Dissertation zur Erlangung des Doktorgrades
Der Fakultät für Chemie und Pharmazie
Der Ludwig-Maximilians-Universität München

Development of proteomics techniques for individual protein and membrane protein mapping

Aiping Lu

吕爱平

aus

Zhejiang, V.R.China

2008

Erklärung

Diese Dissertation wurde im Sinne von § Abs. 3 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Dr. Matthias Mann betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, am ....08.12.2008.....

.....

(Unterschrift des Autors)

Dissertation eingereicht am .....

1. Gutachter Prof. Dr. Matthias Mann
2. Gutachter Prof. Dr. Jacek R. Wiśniewski

Mündliche Prüfung am .01.12.2008.....

Table of Contents

Abbreviations

ABC transporter: ATP-binding cassette transporters

AC: alternating current

AMPA: alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid

AQUA: absolute quantitation

ATP: Adenosine-5'-triphosphate

CDITs; culture-derived isotope tag

CHAPS: 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate

CID: collision induced dissociation

CMC: critical micelles concentration

CNBr: cyanogen bromide

DC: direct current

EGF: epidermal growth factor

ESI: electrospray ionization

FT-ICR or FT: Fourier transform ion cyclotron resonance

FTMS: Fourier transform ion cyclotron resonance mass spectrometer

GABA: gamma-aminobutyric acid

GO: gene ontology

GPCR: G protein-coupled receptor

G protein: guanine nucleotide-binding proteins

GTP: Guanosine-5'-triphosphate

HCD: high energy C-trap dissociation

HPLC: high performance liquid chromatography

ICAT: isotope-coded affinity tags

ICR: ion cyclotron resonance

iTRAQ: isotope tags for relative and absolute quantification

KDa: kilo Dalton

KE: kinetic energy

LTQ: Linear trap quadrupole, refers to linear ion trap

MALDI: matrix-assisted laser desorption/ionization

MDR: multidrug resistance

MMTV: Mouse Mammary Tumor Virus

MS: mass spectrometry

NBFs: nucleotide-binding folds

NMDA: N-methyl D-aspartate

ORFs: open reading frames

PEG: polyethylene glycol

PM: plasma membrane

ppb: part per billion

ppm: part per million

RF: radio frequency

rms: root mean square

SDS: sodium dodecyl sulfate

SILAC: stable isotope labeling by amino acid in cell culture

SIM: selected ion monitoring

SLC: solute carrier

SWIFT: stored waveform inverse Fourier transform

TGF-beta: transforming growth factor-beta

WGA: wheat germ agglutinin

XIC: extracted ion chromatogram

1 Introduction

1.1 Histone

Histones are the chief protein components of chromatin. Five major histone classes are known, H1, H2A, H2B, H3 and H4. Two copies of H2A, H2B, H3 and H4, are assembled into an octamer wrapped around by DNA to form a nucleosome core[1]. The repeating nucleosome cores further assemble into higher-order structures which are stabilized by the linker histone H1(ref [2]). The nucleosome, including nucleosome core, linker DNA and H1, is the principal packaging element of DNA within the nucleus (Fig. 1).

1.1.1 Core histones

The nucleosome core is formed of two copies of H2A, H2B, H3 and H4 and the octamer structure can be divided into H3-H4 tetramer and two H2A-H2B dimers[1]. Each histone consists of a globular domain and an N-terminal tail domain. The globular domain interacts with other histones and DNA in the nucleosome, and the N-terminal tail is flexible to protrude from the lateral surface of the histone octamer[3]. The flexible tails undergo a diverse array of post-translational modifications which correlate with specific transcriptional states[3]. For example, acetylated lysines in the histone N-terminal tail can selectively interact with the bromodomains in different transcription factors to increase transcriptional activity[4-6]. H3S10 phosphorylation is necessary for proper chromosome condensation and segregation[7], and H3K9 methylation interacts with the chromodomains in heterochoromatin protein HP1 and stabilizes higher order chromatin structure[8, 9]. Because of these correlations, a histone code hypothesis has been proposed[10-13]. The hypothesis suggests that histone proteins and their associated covalent modifications would affect the accessibility of DNA within the chromatin structure, thereby playing a major role in the regulation of gene expression. The hypothesis is raised based on the post-translational modifications on the flexible histone tail, however, recent proteomics

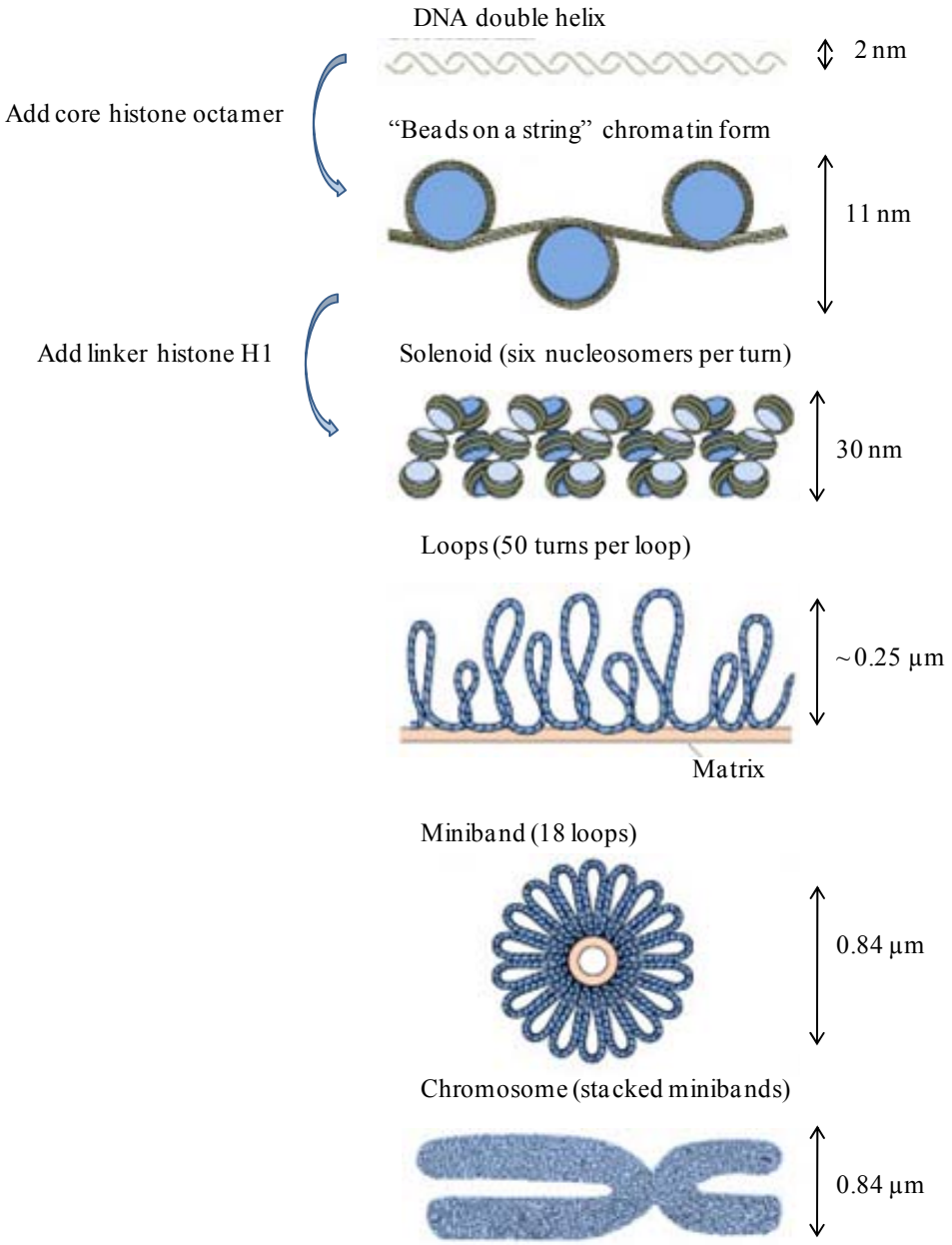


Figure 1. A model of chromosome structure. The double stranded DNA is wound twice around core histone octamers to form 10-nm nucleosomes. With the help of linker histone H1, nucleosomes are arranged into 30-nm fibers. The 30-nm fibers form long DNA loops. The loops then form minniband units of a chromosome. The picture is modified from Biochemistry by Reginald H. Garrett and Charles M. Grisham, second edition.

analyses have revealed a large number of that lie in the structured globular histone core[14-17]. Hence, a model called "regulated nucleosome mobility" was proposed[18]. The model suggests that there are two states of nucleosomes, a mobile state and a relatively stationary state. The post-translational modifications on the globular histone domain may influence the equilibrium of the two states.

1.1.2 Linker histone

The linker histone H1 is involved in assembling the "beads on a string" structure into higher order chromosome structure[2]. Compared to core histone, the linker histone is more heterogeneous and the temporal and spatial expression of different H1 membranes are also different[19]. Mammalian cells contain seven major variants of histone H1: H1.0, H1.1, H1.2, H1.3, H1.4, H1.5 (ref [20-23]), and H1X[24, 25]. In most human cells, H1.2 and H1.4 are the predominant variants[26]. All H1 variants in higher eukaryotes have the same general structure, consisting of a flexible N-terminal tail, a central conserved globular domain, and a lysine rich C-terminal tail[27]. While the globular domain is conserved both orthologous and paralogous, the N-terminal and C-terminal tails is conserved orthologous but not paralogous, i.e. it differs between isoforms in the same species[19]. However, the tri-partite structure is not conserved in lower eukaryotes. For example, in S. cerevisiae, the sole linker histone Hho1p possesses two globular domains[28].

The binding of histone H1 to the nucleosome core structure is largely mediated by the globular and the C-terminal domains, and might be regulate in part by the N-terminal domain[19]. This binding is highly dynamic. The equilibrium constant of the interaction is in favor of association of linker histone to DNA, therefore, most of the chromatin is indeed covered by histone H1(ref [29]). Both *in vitro*[30, 31] and *in vivo*[32] experiments show that the binding affinity of different H1 variants to the chromatin can be hugely different. Although the globular domain plays a key role in regulating the binding of H1 to native chromatin[33], since they are highly conserved between different histone variants, it was demonstrated that the affinity variability of H1 variants to chromatin depends on the length, the density of positively charged residues and the S/TPXK motifs of the C-

terminus[32]. Mutations in the S/TPXK motif or phosphorylation at these sites may disrupt the ability of H1 to bind to chromosomes[34].

The biological function of H1 has been studies by different approaches and was comprehensively reviewed by Izzo *et al*[19] and Enigmatic *et al*[35]. Although all knockout experiments performed in either single cell organism or higher eukaryotes show that H1 is not essential for survival, depletion of different H1 variants in mice do show that different H1 variants are essential for normal development in mammals[36] and different H1 variants play a specific roles in the control of gene expression and chromatin structure[37]. More interestingly, a connection between H1 depletion and core histone N-terminal tail modification and DNA methylation has been observed[38, 39]. By studying H1 function *in vivo* and *in vitro* at certain model promoters, e.g. Mouse Mammary Tumor Virus (MMTV) promoter[40-42], specific functions of some histone variants could be obtained. A summary of so far known functions of different histone variants is shown in Fig. 2.

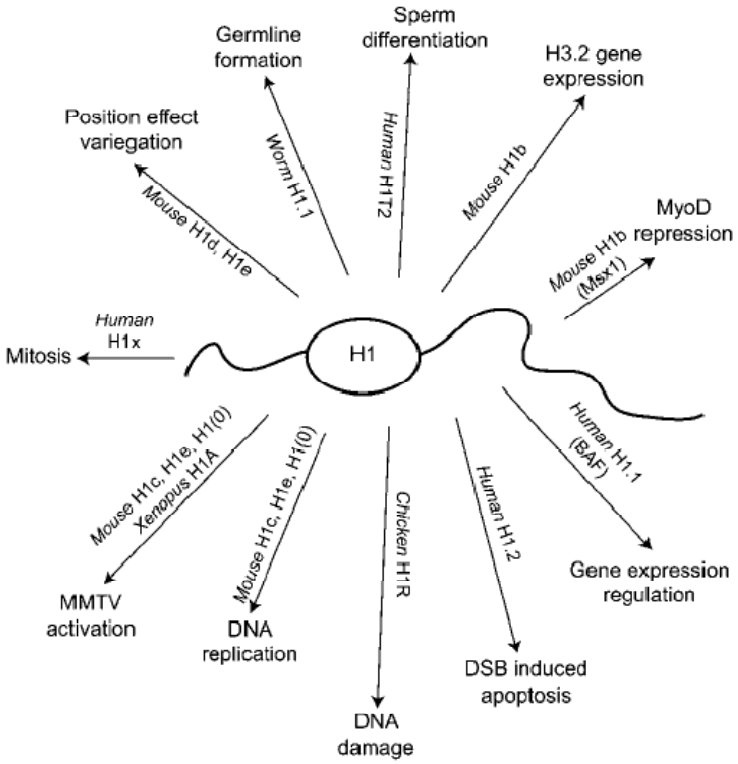


Figure 2. Overview of the multiple function of H1 (from Izzo *et al*[19]).

As for core histone, post-translational modifications on linker histone H1 have also been studied. The first recognized covalent modification of H1 was phosphorylation[43]. By using traditional antibody detection methods, nine phospho-sites and one methylated site have been identified in around 20 years[35]. Applying mass spectrometry (MS)-based proteomics approaches greatly speeds-up the identification, especially for other types of modifications besides phosphorylation. However, the first published papers were still concentrated on phosphorylation of H1 in lower eukaryotes in different cell lines[44-46]. The first comprehensive mapping of modifications of H1 variants in different tissue types was published by our group[47]. In this paper, all previously reported phospho-sites, but also a host of novel sites were identified, not only for phosphorylation but also formylation, acetylation, methylation and ubiquitination. Many of the modification sites locate in the globular domain. In the paper, data also showed different lysine methylation patterns in cell culture and mouse tissue sample, which inspired us to study the H1 methylation in human tissues. The already known methylation site on K26 (H1.4) escaped identification by MS coupled with online high performance liquid chromatography (HPLC). This prompted us to use a complementary method, which is described in the third part of the thesis.

1.2 Purification and isolation of membrane proteins

Membranes provide a selectively permeable barrier physically separating the cell from its environment (plasma membrane, PM) or intracellular organelles from each other and the cytosol (intracellular membranes). Proteins embedded in the membranes play an important role in fundamental biological processes, such as cell signaling, cell-cell communications, cell adhesion, intracellular organelle compartmentalization, ion and solute transport and energy generation. It has been estimated that 20-30% of the ORFs (open reading frames) in the various genomes encode membrane proteins[48] and around 70% of all druggable proteins can be classified as membrane proteins[49, 50]. Although MS-based proteomics has made rapid progress in the analysis of soluble proteins in recent years, the analysis of membrane proteins lags behind due to their high hydrophobicity, thereby causing problems with the normally used aqueous buffers. The biological importance of membrane proteins has induced researchers to develop new technologies for membrane protein purification and isolation.

1.2.1 General introduction of membrane proteins

The main components for biological membrane are phospholipids and proteins, nearly at a 1:1 mass ratio in most animal cell membranes. Although the basic structure and function of the biological membranes is provided by the amphipathic phospholipid bilayer, membrane proteins provide unique compartment specific functions and communication between separated environments. Different membrane proteins are associated with the membranes in different ways, as illustrated in Fig. 3. Based on the strength of the association with the membrane, membrane proteins can be classified into two groups, integral membrane proteins (1 – 6, Fig. 3) and peripheral membrane proteins (7 and 8, Fig. 3). Integral membrane proteins are permanently attached to the membrane, and they contain transmembrane segments, which span the entire membrane (1 – 3, Fig. 3) and integral monotopic protein which are permanently attached to the membrane from only one side (4 – 6, Fig.3). Peripheral membrane proteins are temporarily attached either

to the lipid bilayer or to integral proteins by a combination of hydrophobic, electrostatic interactions, hydrogen bonds or van der Waals interactions.

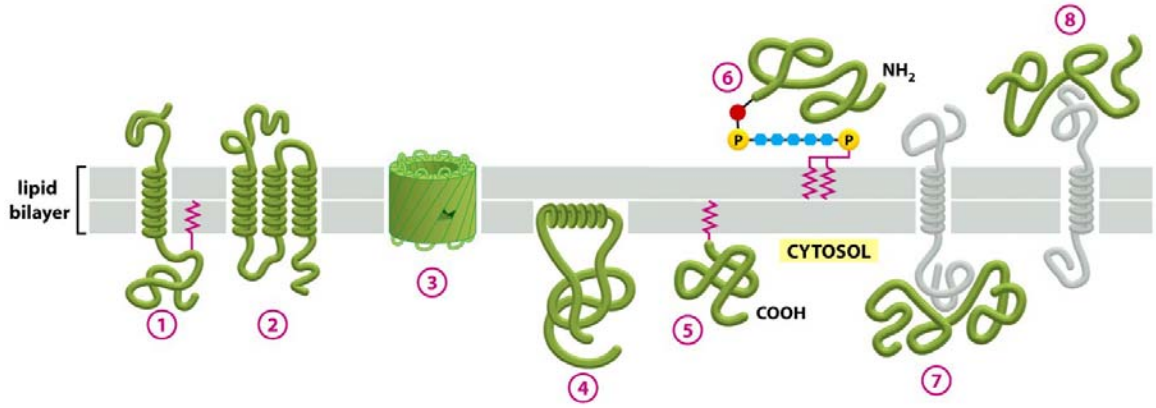


Figure 3. Various ways in which membrane proteins associate with the lipid bilayer. (1) extend across the bilayer as a single α helix, (2) as multiple α helices, or (3) as a rolled-up β sheet, (4) anchored to the cytosolic surface by an amphipathic α helix, (5) attached to the cytosolic monolayer by a covalently attached lipid chain, (6) attached to the non-cytosolic monolayer via an oligosaccharide linker (7, 8) attached to the membrane only by non-covalent interactions with other membrane proteins. From Molecular Biology of the Cell by Alberts *et al* (fifth edition).

1.2.2 Functions of membrane proteins

Membrane proteins transfer either molecules or information cross the lipid bilayer. It takes many different membrane proteins, coordinating in a well organized manner, to enable this passage to work properly. According to their function, membrane proteins can be grouped into different families. Receptors and membrane transport proteins, including ion channels, solute carriers and ATP-binding cassette transporters, are two main families of integral membrane proteins.

1.2.2.1 Receptors

Receptors on the cell surface convert an extracellular ligand-binding event into intracellular signals, by either structure transformation or post-translational modifications, to activate the target cell to respond to environmental stimulus. According to their transduction mechanisms, most cell surface receptor proteins belong to one of three classes: Ion-channel-coupled receptors, G-protein-coupled receptors and enzyme-coupled receptors.

Ion-channel-coupled receptors

Ion-channel-coupled receptors, also known as ligand-gated (transmitter-gated) ion channels or ionotropic receptors, are a group of intrinsic transmembrane ion channels that are opened or closed in response to binding of a chemical messenger. They are involved in rapid synaptic signaling between nerve cells and other electrically excitable target cells such as nerve and muscle cells. Typical examples contain the nicotinic acetylcholine receptor[51], ionotropic glutamate receptors including AMPA (alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptor[52], kainate receptor[53] and NMDA (N-methyl D-aspartate) receptors[54], and GABA (gamma-aminobutyric acid) receptors a and c[55]. Most ion-channel-coupled receptors belong to a large family of homologous, multipass transmembrane proteins.

