecki & Karolina, we started collaborating relatively late, but it was fast and successful. Thanks.

***

Daarnaast wil ik alle Nederlandse vrienden bedanken die me in de afgelopen jaren hebben gesteund. Soms door goede gesprekken in/vanaf Nederland, en nog leuker: door ontspannende weekendjes in Beieren. De drie jaar heeft bovendien geresulteerd in twee mooie kaartenwanden en ontelbare foto's. Voor het uitzoeken en inplakken zal nu eindelijk ook tijd zijn!

Iedereen die me in de afgelopen jaren heeft opgezocht, heel erg bedankt: Nina en Martijn, Jitske en Jelle, Harmen, Erik T en Jelle, Esther, Erik S, Erik H en Maaike, Mariska, Margiet, Jorrit, Ellen, Wietse en Arjan… Behalve dat het altijd leuk was om de Beierse Alpen of omgeving beter te leren kennen, bood het de ideale gelegenheid om lekker uitgebreid bij te kletsen. Onze vriendschap heeft niet geleden onder de afstand. Dankjewel!

In het bijzonder wil ik jou bedanken, Nina. Jij was de eerste die me opzocht, (en als het aan jou lag ook de laatste...) Altijd stond je klaar met een luisterend oor, raad en daad. Jij kent München inmiddels bijna net zo goed als ik!

Goede herinneringen heb ik ook aan Erik T en Jelle, die spontaan bleven slapen, nadat bleek dat zomerbanden toch niet de beste keus was om (op weg naar wintersport) de ergste sneeuwbuien van de eeuw te trotseren.

Spannend was het eveneens met jou, Erik S, toen we pas bij schemering aankwamen in de Alpenhut. Een tweedaagse trektocht, vol gesprekken over succes in werk & liefde. We hebben veel gemeen. Ik ben blij dat ook jij mijn paranimf wilt zijn!

Zeer dankbaar ben ik ook mijn familie en schoonfamilie voor alle steun en mede-leven in de afgelopen jaren.

Mijn lieve zusjes, Judy en Maureen wil ik even speciaal noemen. Met kaartjes, telefoontjes en bezoekjes stonden jullie jullie altijd klaar als dat nodig was. Fijn om elkaar nu weer wat vaker in levende lijve te zien!

Pap & mam, ook jullie stonden altijd voor me klaar.. Jullie hebben me regelmatig opgezocht (met & zonder fiets) en hoorden dan altijd geduldig al mijn enthousiaste verhalen aan! Fijn dat jullie me zo gesteund hebben in alles wat ik deed! Jullie liefde betekent heel veel voor me!

En de meeste dank ben ik verschuldigd aan Doede. Jouw steun en liefde bleek echt onvoorwaardelijk! Hoe vaak ben je niet naar München gereden... soms midden in de nacht… om mij te bezoeken. En dan moest ik nog wennen aan het samenzijn ook! Toch denk ik dat we de afgelopen jaren goed doorstaan hebben: we hadden door de week allebei tijd voor ons werk en hobby's en namen in de weekenden echt tijd voor mekaar. Doede, je wachten is beloond: het boekje is af en ik ben weer terug in Nederland om samen met jou van het leven te genieten!

Een dikke knuffel (a big hug),

Leonie