G-protein-coupled receptors (GPCR)

G-protein-coupled receptors, also known as seven transmembrane domain receptors or heptahelical receptors, are involved in a wide variety of physiological processes, such as visual, olfactory and gustatory sensation, intermediary metabolism, cell growth and differentiation[56]. The three principle components of GPCR signaling are the heptahelical receptor, heterotrimeric G protein and effector protein (typically an enzyme or ion channel). Once the receptor binds to the ligand, it triggers the heterotrimeric G protein to

dissociate into a GTP-bound Gα subunit and Gβγ heterodimer, either of which can activate a cytosolic effector. The activation of the effector can change the concentration of one or more small intracellular mediators (if the effector is an enzyme), or it can change the ion permeability of the PM (if the effector is an ion channel) and the small intracellular mediators (second messengers) further transmit signals intracellularly. A typical example of a GPCR is the metabotropic glutamate receptors[57, 58]. All GPCR belong to a large family of homologous, multipass transmembrane proteins.

Enzyme-coupled receptors

Enzyme-coupled receptors function either directly as enzymes or associate with enzymes that they activate. The great majority of these enzymes are protein kinases, such as receptor serine/threonine kinases including transforming growth factor-beta (TGF-beta), receptor tyrosine kinases including epidermal growth factor (EGF) receptors and insulin receptors. The enzyme-coupled receptors are usually single-pass transmembrane proteins that have their ligand-binding site outside the cell and their catalytic or enzyme-binding site inside.

1.2.2.2 Membrane transport proteins

Although small nonpolar molecules, such as hormone, steroid, N_2, O_2, and CO_2, diffuse rapidly across the cell membrane, cell membranes are highly impermeable to large molecule such as amino acids, sugars and nucleotides, and charged molecules, no matter how small they are. Special membrane transport proteins are responsible for transferring such solutes across cell membranes. The proteins may assist in the movement of substances by facilitated diffusion or active transport. Channel proteins and carrier proteins are the two major classes of membrane transport proteins.

Ion channels

Ion channels are pore-forming proteins that help establish and control the small voltage gradient across the PM of all living cells. Typical assemblies of ion channel usually involve a circular arrangement of identical or homologous proteins closely packed around a water-filled pore through the lipid bilayer. The pore-forming subunit(s) are called the α subunit, and the auxiliary subunits are called β, γ or δ subunit. For example, the voltage-gated sodium channel in mammalian neurons is composed of a 260 kDa α subunit which forms the pore and one or more auxiliary β subunits[59], and the voltage-dependent calcium channels are formed as a complex of α1, α2δ, β1-4, and γ, where the α1 subunit has 24 putative transmembrane segments and forms the ion conducting pore[60].

Classified by the nature of their gating, ion channels can be divided into voltage-gated ion channels and ligand-gated ion channels. As the name indicates, voltage-gated ion channels activate or inactivate depending on the voltage gradient across the PM, whereas ligand-gated ion channels activate or inactivate depending on binding of ligands to the channel. Ligand-gated ion channels also named ion-channel-coupled receptors have already been discussed before. Typical voltage-gated ion channels encompass voltage-gated potassium channels[61], voltage-gated sodium channels[59], voltage-gated calcium channels[60] and voltage-gated proton channels[62].

Solute carriers

Solute carriers are proteins that transport a specific substance or group of solutes through intracellular compartments or in extracellular fluids (e.g. in the blood) across the cell membrane. Unlike ion channels, which interact with the solute to be transported very weakly and can only allow solutes to cross the membrane passively, solute carrier proteins bind the specific solute to be transported and undergo a series of conformational changes to transfer the bound solute across the membrane. This transport can be either facilitated diffusion or active transport. The SoLute Carrier (SLC), for example, includes over 300 members organized into 47 families[63].

ATP-binding cassette transporters (ABC)

ABC-transporters utilize the energy of ATP hydrolysis to transport various substrates across cellular membranes. They are classified as ABC transporters based on the sequence and organization of their ATP-binding domain(s), also known as nucleotide-binding folds (NBFs)[64]. It is known that ATP binding leads to dimerization of the two ATP-binding domains, and ATP hydrolysis leads to their dissociation. These structural changes in the cytosolic domains are thought to be transmitted to the transmembrane segments, driving cycles of conformational changes that alternately expose substrate-binding sites to one or the other side of the membrane. ABC transporters are known to play a crucial role in the development of multidrug resistance (MDR)[64, 65].

1.2.3 Purification of membrane proteins for shotgun proteomics

Traditional MS-based proteomic analyses utilized two-dimensional gel electrophoresis to separate complex protein samples[66]. However, membrane proteins are normally underrepresented on the gel due to their alkaline and poorly soluble properties. Moreover, they are generally not very abundant, so that they cannot even be detected in standard gels[66]. Bottom-up shotgun proteomics approaches based on separating digested peptides with HPLC prior to MS acquisition provides a powerful alternative to 2D gel based proteomics. Sample preparation for shotgun membrane proteomics normally includes soluble proteins and membrane associated protein removal, delipidation, membrane protein solubilization and digestion. Subcellular fractionation is also important for PM proteomics, one of the most interesting proteomes. High salt buffer and high pH sodium carbonate (or sodium hydroxide) buffer are used in nearly all membrane fraction preparation protocols to remove soluble and peripheral membrane associated proteins[67, 68]. The most widely used delipidation approach in the literature is methanol/chloroform precipitation[69-71]. Approaches for membrane fraction purification and membrane protein solubilization are quite diverse. Efforts to improve data quality of membrane proteomics

analysis by biochemical researchers generally concentrate on these two issues, as well as the separation methods for digested peptides before MS acquisition. Since the peptide separation is a technique issue regardless of the properties of membrane proteins, here we focus our discussion on fractionation and solubilization of membrane proteins. Furthermore, the developments of MS instruments and corresponding data analysis software do of course increase the data quality tremendously, as discussed in the next section.

1.2.3.1 Fractionation of membrane protein

Several approaches have been established for fractionation of membrane proteins, in particular the most interesting PM proteins. In a wide sense of the word, centrifugation is essential in every step of purification. Special buffered centrifugation, e.g. sucrose gradient centrifugation[72], can be used alone for subcellular fractionation. Aqueous two-phase systems and affinity purification are two other methods especially setup for PM purification[67, 68]. The general principle and application of the methods are described below.

Centrifugation

Centrifugation separates subcellular compartments according to sedimentation velocities and/or buoyant densities[73]. By applying a well defined g-force for a certain length of time and suspension buffers, differential centrifugation is a rapid means for subcellular fractionation. A two-step differential centrifugation is widely used for crude membrane preparation prior to any other further subcellular purification[74]. In the first step, low speed centrifugation, e.g. 1000 g, is used to remove cell debris and intact nuclei. In the second step, ultracentrifugation, e.g. 100,000 g, is used to remove soluble cytosolic proteins.

Sucrose gradient centrifugation is another widely used method for different organelle membranes or intact organelles separation according to their buoyant density[72]. The gradient can be either a step or continuous function. Although a continuous gradient provides better resolution, the step gradient stacks the sample in a thin layer in the interface of different sucrose concentration, leading to a higher concentration and yield. Furthermore, compared to the continuous gradient, no gradient forming device is needed for the step gradient and the centrifugation time is 1h instead of 10h. The simple work flow and the higher yield make the step gradient more popular for large scale proteomics study with crude membrane preparation[70, 75-81].

Besides sucrose, which is the most commonly used gradient medium, there are other alternatives such as Percoll[82, 83], Ficoll, Nycodenz, or glycerol (for a review see[73]). Taking Percoll as an example, the density of the mixture is chosen to be smaller than the particles at all points during the separation and the run is terminated before the separated zones reach the bottom of the tube. In one report, the membrane content of the identified proteins enriched by Percoll gradient was up to 60% (ref [82]).

Aqueous polymer two-phase systems

The aqueous polymer two-phase system is widely used for analytical PMs purification[78-80, 84-89]. If two structurally distinct water-soluble polymers are mixed above a critical concentration in aqueous solution, the polymers will eventually separate into two phases. The most commonly used aqueous polymer two-phase systems for membrane separation is the polyethylene glycol (PEG)/ dextran system. The partitioning behavior of the protein in the PEG/dextran system depends on the concentrations of the polymer. When the concentrations are close to the critical point of the PEG/dextran two-phase system, membranes tend to partition in the top PEG phase. Increase in polymer concentrations results in larger differences in the composition of the two phases. As a result, membranes tend to partition to the interface or the bottom phase. This dependency of the partitioning behavior on the polymer concentrations can be exploited to selectively enrich PMs in the top phase. In plants and in animals, PMs show the highest affinity for the more

hydrophobic top phase, followed by Golgi vesicles, lysosomes, the endoplasmic reticulum and mitochondria[85]. The degree of polymerization of PEG also affects the phase separation and the partitioning of molecules during extraction, increase of the molecular weight of PEG results in decreased partition of the membranes into the top phase, whereas increase of the molecular weight of dextran gives rise to increased partition of the PMs to the top phase[85]. In addition, salt and ligand in the aqueous system affect the partition behavior of membrane. Since there are so many variables affecting the partition behavior, application of the aqueous two-phase system requires carefully controlled conditions and optimization for specific tissue sources.

The protocol for aqueous tow-phase system seems to be different from laboratory to laboratory, and even different with time in the same laboratory. Because the isolation of PMs cannot be achieved through a single-step procedure, multiple extraction procedures such as countercurrent distribution experiments as shown in Fig. 4 are applied[84, 85]. In the protocol shown, the PMs are enriched in the top phase of G, and the recovery is around 18% in an optimized system[85]. In a simplified workflow, only the first two steps of Fig. 4 are used[78, 79]. After several rounds of the protocol in Fig. 4, the PMs enriched with PEG can be mixed with fresh WGA (wheat germ agglutinin)-dextran, where PMs are selectively pulled into the WGA-dextran enriched bottom phase[86]. For the optimized conditions as reported by Schindler *et al*, the recovery of PMs is 15% and the PM content of the identified 525 proteins is about 27-38%, depending on the prediction software and criteria used[84]. In a rat liver membrane proteomics project published by Cao *et al*, 23% of 883 identified proteins were GO (gene ontology) annotated to be integral membrane proteins or membrane associated proteins[79].

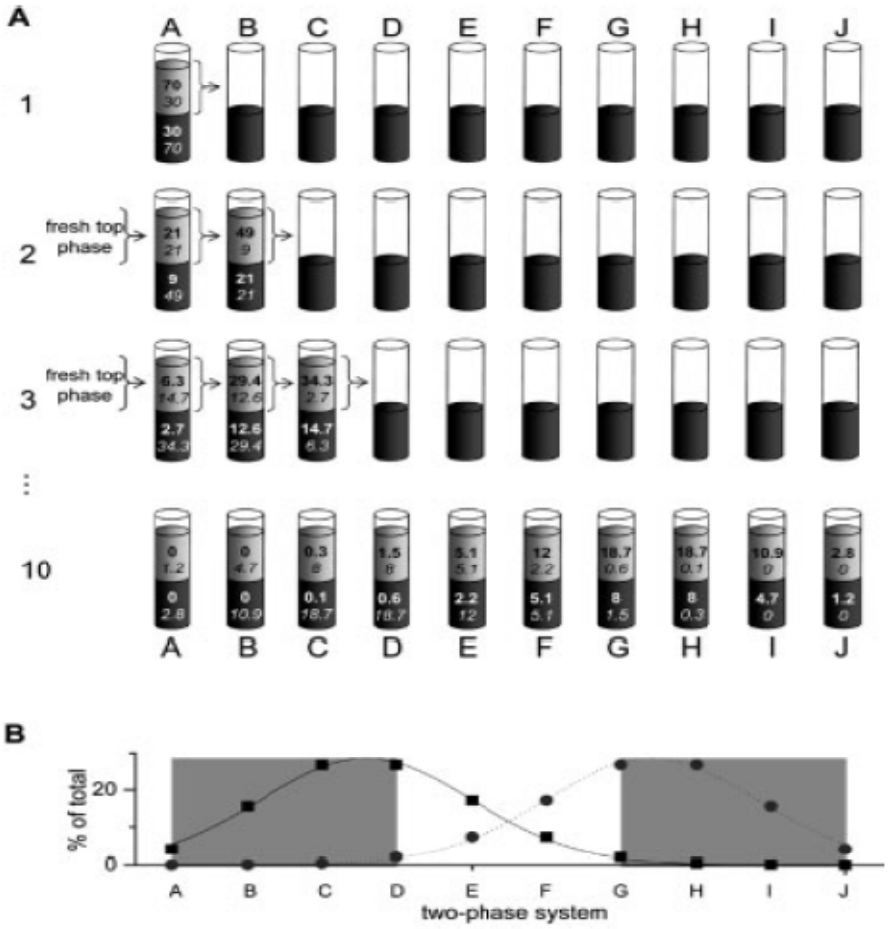


Figure 4. Protocol example for aqueous two-phase partition. A) Tissue is homogenized in a two-phase system and after phase separation, the top phase is transferred onto the fresh bottom phase 'B', and the bottom phase of the primary two-phase system 'A' is covered with a fresh top phase. In the subsequent steps, the top phases are transferred one bottom phase along, always transferring the latest top phase onto a fresh bottom phase and re-extracting bottom phase 'A' with a fresh top phase. By doing so, one more two-phase system is extended in each step, e.g. 10 two-phase systems ('A'–'J') are obtained after the 8th step, each containing a top and a bottom phase. In this protocol, the PM proteins are enriched in the top phase of G and the calculation is based on the assumption that 70% PM fraction partition to the top phase and 30% to the bottom, the value is assumed to be opposite for contaminate membrane fraction. B) the outcome for each fraction illustrated with PMs in circles and intracellular membranes in squares. From Schindler *et al* [85].

Affinity enrichment

The affinity purification of PMs utilizes the properties of different extracellular domains of integral membrane proteins. Two commonly used methods are biotinylation affinity purification[90-95] and Lectin affinity purification based on glycosylation[96]. Compared to biotinylation affinity purification, no labeling procedure is needed for glycosylation affinity purification. The biotin group is normally reacted *in situ* with the ε-amino group of lysine, Two commercially available biotinylation reagents sulfo-NHS-SS-biotin[90-94] and sulfo-NHS-LC-biotin[95] are used. Both reagents have internal disulfide bonds, which allow for the cleavage from the avidin resin by treatment with reducing agent. It was reported that the biotin affinity purification procedure was able to achieve a 1600-fold relative enrichment of PM versus mitochondria and a 400-fold relative enrichment versus endoplasmic reticulum[91]. The best enrichment results by biotin affinity purification obtained so far is 526 integral PM proteins out of 898 identified proteins, which is 58.6% (ref [90]). The data for glycosylation purification is very limited, and the enrichment is not as efficient. Although the biotinylation enrichment seems very promising for cell line PMs purification, it is not suitable for tissue samples. The connection between adjacent cells and the surrounding matrix make the PM proteins inaccessible for labeling.

1.2.3.2 Solubilization and digestion of membrane proteins

Because of the very hydrophobic properties, completely dissolving membrane proteins in aqueous solution is challenging. Detergents therefore play an indispensible role in membrane protein solubilization but, since they are not compatible with MS, they are routinely removed by 1 or 2D PAGE. Because of the limitations of gel based approaches, e.g. proteins may not be completely digested, or peptides may not be extractable from the gel, gel-free but detergent based proteomics would be ideal, as well as other dissolving methods, such as organic/acid solubilization and 'on membrane digestion' (for a review, see [97]).

Detergent solubilization

Detergents are indispensible in membrane proteomics. Nearly all above mentioned projects except[82, 94] used detergents as the membrane protein solubilization reagent. Detergents are amphipathic molecules that contain a polar group (head) at the end of a long hydrophobic carbon chain (tail). The majority of the lipids that make up the membrane contain two hydrophobic groups connected to a polar head, which can be viewed as biological detergents. Detergents solubilize membrane proteins by mimicking the lipid-bilayer environment. The critical micelles concentration (CMC) is an important parameter to be taken into consideration when using detergents. CMC is defined as the concentration of detergents above which micelles are spontaneously formed. Therefore, by applying detergents at a concentration above the CMC, hydrophobic regions of membrane proteins, normally embedded in the membrane lipid bilayer, are now surrounded by a layer of detergent molecules and the hydrophilic portions are exposed to the aqueous medium. This micelle structure keeps the membrane proteins in solution.

Based on the nature of the hydrophilic head group, detergents can be broadly classified as ionic, non-ionic, and zwitterionic detergents, which are exemplified by SDS, Triton and CHAPS respectively. Different detergents may have different preferences in dissolving different group of membrane protein, but so far no directly comparison of applying different detergent in the proteomics area has been reported. Normally researchers choose the detergent according to their own experience or sometimes make use of a combination of different groups of detergents. Since even a small concentration of a detergent completely dominates mass spectra and precludes peptide or protein analysis, detergents have to be efficiently and thoroughly removed from proteins or peptides in MS-based proteomics analysis. Different methods have been described for separation of proteins from detergents including gel filtration, ion-exchange and hydrophobic adsorption chromatography, density gradient centrifugation, dialysis, ultra filtration, phase partition, and precipitation (for a review see[98]).

Organic solvent solubilization

Another alternative way to dissolve or extract membrane proteins is by performing intermittent vortexing and sonication in 60% organic solvent (e.g. methanol[94, 99-101]). Trypsin digestion is immediately carried out in the organic-aqueous solvent mixture. The total number of identified proteins in the above studies range from 117 (ref [101]) to 786 (ref [94]). In the best isolations, 42% proteins have at least one transmembrane domain. One paper reports that organic solvent solubilizes membrane protein more efficiently than SDS[102]. However, since only 299 proteins were identified in that experiment using SDS extraction, the protocol itself is clearly not optimal.

Acid solubilization.

A high concentration (up to 90%) of formic acid was also reported to be effective in solubilizing membrane proteins[103]. In this method, cyanogen bromide (CNBr) is used to cleave many embedded membrane proteins at the C-termini of methionine under acidic conditions. Formic acid is used instead of hydrogen chloride (0.1M) because it dissolves most proteins and its reductive property keeps methionine from oxidation, which is inert to CNBr attack. Formic acid causes the formation of formyl esters to serine or threonine – a potential problem for further analysis. The large fragments from CNBr cleavage are digested with Lys-C or trypsin to obtain a suitable length for mass spectrometric analysis.

On membrane digestion.

Beyond the above mentioned methods, in which protein digestions are done after dissolving the proteins in different buffers, on-membrane digestion was also reported. Whereas proteinase K is used to cleave exposed soluble domains of integral membrane proteins at high pH[104], Lys-C has also been shown to digest proteins 'on membrane' when using 4M urea[82]. In Wu *et al*, whole brain homogate was digested with proteinase K at high pH leading to the identification of 1610 proteins containing 454 (28%) proteins

have at least one transmembrane domain[104]. In our group's work digestion of PM fractions after percoll sedimention in total resulted in 1685 protein identifications, of which the membrane proteins content was 60% (ref[82]).

When membrane proteins are digested, the peptides are further separated by chromatography prior to MS analysis. Multi-dimensional HPLC can also be applied. The MS instruments are different from laboratory to laboratory, and the algorithm for MS data analysis can be also different. All of the post-membrane fractionation strategies may affect the final data quality. Therefore, it is difficult to base comparison of different purification and solubilization methods for membrane proteomics on the number of identified proteins alone. The proportion of membrane proteins should give a more direct evaluation of the fractionation protocol. However, care still needs to be taken because different transmembrane prediction algorithm or different database annotations may give different membrane proportions.

1.3 Mass spectrometry based proteomics

Proteomics is a collective term for large-scale approaches to protein science[105]. Modern proteomic methods include mass spectrometry, protein microarrays, large-scale two-hybrid analyses, high-throughput protein production and crystallization[105]. Since the development of two soft ionization methods, electrospray ionization (ESI)[106] and matrix-assisted laser desorption/ionization (MALDI)[107], MS has gradually become the most popular platform to study protein expression, post-translational modifications and interactions. Recent progresses of MS instrumentation and data analysis software, together with different quantitation methods, not only enable close-to-complete proteome measurements, but also acquire high-content quantitative information about biological samples of enormous complexity[108].

1.3.1 General workflow of MS-based bottom-up proteomics

Proteomics is able to handle different sample mixture or different sub-fractions of interesting sample in batch. Both proteins and peptides can be analyzed directly by MS, which is named top-down and bottom-up proteomics, respectively. Top-down proteomics (reviewed in[109]) measures the molecular weight of intact proteins, and therefore can in principle provide complete information of post-translational modifications. It is especially suitable to analyze the proteins with all PTMs in a certain state or PTM dynamics during different cell states[109]. With increasing molecular weight, top down analysis becomes more and more difficult. It requires more complex instrumentation and expertise when the molecular weight is greater than 20 KDa. Only a few groups in the world have reported identification of more than a few proteins by top-down in one study [109]. Furthermore, automated hardware and software dedicated to top-down approaches are currently in an underdeveloped state[109], which makes high throughput and routine analysis impossible. Compared to top-down, bottom-up proteomics identifies proteins regardless of the intact mass of proteins. Digested peptides can be easily ionized, and much more sensitively detected with modern MS instrumentation. In fact, bottom-up

requires 1–2 orders of magnitude less material than top-down[109]. Bottom-up proteomics is much more competitive in analyzing complex biological mixture.

A typical workflow of MS-based bottom-up proteomics is depicted in Fig. 5. Different tissue samples, cell cultures or body fluids can be analyzed. After tissue homogenization or cell harvest, the crude mixture normally needs to be purified in order to remove contaminating nucleic acids, fat, detergent etc, or subfractionated in order to increase the dynamic range of identification. Isolated protein samples are cleaved by endoproteinase, typically by trypsin or lys-C. After digestion, the peptide mixture is desalted, for example by C_{18} reverse phase StageTips[110]. To increase the dynamic range, peptides are loaded on an HPLC system, and the separated peptides from chromatography are directly sprayed into the MS instrument for data acquisition. The typical flow rate of the HPLC is 200 – 500 nl/min. Raw data files containing the information of ion current intensity and MS and MS/MS spectra are analyzed by bioinformatics software to extract MS and MS/MS lists. The files containing the lists are searched against the in silico predicted spectra from sequence databases by search engines such as MASCOT or SEQUEST. Proper modifications, cleavage enzyme, and maximum mass deviation for precursor and fragmentation ions should be defined before the search. By setting some confidence criteria, a list of interesting proteins can be obtained. The protein list is not the end product of proteomics. With a suitable experiment setup, the inventory may contain information about protein subcellular localization, function in signaling pathways, or protein-protein, protein-nucleic acid interactions, which then needs to be validated by other biochemical approaches.

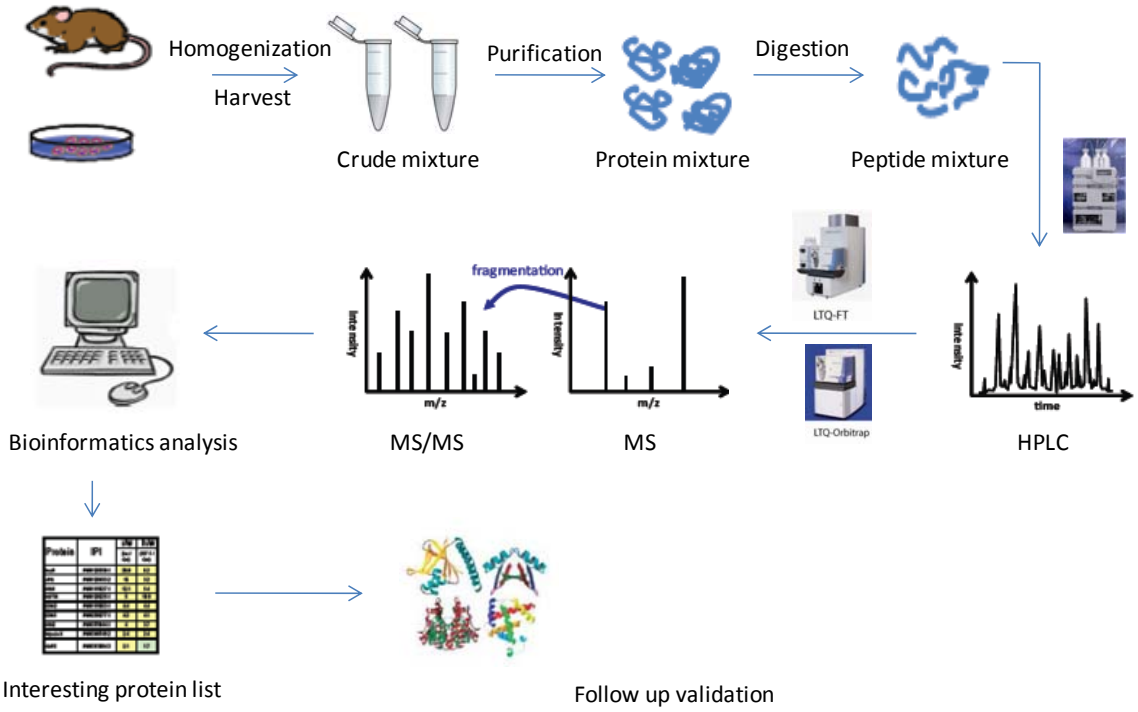


Figure 5. General workflow of MS-based bottom-up proteomics. Samples from tissue or cell culture are homogenized and fractionated for endoprotease digestion. Peptides are separated by on-line HPLC and analyzed by mass spectrometry. The MS and MS/MS lists extracted from MS raw files are searched against protein sequence databases. By setting certain criteria depending on the experiment setup, a list of confident protein identifications is obtained. The functions of interesting protein need to be validated by other biochemical methods.

1.3.2 MS instruments

A mass spectrometer is an instrument that measures the mass-to-charge ratio (m/z) of gas phase ions. Every mass spectrometer nowadays consists of an ion source, at least one mass analyzer and detector, and the data system. Since only ions in gas phase can be analyzed in a MS, protein or peptide ions have to be evaporated in the ion source, typically by one of two soft ionization methods: electrospray ionization (ESI)[106] and matrix-assisted laser desorption/ionization (MALDI)[107]. Ions are separated in the mass analyzer according to either their momentum in a magnetic sector[111], kinetic energy in electrostatic sector instruments[112], velocity in time-of-flight instruments[113], path stability

in linear quadrupoles[114] and quadrupole ion traps[115], and frequency in and ion cyclotron resonance (ICR) mass spectrometer[116] as well as Orbitrap mass spectrometer[117]. Finally, ions are registered under high vacuum conditions in detectors and signals are converted to a readable or graphic display by the data system. Often the electron signals are multiplied via a secondary electron multiplier.

Standard parameters to evaluate a mass spectrometer include resolution, mass accuracy, mass range or upper mass limit, and ion dynamic range[118].

1.3.2.1 Ionization

Although the first MS was constructed by J.J. Thomson in 1912 (ref [119]), the introduction of MS to biological research was not successful until late 1980s, when the two most popular soft ionization methods ESI and MALDI were developed.

Electrospray Ionization (ESI)

The idea of using electrospray dispersion to produce gas phase ions from solution was first introduced by Dole and colleagues in 1968 and later applied to large biomolecules by Fenn and his co-workers in 1989, with the discovery that large molecules produce multiple charged ions in the electrospray ion source[106]. This technique was further improved by Mann and coworkers by introducing nanoelectrospray ionization (nanoESI) compatible with nano-flow rates for minute amount of samples[120], which highly increase the detection sensitivity.

By using electrospray ionization, the analyte solution is placed or pumped inside a fine capillary or needle, to which either a positive or negative high voltage is applied to produce an electric field. The direction of the electric field gradient is dependent on the analytes with the ion potential decreasing along the spray direction. Anions and cations are separated on the surface of the liquid and charges accumulate at the end of the needle. As a result, the liquid protrudes from the needle tip in what is known as a “Taylor cone”

(Fig. 6)[121]. When the Coulombic repulsion between the accumulated charges is equal to the surface tension of the liquid, tiny droplets that contain the excess positive or negative charges detach from the tip and move towards the opposite lens. As the droplets move, more and more solvent evaporates and gradually the droplets become so small that each one contains only a single solute molecule (according to the Dole 'charged residue' mechanism[106]).

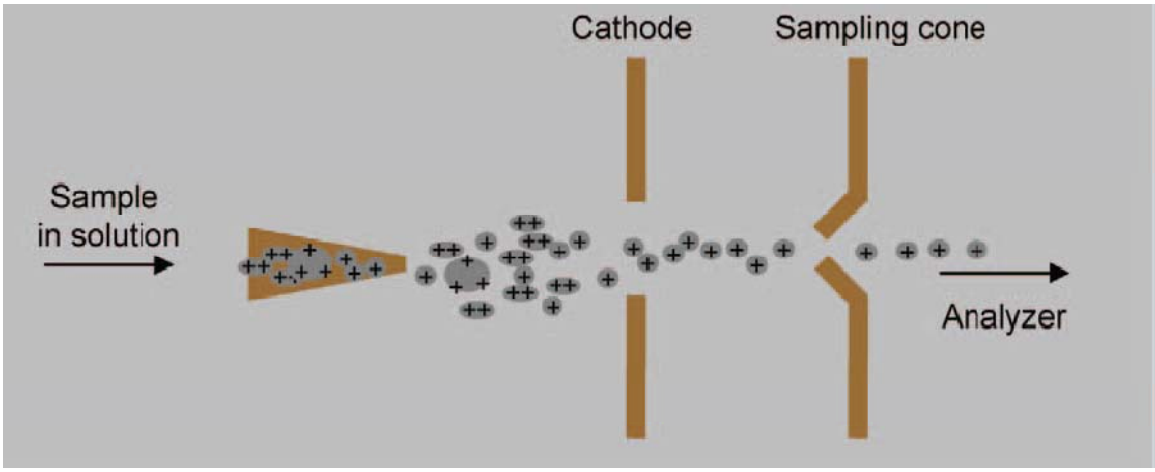


Figure 6. Schematic presentation of the electrospray process for positive charged ions. High potential applied between the needle and the lens forces the highly charged droplets to detach from the tip of the needle. As the droplets travel to the lens, analytes become ionized due to the evaporation of the solvent. The picture is from Wiśniewski[122].

Matrix Assisted Laser Desorption/Ionization (MALDI).

Lasers have been used to generate ions in mass spectrometers since the early 1960s[123]. Early studies of biomolecules with lasers had a mass limitation at about 200 Da, because direct irradiation by the laser is destructive to the thermolabile analytes. A breakthrough for laser ionization came when Karas and Hillenkamp reported the use of matrix assisted ultraviolet laser desorption for non-volatile compounds in 1987 (ref[107]), with which it is possible to analyze large, nonvolatile biomolecules, such as peptides, proteins, oligonucleotides, and oligosaccharides.

In MALDI, the analytes are dissolved in high organic solvent together with matrix, and matrix-analyte co-crystallization occurs on the target, where analyte molecules are

embedded throughout the matrix so that they are completely isolated from one another (Fig. 7). One mechanism to explain the ionization mechanism is that, when crystals are irradiated by the laser, the matrix is excited and expanded into the gas phase, entraining intact analyte in the expanding matrix plume (Fig. 7)[124]. Little internal energy is transferred to the analyte molecules and they may even be cooled during the expansion process. Ionization reactions can occur at any time during this process. Although no unified model exists, it is believed that analyte ions are formed by either proton or electron transfer from or to matrix ions. The reactions maybe very complex and matrix ions may form as both primary ions and secondary ions[124].

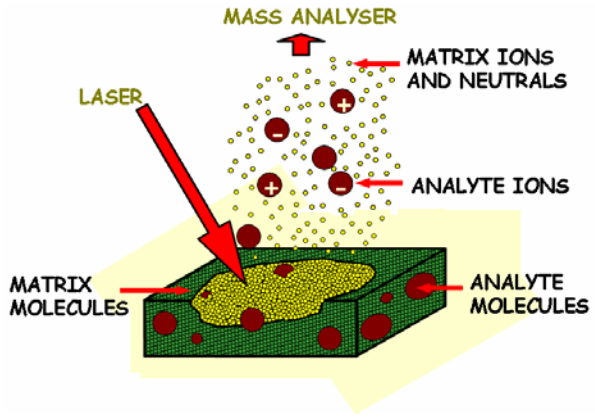


Figure 7. Schematic overview of MALDI source. Analytes are embedded evenly in matrix crystal and entrained into matrix plume when irradiated by laser.

1.3.2.2 Linear ion trap (LTQ)

The basic construction of a commercial linear ion trap, the LTQ, is depicted in Fig. 8. The rods of the quadrupole structure have hyperbolic profiles and are cut into three axial sections, with the length of the central section around three times that of the other two sections. Applying the proper direct current (DC) to three sections allows trapping of the ions along the axis in the central section of the device. Radial trapping of the ions is achieved by applying two phases of the primary radio frequency (RF) voltage to the rod

pairs and ion isolation, activation and ejection is achieved by applying two phases of supplemental alternating current (AC) voltage across the X rods[125].

The LTQ can perform MS analysis alone or work together with other analyzer as a hybrid mass spectrometer. Compared to the 3D Paul trap, LTQ has the advantages of increased ion storage capacity and improved trapping efficiency. Compared to the FT-ICR (Fourier transform ion cyclotron resonance) and the orbitrap described below, it has high sensitivity and fast acquisition, typically 3,000-5,000 ions are enough for LTQ to get good signal within a few milliseconds, while for FT-ICR and orbitrap, normally 1,000,000 ions are necessary and the scanning time is nearly 1 second for each spectrum. But the sensitivity and fast acquisition comes at the expense of low resolution and less accuracy.

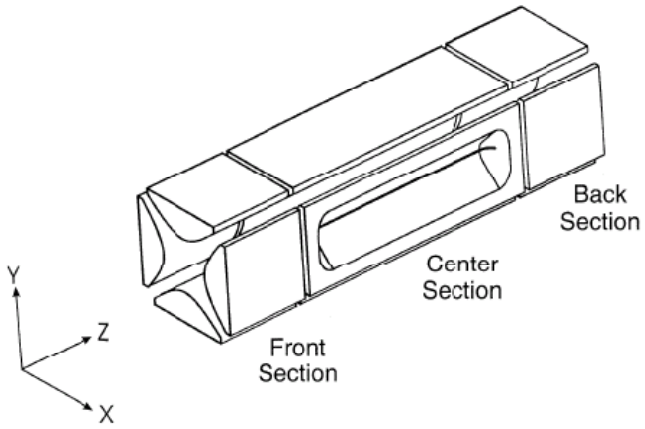


Figure 8. Basic design of the two-dimensional linear ion trap. From Schwartz J.C. *et al*[125].

1.3.2.3 Fourier transform ion cyclotron resonance (FT-ICR)

The fundamental principle of FT-ICR is the motion of a charged particle in a spatially uniform magnetic field[126]. An ion injected in the plate perpendicular to the direction of magnetic field without any collision with a velocity v, will move in a circle of a radius

$$r = \frac{mv}{qB} \qquad (1)$$

at the frequency of

$$f = \frac{qB}{2\pi m} \quad (2)$$

where B is the magnetic field strength, q is the ion charge, m is the ion mass. A remarkable feature in the above equations is that the ICR frequency *f* is independent of their velocity. Therefore, ions of a given mass-to-charge ratio have the same ICR frequency regardless of their initial energy. The insensitivity of the cyclotron frequency to the kinetic energy of an ion is one of the fundamental reasons why the FTMS instrument is able to achieve ultra-high resolution[127].

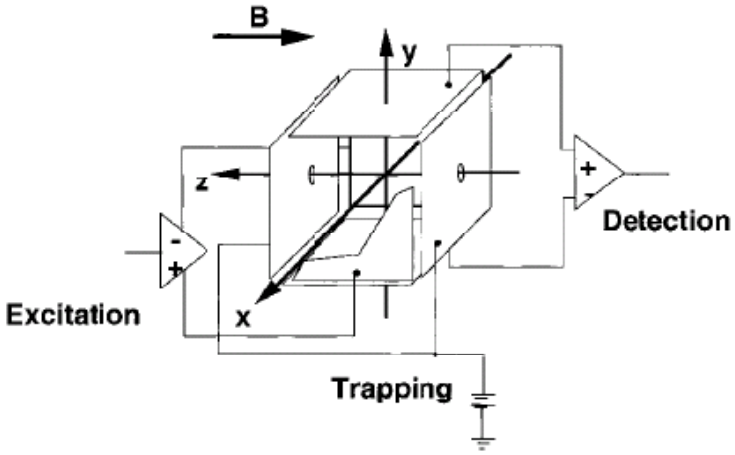


Figure 9. Electronic circuit for ion excitation [126].

Single ion excitation

An ion's initial room temperature thermal cyclotron radius shown as equation (2) is typically on the order of 100 μm, which is too small to be observed[127]. A spatially uniform electrostatic field oscillating sinusoidally at the same frequency of ions of interest is applied perpendicular to the magnetic field to excite ions, along the x-axis (Fig. 9). The post-excitation ion cyclotron radius (r) and kinetic energy (KE) can be calculated by

$$r = \frac{V_{p-p} T_{excite}}{2B} \quad (3)$$

$$KE = \frac{q^2 v_{p-p}^2 (T_{excite})^2}{8d^2 m} \quad (4)$$

Where $V_{p\text{-}p}$ is the peak-to-peak voltage difference between the two plates along x-direction, and d is the distance between two plates, T_{excite} is the oscillating resonant excitation duration[126]. Equation (3) shows that the post-excitation ion cyclotron orbital radius is independent of m/z. Thus, all ions of a given m/z range can be excited to the same ICR orbital radius by application of an RF electric field whose magnitude is constant with frequency. In contrast, equation (4) shows that, for a given excitation electric field amplitude and duration, post-excitation ion kinetic energy is independent of magnetic field strength B. Combing equation (3), we can see that, if we want to excite ions to a give radius, the post-excitation energy increases proportional to the square of B.

Broadband excitation

Broadband excitation in FTICR is performed in SWIFT mode, where SWIFT stands for "stored waveform inverse Fourier transform". SWIFT waveforms start by defining the mass-domain of desired excitation profile, and then convert it to a frequency-domain spectrum, by performing an inverse Fourier transform to generate the desired time-domain excitation waveform[126]. A combination of ion excitation and ion ejection can be achieved simultaneously by SWIFT waveforms.

Ion detection

All ions of the same m/z are excited coherently, and undergo cyclotron motion as a packet. When the ion packet passes the two detection plates alternatively, a sinusoidal image signal is produced, which can be amplified, digitized and stored for processing by a computer. Ions of many masses can be detected simultaneously with FTMS. The maximum resolution that can achieved for a data set is

$$R = \frac{fT}{2} \quad (5)$$

Where R is resolving power, f is the cyclotron frequency and T is the duration of a transient[127].

FTICR MS is the instrument that can provide the best mass accuracy and highest mass resolution so far. By using a 14.5 magnet, it is possible to achieve external calibration broadband mass accuracy less than 300 ppb rms, and a resolving power of 200,000 at m/z 400 (ref [128]).

1.3.2.4 Orbitrap

The orbitrap is an ion trap that traps ions in an electrostatic field, instead of RF or magnet fields[129]. The trap consists of an outer barrel-like electrode and a central spindle-like electrode along the axis (Fig. 10). The outer electrode is split at the middle allowing ions to be injected into the trap. A DC voltage is applied between the outer and inner electrodes.

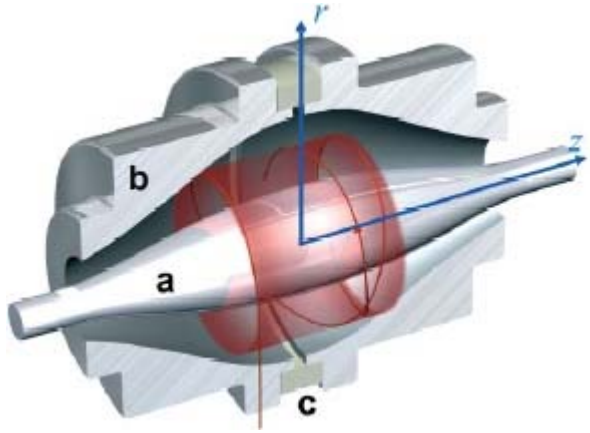


Figure 10. Cut-away structure of the Orbitrap mass analyzer. From Scigelova *et al*[129].

Both electrodes are specially shaped, and the axially symmetric electrodes create a combined "quadro-logarithmic" electrostatic potential:

$$\mathrm{U}(r,z) = \frac{k}{2}\left(z^2 - \frac{r^2}{2}\right) + \frac{k}{2}\cdot(R_m)^2 \cdot \ln\left[\frac{r}{R_m}\right] + \mathrm{C} \qquad (6)$$

Where r and z are cylindrical coordinates, k is a constant, and R_m is the characteristic radius [117]. Stable ion trajectories involve both an orbiting motion around the central electrode and simultaneous oscillations in the z-direction. The potential in the z-direction is exclusively quadratic. Ion mass/charge ratio m/z is simply related to the frequency of ion oscillation along the z-axis

$$\omega = \sqrt{(\mathrm{z/m}) \cdot k} \qquad (7)$$

This axial frequency is used for ion detection, because it is completely independent of energy and of the spatial spread of the ions[117]. This feature in analogy to FTICR, is a fundamental reason for ultra-high resolution.

The orbitrap is another instrument, besides FTICR MS, that can provide very high mass accuracy and mass resolution. By applying a 'lock mass' for internal calibration, sub-ppm (rms) mass accuracy were achieved[130]. High-mass resolution up to 150,000 for ions produced by laser ablation has been demonstrated[131].

1.3.2.5 Hybrid mass spectrometers

As mentioned above, the LTQ analyzer has the advantages of high sensitivity and fast scanning, while FT-ICR and orbitrap have unparalleled high mass accuracy and resolution. Thermo Fisher, a manufacturer of mass spectrometers, combined the advantages of both to create the hybrid LTQ-FT and LTQ-Orbitrap, which are the instruments on which all experiments in the study were performed.

The configuration of the LTQ-FT and LTQ-Orbitrap is shown in Fig. 11. For normal shotgun proteomics analysis, survey scans are acquired in either FT or orbitrap, while MS/MS spectra are acquired in the LTQ. For some special purposes, e.g. post-translational modifications analysis or de-novo sequencing, MS/MS can be also acquired in the FT or orbitrap to assure high quality data.

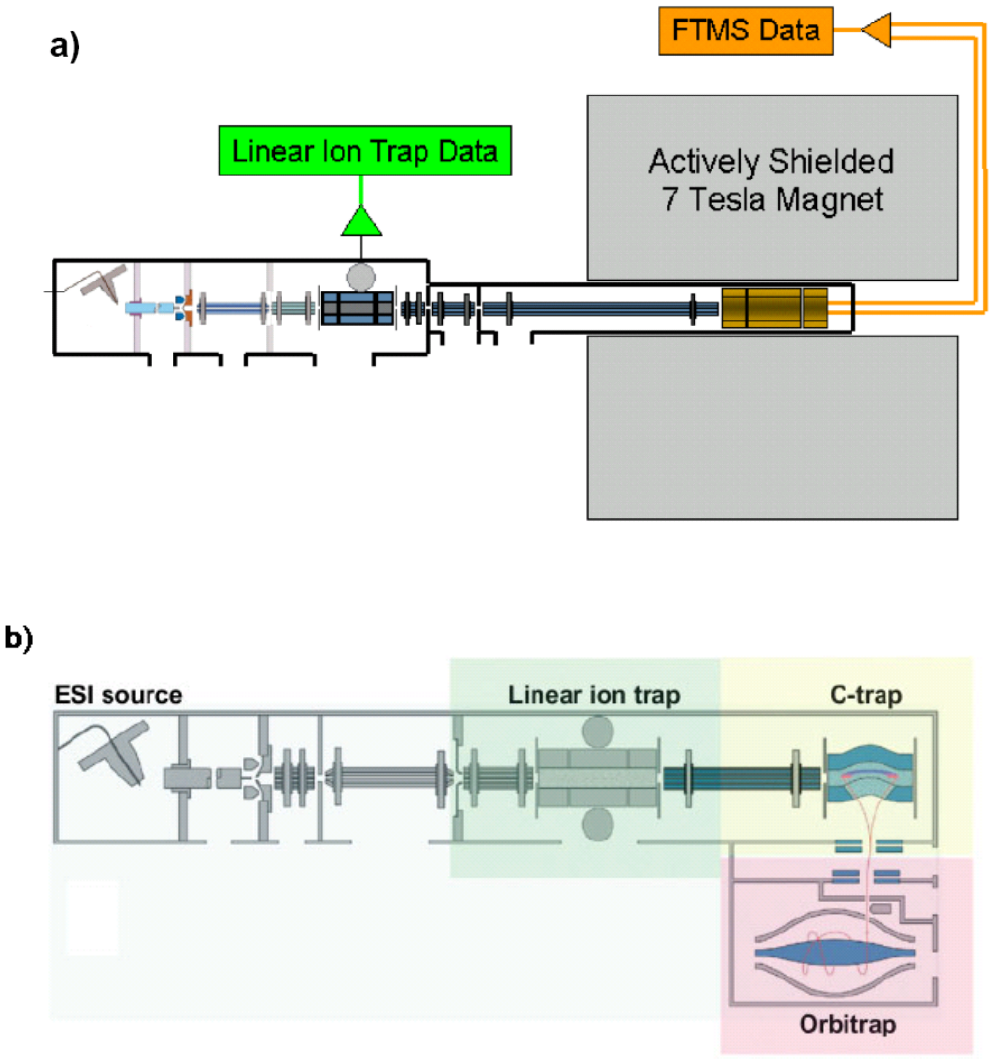


Figure 11. A schematic of the hybrid instruments: a) LTQ-FT (from Schrader *et al*[132]), b) LTQ-Orbitrap (from Scigelova *et al*[129]).

1.3.2.6 Peptide fragmentation

Nomenclature

The nomenclature of peptide fragment ions adopted nowadays is suggested by Roepstorff and Fohlman in 1984 (ref [133]) and modified by Biemann in 1988 (ref [134]). There are three types of fragment ions generated by the cleavage of either of the bonds Cα-C, C-N, N-Cα

in the peptide main chain. According to this nomenclature, the fragment ions are labeled a_n, b_n, c_n when a positive charge is kept by the N-terminal side and x_n, y_n, z_n when the positive charge is kept by the C-terminal side (Fig. 12). There are two extra hydrogen atoms transferred to c_n and y_n ions, one responsible for the protonation and the other one originating from the other side of the peptide, but not for a_n, b_n, x_n, z_n ions. The subscript, n, indicates the number of amino acid residues contained in the fragment ion.

There is a marked difference between the fragmentation observed at high and low CID energy[124]. At low energy, the observed fragments are mostly b_n and y_n. At high energy, all the possible fragments can be generated. Besides that, multiple cleavages occur. There are two types of informative multiple cleavage fragments. One is the immonium ion, which results from multiple cleavage of the peptidic chain and appears among the low masses in the spectrum. They yield information concerning the amino acid composition of the peptide. The other type results from the cleavage of the peptidic chain and amino acid lateral chain. There are three observed types of fragments, termed d_n, w_n, v_n respectively, as shown in the middle part of Figure 15. d_n and w_n result from the cleavage of the bond between the β and γ carbon atoms of the side chain of the C-terminal amino acid or of an N-terminal amino acid of z_n respectively, and are useful to distinguish the isomers Leu and Ile. v_n results from the complete loss of the side chain of the N-terminal amino acid residue of y_n.

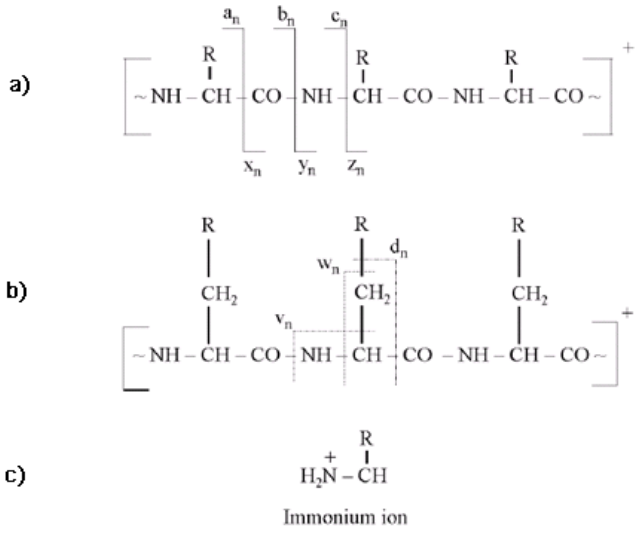


Figure 12. Illustration of nomenclature for peptide fragment ions. a) fragments results from the cleavage of a bond in the peptide chain; b) fragments result from double cleavage of both the peptidic chain and amino acid lateral chain; c) the structure of immonium ions. The nomenclature was proposed by Roepstorff and Fohlman in 1984 (ref [133]) and modified by Bieman in 1988 (ref [134]).

Collision-induced dissociation (CID)

In traditional metastable analysis using magnetic sector instruments, the ions leaving the source can be classified into three categories: (1) stable ions, with a lifetime greater than 10^{-6} s, reach the detector before any fragmentation has occurred; (2) unstable ions, with a lifetime smaller than 10^{-7} s, fragment before leaving the source; (3) metastable ions, with an intermediate lifetime, are stable enough to be selected by the first analyzer and fragment before they reach the second analyzer. CID converts part of the kinetic energy of ions into internal energy by collision with inertial gas, and thus shorten the lifetime of precursor ion from type (1) to type (3).

By applying the energy and momentum conservations in the collision process, the maximum energy fraction that can be converted into internal energy is given by the following equation[124]:

$$E_{cm} = E_{lab} \frac{M_t}{M_i + M_t} \qquad (8)$$

Where M_i is the ion mass, M_t is the collision gas mass, E_{lab} is the ion kinetic energy and E_{cm} is the maximum energy fraction converted into internal energy.

Two collision regimes should be distinguished: low-energy collision with collision energy between 1 and 100 eV, occurs in quadrupole or ion trap instrument; and high-energy collision with ~keV collision energy, occurs in electromagnetic or TOF instrument. The fragmentation patterns observed for low and high collision energy is different. Low energy CID produces fragmentation at the peptide bonds, whereas high energy CID also gives rise to peptide side-chain cleavage.

High-energy C-trap Dissociation (HCD)

HCD is a fragmentation technique that is special to the LTQ-Orbitrap instrument[135]. As shown in Fig. 11b, the C-trap is normally used to store ions on their way from the ion trap to high-resolution analysis in the orbitrap. After applying RF voltage, ions can be fragmented inside the C-trap and further transfered to the orbitrap for acquisition. Compared to the normal CID fragmentation acquired in the LTQ, it has the advantage of high mass accuracy and high resolution. Furthermore, it contains the fragmentation in the low-mass region, which is less than one third of the precursor ion and is normally missing in CID MS/MS spectra obtained in an ion trap. Higher energy is applied compared to CID, therefore, HCD produces immonium ions, which can be used as a diagnostic peak for certain amino acids or modification. This technique is especially useful for post-translational modifications or de-novo sequencing analysis. For example, trimethylation and acetylation with a mass difference of 0.0364 Da can be differentiated by HCD.

1.3.3 Quantitative Proteomics

Proteins are the functional units of biological processes. Cells organize different cellular events not by just turning protein expression on and off, but by precise regulation of protein levels in space and time. Taking tissue differentiation as an example, only a small fraction of all proteins are thought to be expressed in only one tissue[136]. Therefore, protein identification is often not sufficient to decipher different biological phenomenon, whereas quantitative proteomics can yield function from relative protein levels. .

Unlike other proteomics methods which utilize dyes, fluorophores or radioactivity for quantitation, MS-based proteomics quantitation is based on labeling, which can be either metabolic labeling or chemical labeling (Fig. 13). Recent developments of MS data analysis software make label free quantitation possible. Absolute quantitation can also be achieved by spiking standard peptides of known amount. For reviews see[137, 138].

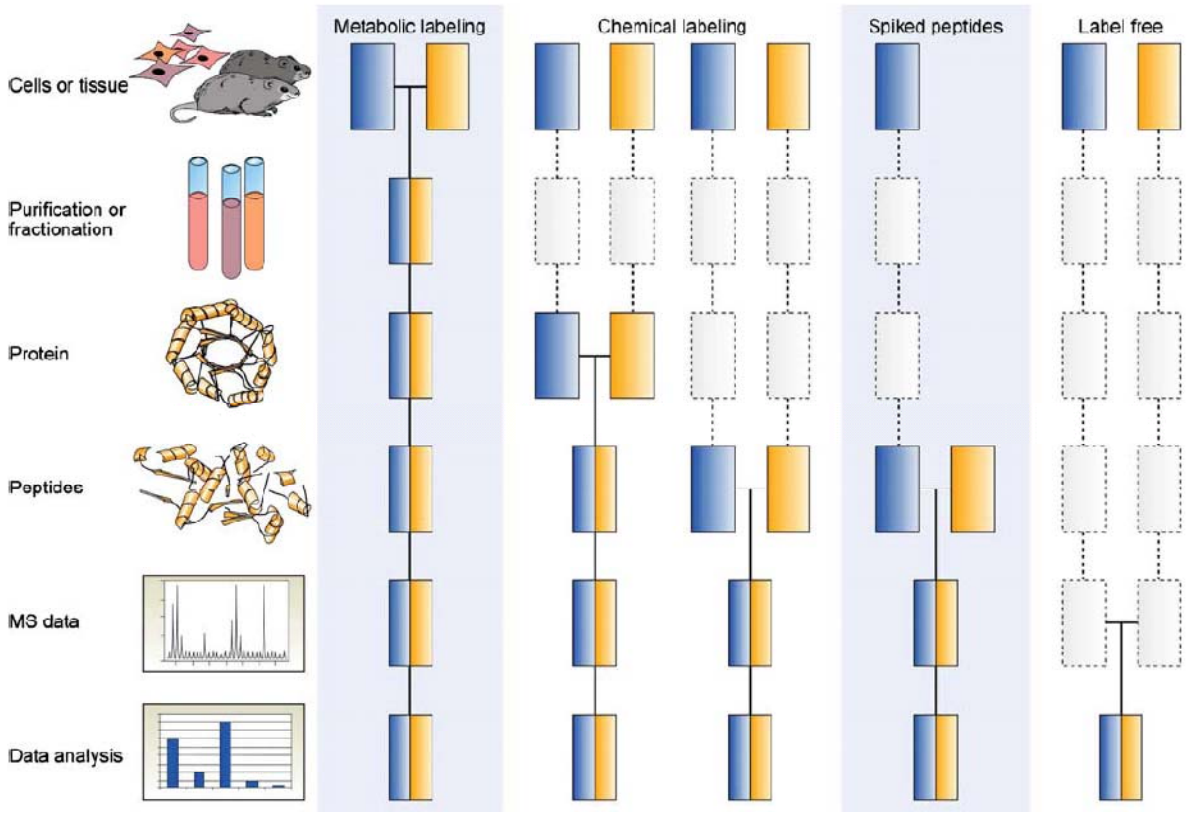


Figure 13. Common quantitative mass spectrometry workflows. Boxes in blue and yellow represent two experimental conditions. Horizontal lines indicate when samples are combined. Dashed lines indicate points at which experimental variation and thus quantification errors can occur. From Bantscheff M. *et al*[138].

Metabolic labeling

Metabolic labeling introduces a stable isotope signature into proteins during cell growth or division, therefore, it allows mixing different samples at the level of intact cells, which eliminates the errors introduced during differential sample preparation. This type of labeling was initially described for total labeling of bacteria using ^{15}N-enriched cell culture medium[139], however, the uncertain number of N in different peptides make quantitation extremely difficulty. The introduction of the stable isotope labeling by amino acid in cell culture (SILAC)[140] made metabolic labeling very popular and the most precise method for quantitation measurement[138].

The amino acids chosen for SILAC labeling should be essential for the cell line. Lysine and arginine are the two most common targets for labeling, which together with the popular trypsin digestion for proteomics, ensures every peptide except the C-terminal one can in principle be used for quantitation. SILAC labeling utilizes arginine and lysine with heavy elements of ^{13}C, ^{15}N, and ^{2}H. The most commonly used forms are $^{13}C_6$-Arg, $^{13}C_6{}^{15}N_4$-Arg, $^{2}H_4$-Lys and $^{13}C_6{}^{15}N_2$-Lys. Up to three different biological conditions can be directly compared in a single SILAC experiment (Fig. 14) and the number of different conditions for comparison can be extended indirectly[141].

The quantitation in SILAC experiments is normally done by comparison of the extracted ion current (XIC), which is the integrated value of a mass-specific signal in the LC MS data[137]. One advantage of SILAC labeling with ^{13}C atoms is that the light- and heavy-labeled peptides co-eluted very well[137], which minimizes artifacts due to pairing wrong peptides.

SILAC based quantitation has proven to be powerful for many interesting biological discoveries[142]. The only limitation is that it is not feasible for *in vivo* tissue sample analysis of large mammals such as humans.

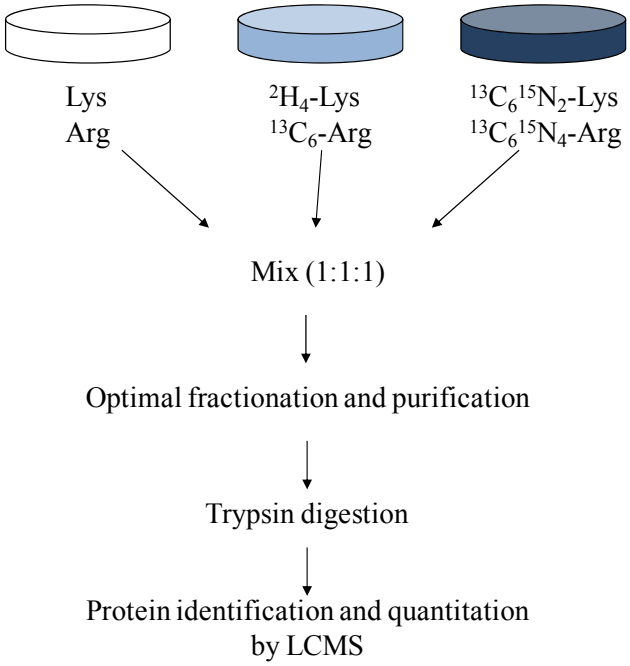


Figure 14. General triple SILAC experimental flowchart.

Chemical labeling

Chemical labeling can be performed either on the intact protein level or at the peptide level. Reported methods for chemical labeling protein or peptide are quite diverse (for reviews see[137, 138]), including isotope-coded affinity tags (ICAT)[143], HysTag[144] or isotope tags for relative and absolute quantification (iTRAQ)[145]. ICAT is the first chemical labeling method that was widely used. The reagent consists of three elements: an affinity tag (biotin), which is used to isolate ICAT-labeled peptides; a linker that can incorporate stable isotopes; and a reactive group with specificity toward thiol groups (cysteines). Therefore, labeling by ICAT can be used both for affinity purification and quantitation. The same is true for the HysTag, which consists of four functional elements: i) an affinity ligand (His6-tag), ii) a tryptic cleavage site (-Arg-Ala-), iii) an Ala-9 residue that contains four (d4) or zero (d0) deuterium atoms, and iv) a thiol-reactive group. iTRAQ is a multiplexed set of isobaric reagents that modify the amino terminus or the side chain of lysines of each peptide. The complete molecule consists of a reporter group, a mass balance group which makes the overall mass of reporter and balance group constant, and a peptide-reactive group. Differently modified peptides have the same m/z in the MS spectra but in MS/MS spectra, the unique low-mass reporter ion of each tag (113-121 Da) from different status is available for quantitation. iTRAQ does not provide enrichment, however, it has the powerful advantage that up to 8 different samples can be directly compared in a single experiment[146]. The disadvantages of all the chemical labeling approaches compared with SILAC are that the samples are combined at a relative late state, and that the labeling procedure may introduce manual errors.

Label free quantitation

Label free quantitation compares the XIC value of the same peptide in different chromatography runs, based on their mass and retention time[147, 148]. The straightforward experimental protocol and the fact that in principle any samples can be compared to each other make label free quantitation attractive[138]. For a reliable comparison, label-free quantitation requires robust and reproducible sample-preparation and analytical

systems[122]. The sample preparation steps should be minimized, and all samples should be prepared in parallel, injected into the MS identically one after another to reduce possible errors. The temperature of the chromatography column should be controlled carefully to make the elution time for the same peptide identical. High mass accuracy and high mass resolution MS instruments and reliable data analysis software are also necessary to obtain reliable results[137]. Elution time alignment and intensity normalization may be necessary to improve the data accuracy.

Absolute quantitation

All three above mentioned approaches quantify proteins relatively in different samples. In principle, if we change one of the samples to a known amount of isotope-labeled standard which resembles the protein/peptides of protein to be quantified, it is possible to obtain the absolute amount of the protein of interest. Several methods are now available for MS-based absolute quantitation. Culture-derived isotope tags (CDITs)[149] quantify tissue samples taking the advantage of SILAC by mixing labeled cells with brain samples. Synthetic unlabeld peptides were spiked into the cell mixture for absolutely quantifying brain samples indirectly. In another method, termed Absolute Quantitation (AQUA)[150], synthetic peptides incorporated with stable isotopes are spiked into samples as an internal standard. Post-translational modification can also be incorporated, thereby, modification induced ionization differences can be eliminated. Both CDITs and AQUA utilize synthetic peptides as an internal standard for absolute quantitation. Peptides should be synthesized specially for every interesting protein. While these methods can accurate quantify the peptide amount after digestion, the sample loss during the preparation steps before digestion is in principle not known. In a third method QCAT[151], a tryptic and His tag containing Q peptides is constructed to be a concatemer of one peptide from every interesting proteins. A gene coding for the Q peptide is insert to *E. coli* by vector. After cloning in ^{15}N medium, the Q peptide is purified and combined with samples as a standard. By using QCAT, peptides do not need to be synthesized individually for every protein and the ratio for all peptides is strictly 1:1 (at least before digestion). The Q

peptides can be spiked into samples earlier than the other two methods, which could also reduce systematic errors.

2 Aim of this study

The projects involved in this thesis were conducted with the aim to improve technologies for (1) individual protein characterization (2) global membrane protein identification.

For a comprehensive analysis of post-translational modifications of a specific protein, 100% sequence coverage is in principle a pre-requisite to cover all the possibilities, which remains a challenging task. In the first project, we analyzed peptides by a chip implementation of nanoelectrospray, coupled to an LTQ-Orbitrap, with the goal to obtain high sequence coverage at low protein concentration. In the second project, this method was applied to characterize different linker histone 1 variants and to identified and quantify the occupancy ratio of different methylation sites.

Modern proteomics is able to identify several thousand proteins in single experiment, however, the identified membrane proteins are typically underrepresented compared to the protein sequence database. Detergents are an indispensible tool to dissolve membrane proteins, however, they ruin MS analysis even at minute amounts. In the third project we set up a detergent-based but gel-free protocol for membrane protein purification and we further combined the method with gel filtration chromatography in the fourth project to perform comparative membrane proteomics analysis for rat cerebellum, spinal cord and sciatic nerve.

3 Peptide mapping by static nanoelectrospray

3.1 Publication: Nanoelectrospray peptide mapping revisited: Composite survey spectra allow high dynamic range protein characterization without LCMS on an orbitrap mass spectrometer

This paper presents the results for peptide mapping by using an electrostatic spray robot called TriVersa coupled with LTQ-Orbitrap. The following pages contain the paper published on International Journal of Mass Spectrometry 268 (2007): 158-167.

Available online at www.sciencedirect.com
ScienceDirect
International Journal of Mass Spectrometry 268 (2007) 158–167
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

Nanoelectrospray peptide mapping revisited: Composite survey spectra allow high dynamic range protein characterization without LCMS on an orbitrap mass spectrometer

Aiping Lu [a], Leonie F. Waanders [a], Reinaldo Almeida [b], Guoqing Li [a], Mark Allen [b], Jürgen Cox [a], Jesper V. Olsen [a], Tiziana Bonaldi [a], Matthias Mann [a,*]

[a] *Department for Proteomics and Signal Transduction, Max-Planck-Institute for Biochemistry, Am Klopferspitz 18, D-82131 Martinsried, Germany*
[b] *Advion Biosciences Limited, Rowan House, 26 Queens Road, Norfolk NR9 3DB, UK*

Received 19 February 2007; received in revised form 9 May 2007; accepted 14 May 2007
Available online 21 May 2007

Abstract

Mass spectrometric (MS) determination of the primary structure of proteins, including post-translational modifications, remains a challenging task. Proteins are usually digested to tryptic peptides that are measured either by MALDI peptide mapping or by liquid chromatography online coupled to tandem MS (LC–MS/MS). Here we instead analyze peptides by a chip implementation of nanoelectrospray (TriVersa Nanomate, Advion Biosciences), coupled to a linear ion-trap–orbitrap hybrid instrument (LTQ-Orbitrap, Thermo Fisher). 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 times of more than 30 min/μl 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 6000-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 purified 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.

Keywords: LTQ-Orbitrap; Peptide mass mapping; Peptide sequencing; Dynamic range; Protein modification

1. Introduction

During the last few years, efforts in mass spectrometry-based proteomics [1] have concentrated on the qualitative and quantitative analysis of complex protein mixtures. 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) and MALDI-TOF/TOF are popular methods 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

* Corresponding author. Tel.: +49 89 8578 2778; fax: +49 89 8578 2219.
E-mail address: mmann@biochem.mpg.de (M. Mann).

1387-3806/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2007.05.006

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 s.

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].

2. Experimental

2.1. Sample preparation for 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 DTT (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 4× 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% (v/v) of absolute TFA. BSA peptides were desalted and stored on RP-C_{18} StageTip columns [12] and eluted right before mass spectrometric analysis with 50% methanol/0.5% formic acid.

2.2. Histone H3 sample preparation

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 proteins (HMG) 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 h 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% ACN and directly loaded onto a reverse phase HPLC column (Jupiter C_{18}, 250 × 4.60, 5 μm, 300 Å) (Phenomenex) connected to an Aekta LC-system (Amersham). Individual histones were separated by applying a gradient from 20% to 80% ACN 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-C_{18} StageTip columns, while the other half was desalted and stored using SCX (Strong Cation Exchange) StageTip columns. Peptides on the RP-C_{18} 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.

2.3. Mass spectrometric analysis

All experiments were performed using a linear ion-trap–orbitrap hybrid mass spectrometer (LTQ-Orbitrap, Thermo Fisher Scientific, Bremen, Germany) with a TriVersa Nanomate (Advion Biosciences, Ithaca, USA) 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 (i.d. 2.5 μm) from Advion, providing a flow rate of 20 nl/min. At this flow rate, 1 μl 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 s, longer than the acquisition time per two overlapping segments.

2.4. Data analysis

The Mascot engine was used for mass spectrometry data identification (Matrix Science, London, UK). BSA peaks were searched in IPI_Human_v313 to which the BSA sequence had

been added, using 5 ppm maximum mass deviation (MMD [13]) for precursor ions, 0.5 Da MMD for fragment ions, 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 ion, 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.

3. Results and discussion

3.1. 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 (Fig. 1). As can be seen in the figure, 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 population 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 0.25–1 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 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.

3.2. 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 (Fig. 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 Fig. 2A, were chosen so that the accu-

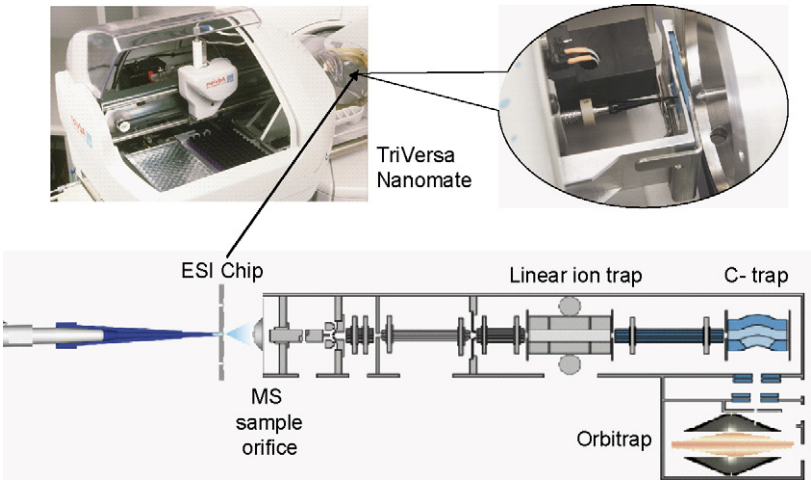


Fig. 1. Schematics of the TriVersa Nanomate coupled to the LTQ-Orbitrap. Samples were applied by coated tips to the nozzle of the electrospray chip in front of the TriVersa instrument. The low flow rate chip (i.d. 2.5 μm) provided a stable flow rate of 20 nl/min, which is in the 'true' nanoelectrospray range.

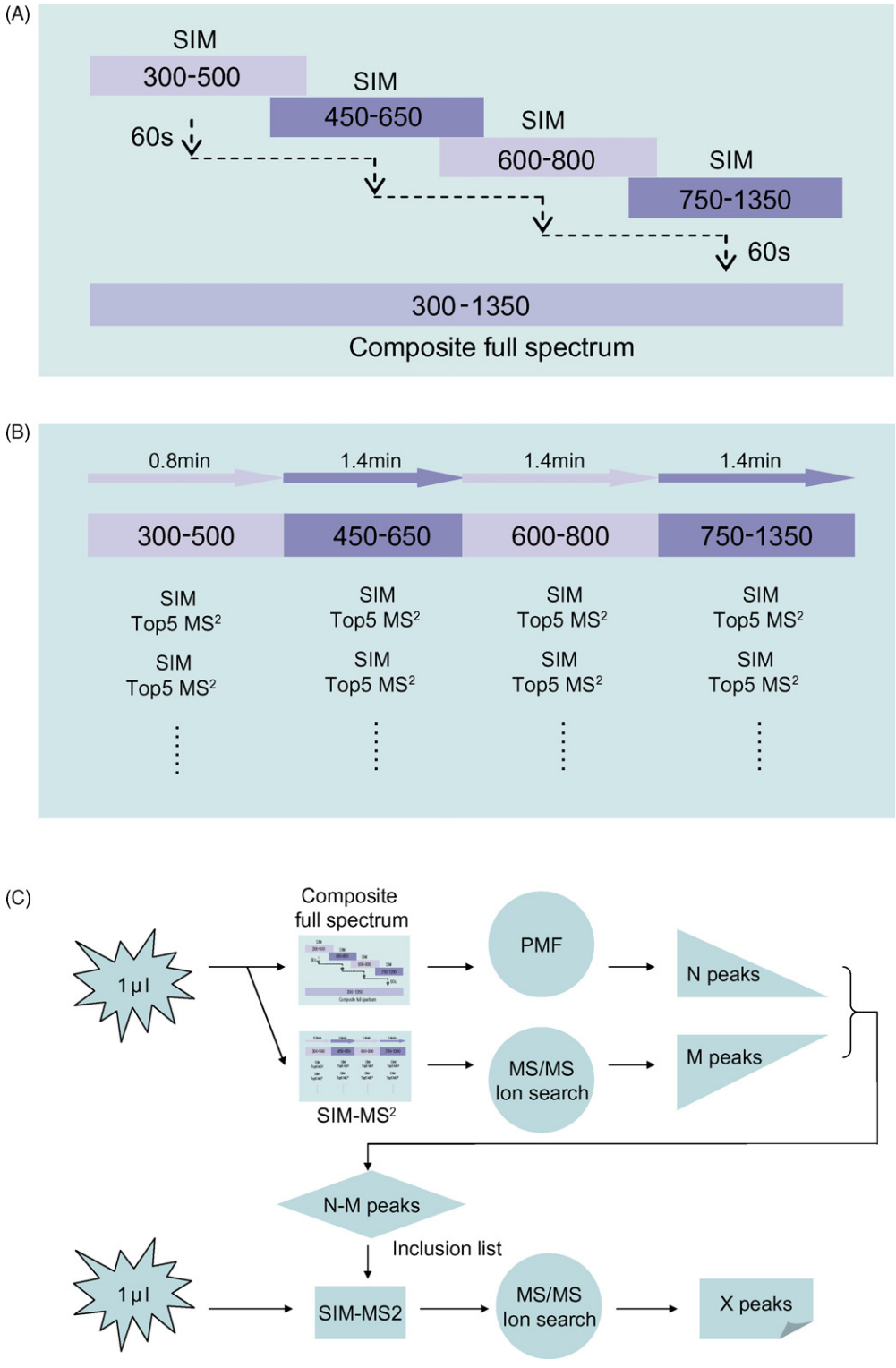


Fig. 2. Schematic description of the acquisition strategy as performed for protein characterization of tryptic digested BSA. (A) Selected ion monitoring (SIM) scans of multiple mass segments were repeatedly analyzed in the orbitrap mass spectrometer and these multiple SIM acquisitions were combined in 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^2 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^2 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 2 μl of very diluted sample solution.

Table 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^2 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 concentrations	Acquisition methods	Sequence identified		
		per Individual Run (%)	per Acquisition Method (%)	Overall (%)
25 fmol/μl	Composite SIM	87	88	88
		79		
		81		
	SIM–MS^2	63	75	
		73		
		72		
	Inclusion list MS^2	1.4	1.4	
5 fmol/μl	Composite SIM	73	82	83
		77		
		78		
	SIM–MS^2	54	65	
		60		
		60		
	Inclusion list MS^2	10.1	10.1	
500 amol/μl	Composite SIM	45	60	66
		42		
		49		
		24	35	
	SIM–MS^2	20		
		24		
	Inclusion list MS^2	10.5	10.5	

mulation 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 100 times.

First identification is based on peptide mass fingerprint (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^2 cycle is repeated for a time adapted to the number of MS/MS candidates (Fig. 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. Fig. 2C presents an overview of the three-step sequence.

Since the acquisition methods for composite full spectra and SIM–MS^2 take only 4 and 5 min, respectively, 1 μl 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.

3.3. 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 Fig. 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 1). As shown in the table, the inclusion list SIM–MS^2 method turned out to be particularly advantageous for lower protein concentrations.

3.4. 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

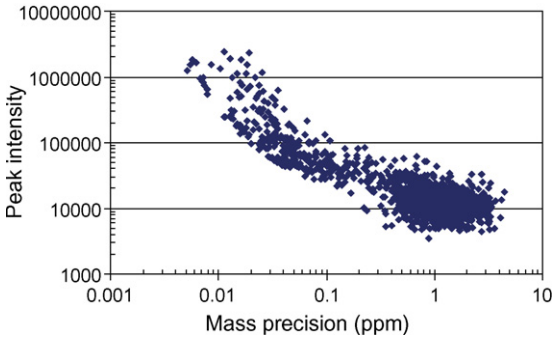


Fig. 3. Mass precision correlated to ion intensity. Plot of the intensity of different ions extracted from the composite spectrum vs. their precision, from thousands of scans depicted on a double logarithmic scale. Note that for most peaks we achieve sub ppm mass precision.

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, by repeated measurement of the same spectrum for 100 times, the standard deviation of the mass value should be decreased by a factor of 10. In Fig. 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, a large percentage of the peptides have a precision well below 1 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.

3.5. More than 6000-fold dynamic range in the composite full spectra

Summing up multiple spectra filters out noise but boosts low intense ions that cannot 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 regions with low intense signals are accumulated for a longer time in order to reach the same specified injection target value.

As depicted in Fig. 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, 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 (Fig. 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.

3.6. Sequence coverage comparable to LC–MS

We compared the sequence coverage obtained after 5 min of SIM–MS^2 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 (i.d. 5 μm) was used for nanoelectrospray, resulting in a flow rate of 200 nl/min. The LC–MS run took in total 53 min, of which the actual gradient lasted for 28 min. Both methods were performed twice. The sequence coverage obtained by these two methods was very comparable, 83.7% for nanoelectrospray SIM–MS^2 and 78.5% for LC–MS/MS. Detailed identification information for each tryptic peptide in the BSA sequence is shown in Fig. 5. Most of the peaks were identified by both methods, 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^2 run.

3.7. 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 post-translational modifications (PTMs) of selected amino acids, such as lysine acetylation, lysine and arginine methylation, serine phosphorylation and attachment of ubiquitin [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 mass spectrometric 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. Furthermore, fragmentation was performed by higher energy injection into the C-trap [19], which 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^2 mode were searched in

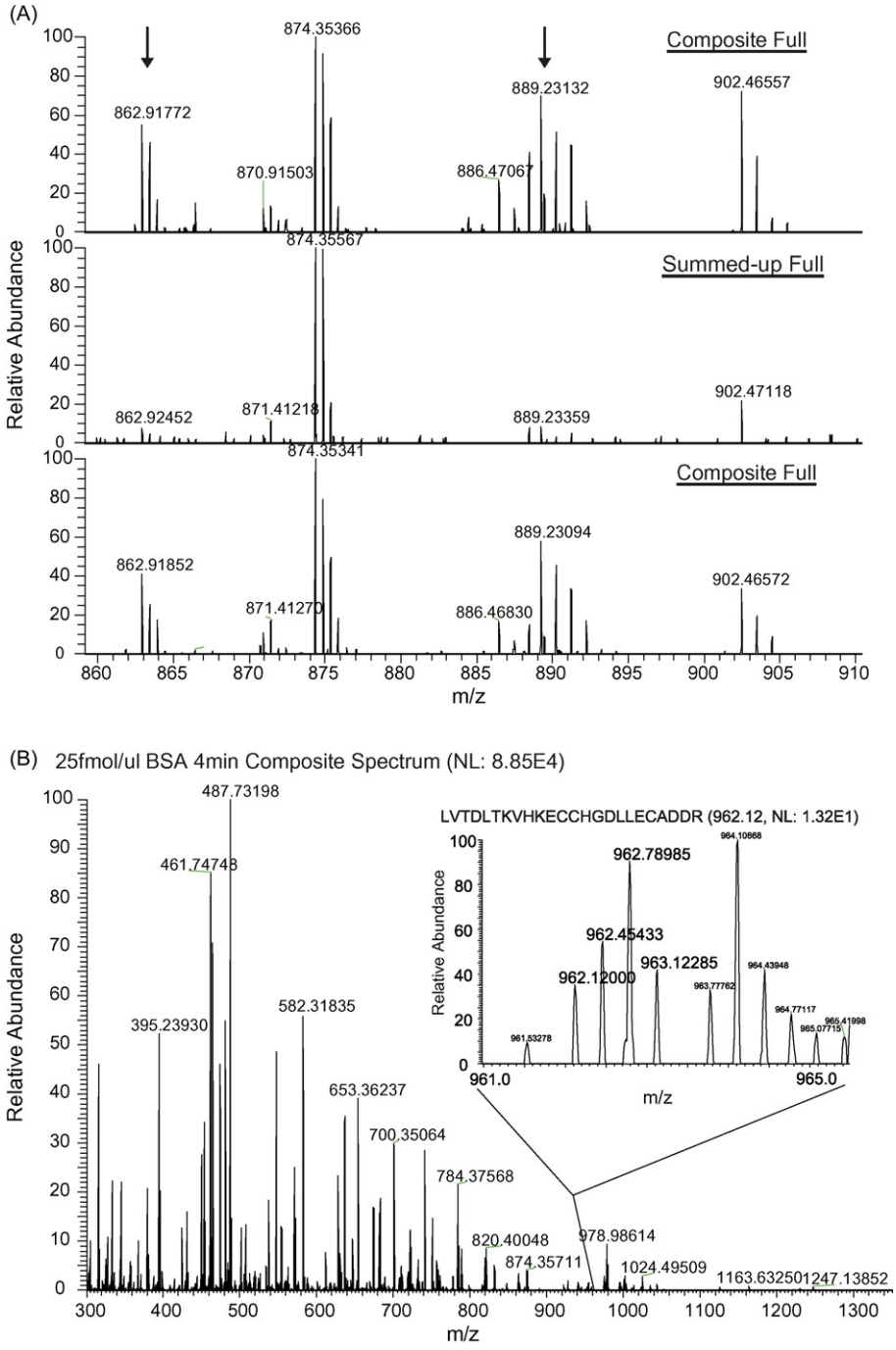


Fig. 4. Advantages of composite spectra. (A) In the composite spectrum ions of low intensity were boosted in comparison to full spectra summed-up for the same time span. 500 attomole/μl BSA sample was sprayed, and the composite spectra and the summed-up full spectra were acquired directly after one another. 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 6000-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 8.85E4, 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.

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 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, the trimethylated and acetylated peptide KSTGGKAPR, the modified sites could not unambiguously be assigned to the sequence since the fragmentation was performed in the LTQ and the mass difference between these modifications (0.0364 Da) is far less than what the LTQ can distinguish. Therefore, a second

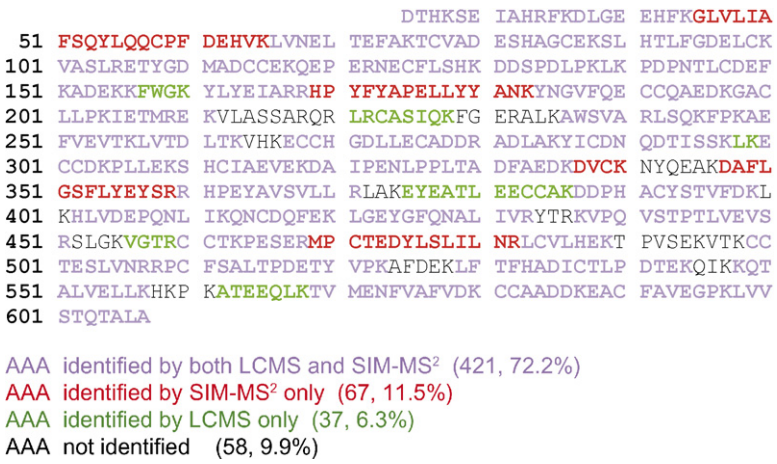


Fig. 5. Comparison of the BSA sequence coverage obtained in SIM–MS^2 vs. LC–MS/MS mode. In both experiments 50 fmol of BSA solution were used and both measurements were performed with the same instrument settings. Peptides in blue are identified in both methods, red peptides only in SIM–MS^2 and green peptides only in LC–MS mode. Black peptides were not identified.

Table 2
Histone H3 modifications (combinations) identified by the SIM–MS^2 method

Modified amino acids	Sequence
K4 monomethylation	TKQTAR
K9 dimethylation + K14 acetylation K9 trimethylation + K14 acetylation	KSTGGKAPR
K23 acetylation K18 acetylation + K23 acetylation	KQLATKAAR
K79 monomethylation K79 dimethylation	EIAQDFKTDLR

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, 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).

3.8. Relative quantitation by deconvoluted peak intensity

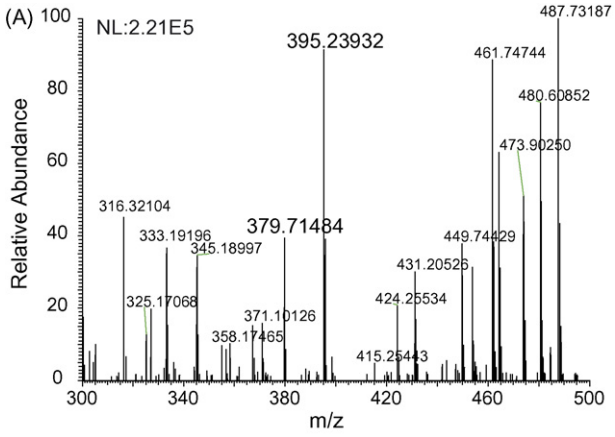
As mentioned above, quantification of peptides and proteins is becoming more and more important. An advantage of the acquisition of a large number of spectra is the increasing precision, not only of the mass value but also of the intensity ratio between ions. This will be illustrated with the ratio between two different BSA peptides as well as with the ratio between a methylated and non-methylated histone peptide.

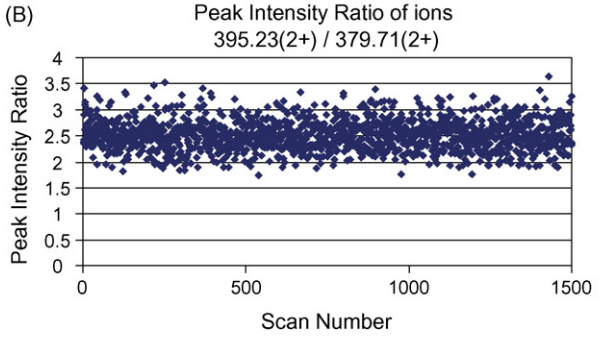
Fig. 6 illustrates the relative quantitation of BSA peptides. Panel A shows the *m*/*z* segment 300–500. Fig. 6B reveals that the ratio of the relative intensities of the peptides with *m*/*z* 395.239 (2+) and 379.715 (2+) varied per spectrum between 1.7 and 3.5. However, with increasing number of accumulated spectra quantitation becomes more and more precise. As shown in Fig. 6C, the 99% confidence interval for quantitation decreases from 13% after accumulating 10 scans to 0.9% after accumulation of 1500 scans (15 min acquisition).

In the case of histone H3 we investigated quantitation of the normal peptide against a slightly modified form. This is possible in nanoelectrospray since both peptides are present in the same scan. In order to avoid inaccuracy due to transmission 'edge effects' in the SIM windows, we chose to quantify based on the full spectrum. The relative quantitation values for identified histone H3 peptide pairs are listed in Table 3. The amounts of several methylated peptides were about 10-fold less than those of the unmodified peptides. Note that this value gives a general idea

Table 3
Quantitation values of identified peptide pairs with the same sequence but different modifications

Identified peptide pairs	Quantitation value with 99% confidence interval
TKQTAR/TK$_{(methyl)}$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$_{(methyl)}$TDLR/	11.391 ± 0.868
EIAQDFK$_{(methyl)}$TDLR/EIAQDFK$_{(dimethyl)}$TDLR	8.206 ± 0.990





(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

Fig. 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. Per single scan the ratio varies from 1.5 to 3.5 (B). With an increasing number of accumulated spectra, the precision of the quantitation ratio increases, as shown in (C).

of the absolute stoichiometry of this methylation site but that it needs to be corrected for the different ionization efficiencies of the modified versus the unmodified peptides [21].

4. Conclusions and perspectives

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, nanoelectrospray 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 offers 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 (orbitrap and LTQ), 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 Da 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^2 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

The authors thank other members of the department for Proteomics and Signal Transduction and of Advion Biosciences for constructive comments and discussion. This work was partially supported by the EU research directorate in the 'Interaction Proteome' grant (LSHG-CT-2003-505520) and in the HEROIC grant (LSHG-CT-2005-018883).

References

[1] R. Aebersold, M. Mann, Nature 422 (2003) 198.
[2] M. Mann, O.N. Jensen, Nat. Biotechnol. 21 (2003) 255.
[3] O.N. Jensen, Nat. Rev. Mol. Cell. Biol. 7 (2006) 391.
[4] M. Kussmann, P. Roepstorff, Meth. Mol. Biol. 146 (2000) 405.
[5] J. Yates, A. McCormack, D. Schieltz, E. Carmack, A. Link, J. Protein Chem. 16 (1997) 495.
[6] J. Peng, S.P. Gygi, J. Mass Spectrom. 36 (2001) 1083.
[7] S.E. Ong, M. Mann, Nat. Chem. Biol. 1 (2005) 252.

[8] H. Steen, M. Mann, Nat. Rev. Mol. Cell. Biol. 5 (2004) 699.
[9] M. Wilm, M. Mann, Anal. Chem. 68 (1996) 1.
[10] M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, M. Mann, Nature 379 (1996) 466.
[11] A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat, S. Horning, Anal. Chem. 78 (2006) 2113.
[12] J. Rappsilber, Y. Ishihama, M. Mann, Anal. Chem. 75 (2003) 663.
[13] R. Zubarev, M. Mann, Mol. Cell. Proteomics 6 (2007) 377.
[14] J.V. Olsen, L.M. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, M. Mann, Mol. Cell. Proteomics 4 (2005) 2010.
[15] H.T. Spotswood, B.M. Turner, J. Clin. Invest. 110 (2002) 577.
[16] T. Kouzarides, Cell 128 (2007) 693.
[17] T. Jenuwein, C.D. Allis, Science 293 (2001) 1074.
[18] B.A. Garcia, S.B. Hake, R.L. Diaz, M. Kauer, S.A. Morris, J. Recht, J. Shabanowitz, N. Mishra, B.D. Strahl, C.D. Allis, D.F. Hunt, J. Biol. Chem. 282 (2007) 7641.
[19] J.V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, M. Mann, Nat. Meth., in press.
[20] C.E. Thomas, N.L. Kelleher, C.A. Mizzen, J. Proteome Res. 5 (2006) 240.
[21] H. Steen, J.A. Jebanathirajah, J. Rush, N. Morrice, M.W. Kirschner, Mol. Cell. Proteomics 5 (2006) 172.

4 Protocol for membrane protein purification

4.1 Publication: Detergent-based but Gel-free Method Allows Identification of Several Hundred Membrane Proteins in Single MS Runs

This paper presents the results of the project on optimize a detergent based but gel free protocol to purify membrane proteins from tissue sample. The project is a joint effort together with Mr. Nagarjuna Nagaraj in our group. The article has been published by Journal of Proteome Research. The following pages contain the version of the article available on the web at

http://pubs.acs.org/cgi-bin/asap.cgi/jprobs/asap/html/pr800412j.html

.

research articles Journal of proteome research

Detergent-Based but Gel-Free Method Allows Identification of Several Hundred Membrane Proteins in Single LC-MS Runs

Nagarjuna Nagaraj,[#] Aiping Lu,[#] Matthias Mann,* and Jacek R. Wiśniewski*

Department of Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, Martinsried, Germany

Received June 6, 2008

Detergents are indispensable solubilizing agents in the purification and analysis of membrane proteins. For mass spectrometric identification of proteins, it is essential that detergents are removed prior to analysis, necessitating an in-gel digestion step. Here, we report a procedure that allows use of detergents and in-solution digestion of proteins. Crude membrane preparations from mouse brain were solubilized with Triton X-100, CHAPS, or SDS, and the detergents were depleted from the membrane proteins using a desalting column equilibrated with 8 M urea. Following digestion with endoproteinase Lys-C, the resulting peptides were analyzed by LC-MS/MS on Linear ion trap-Orbitrap instrument. Applying stringent identification criteria, in single-LC-MS-runs, 1059 $\pm$ 108 proteins, including 797 $\pm$ 43 membrane proteins, were mapped from mouse brain. The identified proteins represented a broad spectrum of neurotransmitter receptors and other ion channels. The general applicability of the method is demonstrated by profiling of membrane proteins from four other mouse organs. Single-run analyses of eye, liver, spleen, and skeletal muscle allowed identification of 522 $\pm$ 9, 610 $\pm$ 7, 777 $\pm$ 8, and 307 $\pm$ 7 membrane proteins. Our results demonstrate that membrane proteins can be analyzed as efficiently as soluble proteins.

Keywords: Membrane proteomics • integral membrane proteins • detergent removal • brain • liver • eye • spleen • muscle • LTQ-Orbitrap

Introduction

The use of detergents in biochemical research ranges from standard procedures, such as SDS-PAGE or pull-down experiments, to complex, specialized applications, such as extraction of integral membrane receptors consisting of multiple subunits. In the field of membrane biochemistry, detergents are indispensable tools for solubilization and fractionation of membrane proteins. However, detergents, even in small concentrations, dominate mass spectra and preclude peptide or protein analysis. Thus, in mass spectrometry (MS)-based proteomics, detergents have to be efficiently and thoroughly removed from proteins or peptides prior to analysis, but this is not an easy task. Different methods have been described for separation of proteins from detergents including gel filtration, ion-exchange and hydrophobic adsorption chromatography, density gradient centrifugation, dialysis, ultra filtration, phase partition, and precipitation (for a review see ref 1). However, they have not become popular in mass spectrometry because of their inability to completely remove the detergents.[2] Moreover, these methods can lead to substantial sample losses as they have been designed to deal with relatively high protein amounts; thus, their applicability to proteomics is limited.

To circumvent the difficulties with removal of detergents, alternative approaches avoiding the use of detergents have been proposed. For example, 60% methanol[3,4] was used for the solubilization of membranes and the extracted proteins were digested with trypsin. In another approach, membranes were solubilized with 90% formic acid and the proteins were chemically cleaved with cyanogen bromide.[5] In addition, digestion of membrane proteins directly in a suspension of fractions enriched in membranes has been described. Wu et al. used proteinase K at high pH to digest protein chains protruding from the membrane bilayer.[6] Using a related concept, we analyzed mouse brain membrane proteins by digesting purified plasma membranes in 4 M urea with endoproteinase Lys-C.[7] We further applied this 'solid-phase digestion' strategy in protein profiling[8] and comparative, semiquantitative mapping of plasma membrane proteins between distinct regions of mouse brain.[9,10]

Despite these developments, detergents are preferred due to their strength in membrane solubilization and are widely used in sample preparation for subsequent proteomic analysis. Unfortunately, so far the only method to efficiently remove detergents once they were introduced involved in-gel digestion after SDS-PAGE or, alternatively, incorporation of detergent-containing protein lysates into a polyacrylamide matrix without electrophoresis.[11]

In this work, we present a novel procedure for detergent removal and digestion of membrane proteins, and compare it with an in-gel protein immobilization and digestion procedure.

* Correspondence address: Jacek R. Wiśniewski and Matthias Mann, Department of Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried near Munich, Germany, E-mail, (J.R.W.) jwisniew@biochem.mpg.de, (M.M.) mmann@biochem.mpg.de; fax, +49 89 8578 2219.

[#] These authors contributed equally.

10.1021/pr800412j CCC: $40.75

Published on Web 10/08/2008

We show that membrane proteins solubilized with Triton X-100, CHAPS, or SDS can be efficiently separated from the detergents by 'desalting' on a column equilibrated with 8 M urea. In this protocol, the proteins are digested with endoproteinase Lys-C and the resulting peptides are bound to a C_{18} membrane and analyzed by LC-MS/MS. Our method results in an almost 2-fold increased protein identification in comparison to the in-gel based approach. We demonstrate that our procedure is useful for profiling of membrane proteins from various tissues including mouse brain, liver, spleen, eye, and muscle tissues.

Materials and Methods

Membrane Preparation and Protein Solubilization. Frozen mouse brain, liver, spleen, leg muscle and eye were purchased from Pel-freez Biologicals, Rogers, AR. Membrane preparation was carried out as described previously.[8] Briefly, 20 mg of tissue was homogenized in 1 mL of high salt buffer (2 M NaCl, 10 mM HEPES-NaOH, pH 7.4, 1 mM EDTA) using IKA Ultra Turbax blender at maximum speed for 20 s. The suspension was centrifuged at 16 000*g*, at 4 °C for 15 min. The resulting pellet was re-extracted twice with carbonate buffer (0.1 M Na_2CO_3, pH 11.3, 1 mM EDTA) as above. After incubation for 30 min, the pellet was washed with urea buffer (4 M urea, 100 mM NaCl, 10 mM HEPES/NaOH, pH 7.4, and 1 mM EDTA). Following urea wash, the pellet was solubilized in 50 μL of 100 mM sodium phosphate buffer, pH 7.0, containing either 2% (w/v) SDS, 0.5% (v/v) Triton X-100, or 3.5% (w/v) CHAPS.

Removal of Detergents and Protein Digestion with Endoproteinase Lys-C. Detergents were removed on a HiTrap desalting column (5 mL, Amersham Biosciences, Uppsala, Sweden). The column was equilibrated with 8 M urea, 25 mM Tris-HCl, pH 8.0. The proteins were digested with 0.5 μg of endoproteinase Lys-C from Wako (Richmond, VA) at 25 °C overnight. Digestion was terminated by addition of 1% (v/v) trifluoroacetic acid. The digested peptide mixture was purified and stored in C_{18} StageTips as described.[12] Usually, 5% of the digestion mixture was loaded on a StageTip containing two membrane plugs.

In-Gel Digestion. Detergent solubilized membrane preparations were mixed with sample buffer, loaded on NuPAGE 4−12% Bis-Tris gel (Invitrogen, Carlsbad, CA), and stacked in the gel by electrophoresis at 50 V for 15 min. The gel was stained with Coomassie blue staining kit (Invitrogen), and entire lanes (usually 0.5 cm in length) were excised and in-gel digested as described.[13,14]

Mass Spectrometric Analysis. Protein digests were analyzed by online capillary LC-MS/MS. The LC-MS/MS setup was similar to that described before.[15] Briefly, samples were separated on an in-house made 15 cm reversed phase capillary emitter column (inner diameter 75 μm, 3 μm ReproSil-Pur C18-AQ media (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany)) using 120 min gradients and an Agilent 1100 nanoflow system (Agilent Technologies, Palo Alto, CA) or 90 min gradients using a Proxeon EASY-nLC (Proxeon Biosystems, Odense, Denmark). The LC setup was connected to an LTQ-Orbitrap (Thermo Fisher, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems). The mass spectrometers were operated in data-dependent mode. Survey MS scans were acquired in the orbitrap with the resolution set to a value of 60 000. Up to 5 most intense ions per scan were fragmented and analyzed in the linear ion trap. For accurate mass measurements, the lock-mass option was employed.[16]

Peak List Generation, Database Searching and Validation. The raw files were processed with MaxQuant, an in-house developed software suite (version 1.0.6.3).[17,18] The peak list files were searched against decoy IPI-mouse database version 3.24 containing both forward and reversed protein sequences, by the MASCOT search engine.[19] The initial parent and fragment ion maximum mass deviation[20] were set to 7 ppm and 0.5 Th, respectively. The search included variable modifications of oxidation of methionine and protein N-terminal acetylation. Peptides with at least seven amino acids were considered for identification and proteins with two or more peptides (at least one of them unique to the protein sequence) were considered valid hits. The false discovery rate for both the peptides and proteins were set a threshold value of 0.01. All proteins identified in this study are listed in Supplementary Table 1 in Supporting Information.

Bioinformatics Analysis. Gene ontology analysis of the identified proteins were performed using the Protein Center platform (Proxeon Biosystems, Odense, Denmark) primarily for cellular component analysis and membrane and transmembrane domain annotations.

Results and Discussion

Previously, we have described a detergent-free method for proteomic analysis of membrane proteins.[7] In that method, membranes are directly digested with endoproteinase Lys-C and the released peptides are analyzed using one-[7,9] or two-dimensional LC-MS/MS.[8,10,21] Even though this method is a powerful tool for mapping of membrane proteins, it has some limitations. Proteolytic digestion is performed on only partially denatured proteins (4 M urea); therefore, the yield of peptides is restricted by accessibility of the cleavage sites to proteases. Moreover, this method cannot be combined with chromatography techniques for separation of membrane proteins before digestion. We wished to develop a method that can be coupled with chromatographic separation like size-exclusion chromatography for in-depth analysis of membrane proteome of tissue samples. The use of detergents for extraction of membrane proteins would circumvent the above-mentioned limitations and, in addition, offer the option of stepwise extraction of membrane proteins which potentially can be used for selective protein solubilization and fractionation of membrane proteins.[22−24]

Removal of Detergents. The major goal of this work was to establish a simple, effective, and robust method for removal of detergents from solubilized membranes, such that mass spectrometric analysis would not be affected. For this purpose, crude membrane preparations from mouse brain were extracted with three different detergents including SDS, Triton X-100, and CHAPS, which are representative of anionic, nonionic, and zwitterionic detergents, respectively. The detergents were used in relatively high concentrations, exceeding their critical micellar concentration (CMC) values several-fold. The membrane preparations were solubilized with 3.5% (w/v) CHAPS, 2% (w/v) SDS, and 0.5% (v/v) Triton X-100. Since size exclusion chromatography has been reported to be highly effective in detergent removal,[25] we considered the use of gel filtration in our experiments. To separate detergents from proteins and to dissociate micelles, while keeping membrane proteins in solution, we employed the strongly chaotropic reagent urea at 8 M concentration. In the presence of 8 M urea, micelles dissociate while membrane proteins stay in solution. Importantly, the detergent migrates into the gel filtration matrix, while proteins elute in the void volume. Thus, when

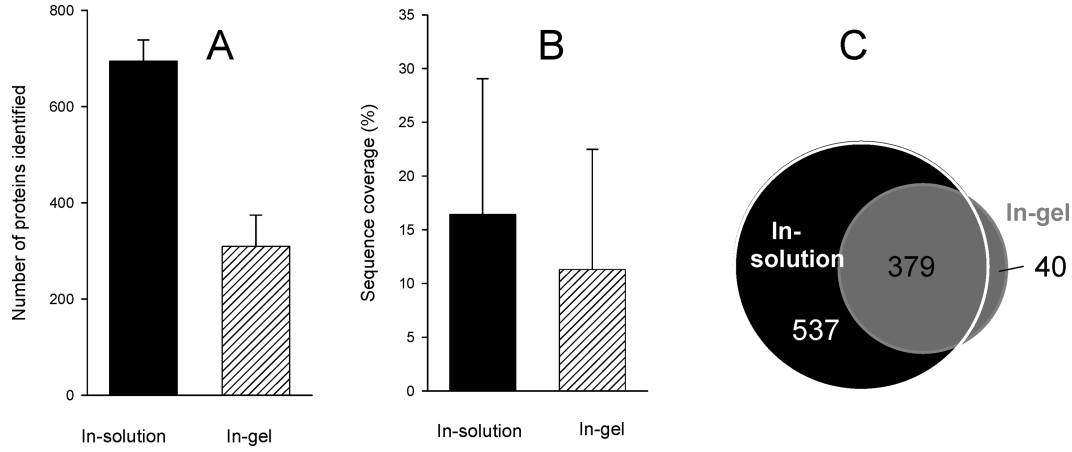


Figure 1. Comparison of membrane protein identification efficiency (A) and sequence coverage (B) using in-solution and in-gel methods. Brain membranes were extracted with 2% SDS. Five aliquots from the same extract were digested either in solution or in-gel. C, Venn diagram comparing the numbers of identifications achieved with both digestion methods.

the gel filtration columns equilibrated with 8 M urea were used, it was possible to efficiently separate proteins from the detergent.

After digestion with Lys-C, and removal of urea on StageTips, the peptides were analyzed by LC-MS/MS. We did not observe any detergent contamination in any of our experiments, demonstrating the efficiency of the depletion. Results were similar for all three tested detergents. Overall, 56.2 ± 2.5% of the identified proteins from the brain membrane preparation contained predicted transmembrane domains. This is an even higher proportion than in our previously reported method for profiling of membrane proteomes where the proteins were digested directly from membranes in a suspension.[8] In that method, about 40% of the identified proteins contained predicted transmembrane segments.[8−10] Note that not all membrane proteins contain transmembrane domains; therefore, the proportion of membrane proteins in our preparations is even higher (nearly 80%; see below).

To demonstrate the efficiency of the protein digestion in urea, we compared it to the in-gel digestion procedure. Five samples for each procedure were prepared using aliquots from the same membrane preparation. We found that the number of proteins identified by the in-solution method was more than twice that identified by the in-gel method (Figure 1A). Furthermore, the sequence coverage of identified proteins was a third higher in the new method (Figure 1B). The total number of membrane proteins identified in the 10 runs was 956. Of those, 537 and 40 proteins were exclusively identified in the in-solution and the in-gel method, respectively (Figure 1C). The proteins that were identified only by the in-solution method do not show any obvious physio-chemical difference compared to proteins identified by in-gel method. These proteins were likely not identified by the in-gel method due to a combination of less efficient peptide extraction by the in-gel method and stochastic 'picking' of peptide peaks for sequencing. However, we observed that we were able to recover more large proteins by the in-solution method. Using our method, we identified 27 proteins consisting of more than 1500 residues, whereas only one protein of this size was found using the in-gel approach. Thus, the results achieved using the new method encompass, rather than complement, the in-gel method.

Comparison to Published Membrane Proteome Analyses. The membrane proteome has been one of the most difficult challenges for 2D gel electrophoresis and usually very few or no transmembrane proteins are reported with that technology. MS-based proteomics has been more successful, but so far, researchers had to reduce protein complexity by applying different protein and peptide fractionation strategies that can facilitate the identification of less abundant proteins. However, multidimensional separation also generates large numbers of fractions that have to be analyzed individually which requires extensive measurement time. For rapid screening of tissue specimens such as clinical biopsy material, high-throughput methods are required. As described above, the combination of the previously developed method for extraction of membranes[8] with the here described detergent-based solubilization of membrane proteins resulted in identification of 530 proteins with predicted transmembrane domains from a mouse brain sample. To assess the relative usefulness of our method, we compared our results from single MS runs to other membrane proteome analyses in the recent literature, which employed extensive fractionation. In terms of identification of membrane proteins and the analysis time required, our method shows significant advantages. For example, a recently reported 3-D-strategy for analysis of membrane proteins allowed identification of only 125 proteins membrane proteins in mouse brain.[26] Other approaches for studying membrane proteins from various sources including *Corynebacterium glutamicum* and human platelet membranes have been carried out. Fischer et al., characterized the membrane proteome of two strains of *C. glutamicum* and reported 326 integral membrane proteins involving multiple fractionation steps and extensive mass spectrometric measuring time.[27] Moebius et al. in their platelets study identified less than 300 proteins with approximately 30% membrane proteins.[28] More recently, improved protocols for analysis of enriched membrane proteins have been published. Analysis of membrane fraction from HeLa cells using a phase transfer surfactant-aided digestion procedure resulted in identification of 764 membrane proteins (53% of total identified proteins) in 12 cumulative LC-MS/MS runs.[29] In another study, methanol was used to improve the efficiency of tryptic digestion which allowed identification of a total of 690 integral membrane proteins in 72 LC-MS/MS runs.[30]

Identification of Brain-Specific Proteins. A single run on the LTQ-Orbitrap instrument identified approximately 800 proteins. More than 70% of these were membrane proteins and 59% had predicted transmembrane domains (Figure 2C). The combination of our sample preparation procedure with high

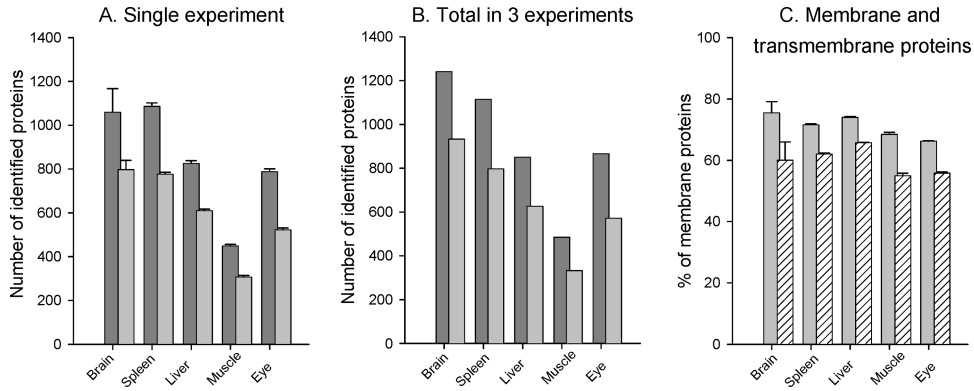


Figure 2. Identification of proteins from membrane preparations of mouse brain, spleen, liver, skeletal muscle, and eye. (A) single experiment; (B) cumulative result from 3 independent experiments; (C) content of membrane annotated proteins and proteins with predictable transmembrane segments. *Dark bars*, total protein number; *light bars*, membrane annotated proteins, *striped bars*, proteins with predictable transmembrane segments.

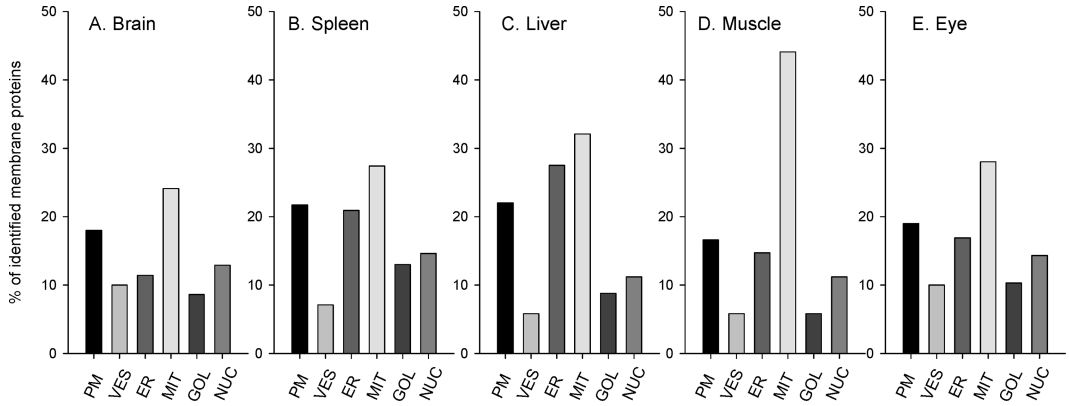


Figure 3. Subcellular localization of membrane proteins from mouse brain (A), spleen (B), liver (C), skeletal muscle (D), and eye (E). *GOL*, Golgi apparatus; *PM*, plasma membrane; *VES*, cytoplasmic vesicles; *ER*, endoplasmatic reticulum; *MIT*, mitochondria, *N*, nucleus.

resolution MS analysis yielded a large number of brain-specific proteins. These include neurotransmitter receptors such as glutamate and GABA receptors and also ion channels such as sodium and potassium channel proteins (Supplementary Table 2 in Supporting Information). The identified glutamate receptors represent ionotropic-AMPA, NMDA, metabotropic, and GluRdelta-2 receptors. Four subunits of $GABA_A$ receptor and two subunits of $GABA_B$ receptors were identified. A complete set of subunits of the voltage-dependent calcium channel was also mapped, including the channel subunit Cacna1a (and its isoform Cacna1e) as well the auxiliary subunits alpha2/delta, beta, and gamma. The channel subunit is a 281 kDa polypeptide with 23 predicted transmembrane helices. Similar in size and number of transmembrane domains is the identified sodium channel protein Scn1a (230 kDa, 23 transmembrane segments). Identification of such large proteins is an important advantage of our method.

Application of the Method to Other Tissues. Having demonstrated the efficiency of our method with mouse brain tissue, we wanted to show its applicability to a wide range of tissues. We selected liver, spleen, eye, and skeletal muscle which represent widely different tissue types. Each sample was analyzed in triplicate (Figure 3). We observed that the number of membrane proteins identified was dependent on the properties of the organs and ranged from 797 ± 43 in the brain to 307 ± 7 in leg muscle (Figure 2A). With respect to the total number of proteins identified, the percentage membrane proteins ranged from 66% in eye to 78% in brain (Figure 2C). The comparatively low number of proteins identified in leg muscle was due to the highly abundant titin, which has a molecular weight of several MDa and takes up most of the sequencing time in the mass spectrometer. Only small differences in the number of identifications were observed between each of the three runs, and therefore, the cumulative number of identified proteins for each tissue was only slightly higher compared to a single run which emphasizes the reproducibility of our method (Figure 2A,B).

Analysis of subcellular location of identified membrane proteins using Gene Ontology revealed distinct origin of membranes, which may reflect abundance of different organelles in the analyzed tissues (Figure 3). Mitochondrial membranes appear to be the most abundant in all analyzed samples ranging from 24% in brain to 44% in skeletal muscle. The high content of mitochondrial proteins in leg muscle reflects the fact that muscles are extremely rich in mitochondria for ATP production. The percentage of proteins annotated as extracellular or cell surface, which are mainly plasma membrane proteins, was very similar between the tissues and at 17−22%. Liver and spleen membranes contained the highest percentage of proteins annotated as endoplasmatic reticulum (Figure 3C,B). The abundance of endoplasmatic reticulum in liver is related to the high level of protein synthesis including major plasma proteins such as albumin. Compared to other tissues, brain and eye have the highest content of cytoplasmic vesicle proteins (Figure 3A,E), which reflects the importance of vesicular transport of neurotransmitters in nerve tissue.

In most cases, purification of organelles is not an easy task, in particular, when only minute amounts of frozen tissue are available. Our results demonstrate that relatively high numbers of membrane proteins belonging to various organelles can be profiled without extensive fractionation, simplifying protein quantitation.

Conclusions

Detergents are powerful agents for solubilization of biological membranes and allow separation of membrane proteins using chromatographic methods such as size exclusion and ion-exchange chromatography. Development of methods for mass spectrometry-based proteomics of biological membranes is currently a subject of many investigations. As detergents are almost indispensable reagents in membrane biochemistry, the majority of relevant studies involve in-gel digestion to remove detergent prior to mass spectrometric analysis. In this work we introduced a simple and highly reproducible method for membrane proteomics that allows use of detergents. Moreover, we showed that the gel-free analysis of membrane proteins yields more than twice the number of protein identifications compared to in-gel digestion. Since our method offers a fast and reproducible means for analysis of membrane proteins, we believe that it may be suitable for high-throughput applications.

Acknowledgment. We thank Dr. Alexandre Zougman for the helpful advice in mass spectrometric analysis and critical reading of the manuscript.

Supporting Information Available: Tables of all proteins identified in this study and selection of neurotransmitter receptors and ion channels identified in mouse brain in two single-run experiments. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Furth, A. J. Removing unbound detergent from hydrophobic proteins. *Anal. Biochem.* **1980**, *109* (2), 207–15.
(2) Speers, A. E.; Wu, C. C. Proteomics of integral membrane proteins-theory and application. *Chem. Rev.* **2007**, *107* (8), 3687–714.
(3) Blonder, J.; Conrads, T. P.; Veenstra, T. D. Characterization and quantitation of membrane proteomes using multidimensional MS-based proteomic technologies. *Expert Rev. Proteomics* **2004**, *1* (2), 153–63.
(4) Blonder, J.; Xiao, Z.; Veenstra, T. D. Proteomic profiling of differentiating osteoblasts. *Expert Rev. Proteomics* **2006**, *3* (5), 483–96.
(5) Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–7.
(6) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R. 3rd, A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.* **2003**, *21* (5), 532–8.
(7) Olsen, J. V.; Andersen, J. R.; Nielsen, P. A.; Nielsen, M. L.; Figeys, D.; Mann, M.; Wisniewski, J. R. HysTag-a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. *Mol. Cell. Proteomics* **2004**, *3* (1), 82–92.
(8) Nielsen, P. A.; Olsen, J. V.; Podtelejnikov, A. V.; Andersen, J. R.; Mann, M.; Wisniewski, J. R. Proteomic mapping of brain plasma membrane proteins. *Mol. Cell. Proteomics* **2005**, *4* (4), 402–8.
(9) Olsen, J. V.; Nielsen, P. A.; Andersen, J. R.; Mann, M.; Wisniewski, J. R. Quantitative proteomic profiling of membrane proteins from the mouse brain cortex, hippocampus, and cerebellum using the HysTag reagent: Mapping of neurotransmitter receptors and ion channels. *Brain Res.* **2007**, *1134* (1), 95–106.
(10) Le Bihan, T.; Goh, T.; Stewart, I. I.; Salter, A. M.; Bukhman, Y. V.; Dharsee, M.; Ewing, R.; Wisniewski, J. R. Differential analysis of membrane proteins in mouse fore- and hindbrain using a label-free approach. *J. Proteome Res.* **2006**, *5* (10), 2701–10.
(11) Lu, X.; Zhu, H. Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins. *Mol. Cell. Proteomics* **2005**, *4* (12), 1948–58.
(12) Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **2003**, *75* (3), 663–70.
(13) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1* (6), 2856–60.
(14) Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **1996**, *379* (6564), 466–9.
(15) Olsen, J. V.; Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (37), 13417–22.
(16) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4* (12), 2010–21.
(17) Cox, J.; Mann, M. Is proteomics the new genomics. *Cell* **2007**, *130* (3), 395–8.
(18) Graumann, J.; Hubner, N. C.; Kim, J. B.; Ko, K.; Moser, M.; Kumar, C.; Cox, J.; Scholer, H.; Mann, M. 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* **2008**, *7* (4), 672–83.
(19) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20* (18), 3551–67.
(20) Zubarev, R.; Mann, M. On the proper use of mass accuracy in proteomics. *Mol. Cell. Proteomics* **2007**, *6* (3), 377–81.
(21) Kristensen, D. B.; Brond, J. C.; Nielsen, P. A.; Andersen, J. R.; Sorensen, O. T.; Jorgensen, V.; Budin, K.; Matthiesen, J.; Veno, P.; Jespersen, H. M.; Ahrens, C. H.; Schandorff, S.; Ruhoff, P. T.; Wisniewski, J. R.; Bennett, K. L.; Podtelejnikov, A. V. Experimental Peptide Identification Repository (EPIR): an integrated peptide-centric platform for validation and mining of tandem mass spectrometry data. *Mol. Cell. Proteomics* **2004**, *3* (10), 1023–38.
(22) Rabilloud, T.; Gianazza, E.; Catto, N.; Righetti, P. G. Amidosulfobetaines, a family of detergents with improved solubilization properties: application for isoelectric focusing under denaturing conditions. *Anal. Biochem.* **1990**, *185* (1), 94–102.
(23) Churchward, M. A.; Butt, R. H.; Lang, J. C.; Hsu, K. K.; Coorssen, J. R. Enhanced detergent extraction for analysis of membrane proteomes by two-dimensional gel electrophoresis. *Proteome Sci.* **2005**, *3* (1), 5.
(24) Chevallet, M.; Santoni, V.; Poinas, A.; Rouquie, D.; Fuchs, A.; Kieffer, S.; Rossignol, M.; Lunardi, J.; Garin, J.; Rabilloud, T. New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. *Electrophoresis* **1998**, *19* (11), 1901–9.
(25) Amons, R.; Schrier, P. I. Removal of sodium dodecyl sulfate from proteins and peptides by gel filtration. *Anal. Biochem.* **1981**, *116* (2), 439–43.
(26) Lohaus, C.; Nolte, A.; Bluggel, M.; Scheer, C.; Klose, J.; Gobom, J.; Schuler, A.; Wiebringhaus, T.; Meyer, H. E.; Marcus, K. Multidimensional chromatography: a powerful tool for the analysis of membrane proteins in mouse brain. *J. Proteome Res.* **2007**, *6* (1), 105–13.
(27) Fischer, F.; Wolters, D.; Rogner, M.; Poetsch, A. Toward the complete membrane proteome: high coverage of integral membrane proteins through transmembrane peptide detection. *Mol. Cell. Proteomics* **2006**, *5* (3), 444–53.
(28) Moebius, J.; Zahedi, R. P.; Lewandrowski, U.; Berger, C.; Walter, U.; Sickmann, A. The human platelet membrane proteome reveals several new potential membrane proteins. *Mol. Cell. Proteomics* **2005**, *4* (11), 1754–61.
(29) Masuda, T.; Tomita, M.; Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J. Proteome Res.* **2008**, *7* (2), 731–40.
(30) Chick, J. M.; Haynes, P. A.; Molloy, M. P.; Bjellqvist, B.; Baker, M. S.; Len, A. C. Characterization of the rat liver membrane proteome using peptide immobilized pH gradient isoelectric focusing. *J. Proteome Res.* **2008**, *7* (3), 1036–45.

PR800412J

5 Analysis of post translational modification of linker histone H1 variants from human breast cancer sample using the peptide mapping method

5.1 Publication: Mapping of Lysine Monomethylation of Linker Histones in Human Breast Cancer

This manuscript presents the result from the project of the post translational modification analysis of linker histone H1 variants from human breast cancer sample. At the time of writing this thesis, it is being re-submitted to the journal of Molecular and Cellular Proteomics. The following pages contain the submitted version of the manuscript.

Mapping of Lysine Monomethylation of Linker Histones in Human Breast Cancer

Aiping Lu[1], Alexandre Zougman[1] Marek Pudełko[2], Marek Bębenek[2], Piotr Ziółkowski[3], Matthias Mann[*1], and Jacek R. Wiśniewski[1*]

Author affiliation:

[1]Department of Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried near Munich, Germany;

[2]2nd Department of Surgical Oncology, Lower Silesian Oncology Center, Plac Hirszfelda 12, 53-413 Wroclaw, Poland;

[3]Department of Pathology, Wrocław Medical University, ul. Marcinkowskiego 1, PL-50-368 Wrocław, Poland.

[*]To whom correspondence should be addressed: Email: jwisniew@biochem.mpg.de mmann@biochem.mpg.de, Fax: +49 89 8578 2219

Abstract:

Linker histones H1 are key modulators of chromatin structure. Tightness of their binding to DNA is regulated by posttranslational modifications. In this study we have analyzed posttranslational modifications of five major variants of H1 in human tissue - H1.0, H1.2, H1.3, H1.4, and H1.5. To improve sequence coverage, tryptic peptides of H1 were separated by HPLC and the individual fractions were analyzed using a peptide on-chip implementation of nanoelectrospray (TriVersa), coupled to a linear ion trap-Orbitrap hybrid instrument. For quantitative analysis of lysine methylation, ionization efficiencies of methylated and non-methylated peptides were determined using synthetic peptides. Our analysis revealed that monomethylation of lysine residues alongside with phosphorylation of serine and threonine residues is the major modification of H1 in tissue. We found that most prominent methylation sites are in the N-terminal tail and the globular domain of H1. In the C- terminal domains we identified only few less abundant methylation sites. Quantitative analysis revealed that up to 25% of H1.4 is methylated at K-26 in human tissues. Another prominent methylation site was mapped to K-27 in H1.5, which resembles K-26 site in H1.4. In H1.0 five less abundant (<1% of H1.0) sites were identified. Analysis of patient matched pairs of cancer and adjacent normal breast demonstrated high variation between individuals. Our study revealed differences in methylation of linker histones between normal breast and its cancer.

Introduction

Diverse posttranslational modifications of core histones (H2A, H2B, H3, and H4) that form nucleosomal cores can alter transcriptional activity of expression of associated genes and affect DNA replication and repair processes (for a review see (1)). The link between the modifications and the epigenetic regulation of chromatin is known as the "histone code" (2). Whereas the modifications of core histones have been extensively studied little is known on the involvement of linker histones in this code.

Linker histones H1 are responsible for condensation of the nucleosomal “string-on-beads” structure into the 30 nm chromatin fiber (3). Mammalian cells contain seven major variants of histone H1: H1.0, H1.1, H1.2, H1.3, H1.4, H1.5 (4-7), and H1X (8,9). Whereas four variants H1.2, H1.3, H1.4, and H1.5 are the most abundant and widely spread in all cells (10), the levels of other three are more variable. H1.1 variant is considered a specific variant for thymus, testis, and spleen, lymphocytic and neuronal cells. H1.0 variant is abundant in terminally differentiated cells but significant levels of this variant were identified in cultured cancer cells (10,11) and detected by immunochemistry in breast carcinoma (12). Expression of *H1.X* variant gene has been observed across the majority of tissues, but so far, the protein was identified solely in cultured cells. Functional specificity of the individual variants remains obscure (13).

For a long time, in contrast to core histones, H1 was considered as to be solely modified by S/T kinases, mainly by CDK1 (14,15). Recently, monomethylation of K-26 in human H1.4 variant (16) and lysine acetylation for the *Tetrahymena* H1 were reported (17). Our recent studies on modifications of H1 in two cells lines and 9 mouse tissues revealed a greater array of posttranslational modifications. In addition to previously described phosphorylation, we found that linker histones are mono-, and dimethylated, acetylated, formylated, and ubiquitinated (10), thus they are sharing some similarity with core histones in respect to posttranslational modifications. Our study also revealed that linker histones from tissues carry more

methylated lysyl residues than histones from cultured cells. In contrast, the extent of lysine acetylation was found to be lessened in tissues. These findings suggested that in artificial systems of rapidly proliferating cells the modifications of the proteins do not necessarily reflect the situation *in vivo*. In this work, we have analyzed and compared posttranslational modifications of the histone H1 variants isolated from breast carcinoma and normal breast. Our analysis was performed on chromatographically separated peptides using a peptide on-chip implementation of nanoelectrospray (TriVersa), coupled to a linear ion trap-Orbitrap hybrid instrument (18). For quantitative analysis of lysine methylation, ionization efficiencies of monomethylated and non-methylated peptides were determined. We show that, besides phosphorylation, lysine monomethylation is the most prominent posttranslational modification of linker histones in human tissue, and that changes in methylation extent at individual lysine residues accompany cancer.

EXPERIMENTAL PROCEDURES

Tissue extraction – Samples of ductal invasive carcinoma (DIC) of grades 2 or 3 and the adjacent normal tissue were retrieved during surgery. Analysis of the samples followed an informed consent approved by the local ethics committee. The protein extraction procedure was carried out as described previously (10). Briefly, frozen tissue was homogenized with 3 vol. (m/v) of 5 % (v/v) $HClO_4$ using an IKA Ultra Turbax blender and the homogenate was centrifuged at 15,000 $\times$ g for 5 min. The supernatants were precipitated by addition of 100% (v/v) CCl_3COOH to a final concentration of 33% (v/v). After 30 min on ice the precipitate collected by centrifugation at 15,000 $\times$ g for 10 min.

Histone H1 purification, digestion and peptide fractionation – The protein pellets were dissolved in water and chromatographed on C_{18} reverse phase column as described previously (10,19). Fractions containing histone H1 were collected and the proteins vacuum dried. Dried protein pellets were reconstituted in 100 mM ammonium bicarbonate and digested with trypsin overnight at 37 °C. The resulting

peptide mixtures were separated on a 50mm x i.d. 2 mm Jupiter 4 µm Proteo 90Å column (Phenomenex) using an H_3CCN gradient in 0.25% CF_3COOH/water at 40 °C. The gradient has a following profile: 0 to 6% of organic solvent from 0 to 8 min; 6% to 20% from 8 to 15min; 20% to 40% from 15 to 22 min. The gradient was run at the flow rate of 0.2 ml/min, and fractionated every 2 min. The first 11 fractions were vacuum dried for mass spectrometric analysis.

Standard peptides – Peptide pairs with either non- or monomethylated lysyl residues were chromatographed using the 4 µm Proteo column as described above. Relative quantitation of the peptide pairs was achieved by integrating the peak areas of the chromatograms at 215 nm.

Mass spectrometric analysis – Identification and relative quantitation of the peptides was performed as described previously (18). Briefly, peptides were analyzed by a chip implementation of nanoelectrospray (TriVersa Nanomate, Advion Biosciences, Ithaca, US), coupled to a linear ion trap-orbitrap hybrid instrument (LTQ-Orbitrap, Thermo Fisher, Bremen, Germany). Data acquisition was performed in three steps. In the first step, full range spectra (full scan) from 75 to 1150 were acquired for 2 min in the Orbitrap with the resolution set to a value of 30000. In the second step, survey MS scans were acquired in the Orbitrap with four overlapping segmented mass range (75-350, 300-600, 550-850, 750-1150, Selected Ion Monitoring, or SIM scans) with a resolution of 60000. Up to 3 most intense ions per SIM scan were fragmented and acquired in the linear ion trap. In the third step, survey MS scans were acquired in two overlapping segmented mass range (75-600, 550-1150, SIM scans) with a resolution of 60000 in the Orbitrap, and up to 3 most intense ions per scan were fragment in the C-trap and acquired in the Orbitrap with a resolution of 15000. All three steps were composed in one instrument acquisition method setup, by different scan events. The total acquisition time for a sample can be different from batch to batch, from 15 min to maximum 30 min. The acquisition time for the first step was always fixed to 2 min, and the acquisition time for SIM-IT MS^2 and SIM-FT MS^2 was set roughly to 1:1. Target values were 1,000,000 for full scan, 300,000–

500,000 for SIM and FT MS/MS scan, 7,000–10,000 for IT MS/MS scan. For accurate mass measurements the lock-mass option was used as previously described (20).

Protein and posttranslational modifications identification – Combined peak lists for histone samples were extracted by MaxQuant software (21). The data were searched against a histone database (298 non-redundant sequences, including different histone proteins/variants, High mobility group protein variants, keratins and the proteases used) with the aid of the MASCOT (Matrix Science, London, UK) search engine (22), using 10 ppm MMD for precursor ion, 0.4 Da MMD for ion-trap fragmentation, and 0.01 Da mass tolerance for C-trap fragmentation. Up to three missed cleavages were allowed. No fixed modification was applied, and eight variable modifications, including acetylation on protein N-terminus, methionine oxidation, mono-, and di-, tri-methylation of lysines and arginines, serine/threonine phosphorylation, lysine acetylation and formylation. Results retrieved from Mascot were validated manually. Several rules were applied during validation: 1) For non-modified peptides with a length longer than 6 amino acids, IT MS/MS was enough for the ones with at least 3 amino acids conservatively identified and a score higher than that corresponding to a significance value of p = 0.05; 2) for non-modified peptides with a length equal to 5 or 6 amino acids, only whole sequence identification from FT MS/MS were accepted; 3) for methylation, accurate precursor ion m/z (14.01 delta mass add-up for one methyl group) and elution time (within ±1 elution faction) compared to the non-metylated corresponding peptide sequence plus FT/IT MS/MS were taken into consideration, mass accuracy of all identified mono-methylation sites was listed in supplementary table 1, and MS/MS spectra for all identified mono-methylation sites can be found in either Fig. 4 or supplementary Fig. 1; 5) for formylation and acetylation, only the modified sites clearly assigned in FT MS/MS were accepted.

Quantitation of posttranslational modifications – Since we performed off-line HPLC separation for the tryptic digested histone peptides, the modified peptide could be eluted either in the same or different fractions with the same peptide sequence

without modifications, leading to different relative quantitation strategies for the two situations. For the peptide pairs eluted in the same fraction, deconvoluted peak intensity from the 2-min-full-range-summed-up spectrum was used for relative quantitation, the same as described previously (18). And for the peptide pairs eluted in different fractions, the total ion current of all peaks of the same peptide with different charge states in different fractions was used for relative quantitation. Relative quantitation ratio of methylation was calculated independently in peptide pairs with the same tryptic cleavage (details were listed in supplementary table 2), and the average ratio of one particular modified site in different cleaved peptides was considered as the relative ratio for the sample. The occupancy quantitation values shown in the results were derived from the relative quantitation ratios with normalization from ionization efficiency enhancement for methylation.

RESULTS

Development of a procedure that allows in-depth analysis of posttranslational modifications of linker histones – Mass spectrometry based identification of posttranslational modifications require digestion of proteins to peptides. Usually trypsin or another residue specific endoproteinase enable cleavage of proteins into peptides of a length allowing their identification over the entire sequence. Linker histones are in 30 % composed of lysine and therefore digestion with trypsin results in generation of a large number of small peptides which escape from detection using conventional LC-setup.

To improve the sequence coverage of analyzed H1 we separated off-line tryptic peptides of H1 and then analyzed individual fractions using a chip implementation of nanoelectrospray coupled to a linear ion trap-Orbitrap hybrid instrument (18). In this approach partially purified linker histones were digested with trypsin and the resulting peptides were separated into 11 fractions on a C_{18} column in the presence of a high concentration of ion-pairing reagent (Fig. 1). A comparison of the data obtained using this approach and using the standard LC-MS setup revealed clear

advantages of our novel approach (Fig. 2). In particular, sequence coverage of the C-terminal domain (CTD) of the histone variants was improved from 20-40% to 70-80% (Fig. 2C and D).

Relative quantitation of linker histone methylation – In respect to abundance, phosphorylation is the most prominent posttranslational modification in linker histones and has been studied extensively using biochemical methods in the past (14,23). In contrast, little is known about the abundance of other posttranslational modifications that have recently been discovered using high performance mass spectrometry (10,16,24). In our previous work, we have already introduced the concept to quantify modifications by peak intensity (18), and we also found that both the TIC and peak intensity are highly reproducible when we sprayed the same sample several times. To reduce the error-factor of solvent related ion suppression, all samples were electrosprayed in the same ionization solvent. Due to differences in ionization efficiency between peptide with and without a modification a simple comparison of ion intensities cannot be used directly as a measure of the fractional abundance of the modification. For this purpose first ionization efficiencies have to be determined using standard peptides. For quantitation of methylations sites in H1 we have synthesized peptide pairs of methylated and non-methylated peptide, corresponding to the tryptic peptides 26-32 and 65-75 of H1.4 and peptide 27-35 of H1.5. Measurements of ion-intensities revealed 1.3-2.0 fold higher ionization efficiencies of the methylated peptides in comparison to their non-modified forms (Table 1). For relative quantitation of methylation at K-26 and K-75 of H1.4 and K-27 of H1.5 the ratio of intensities between modified and unmodified peptide were normalized using the values shown in Table 1. For normalization of the extents of other methylations we used an average value of 1.5. For methylation sites other than K-26 and K-75 of H1.4 and K-27 of H1.5 the averaged normalization factor provides only rough estimation of their abundances.

The quantitation method was tested for linearity and reproducibility using synthetic peptides (Fig. 3). The peptides **mK**AAGAGAAK and KAAGAGAAK were sprayed either

individually or as a mixture over a wide range of concentrations. We found out that the ratio of XIC intensity to MS intensity of the same peptide in the same spray is 1:1 (Fig. 3, panel A) and the intensity values are directly proportional to the peptide concentration over a dynamic range used for peptide quantitation (Fig.3, panel B). Over a wide range of peptide concentrations the observed signal intensity had a standard deviation below 10% (Fig. 3, Panel C).

Mapping of posttranslational modification in human tissue – Samples of breast cancer and normal breast were used to map posttranslational modifications in human tissue. In addition to previously described sites of acetylation, phosphorylation, and formylation sites (10,24) we found several sites of lysine monomethylation in H1.0, H.1.2 H1.3, H1.4 and H1.5 (Table 2). In linker histone H1.0 five methylation sites were identified. We found one at K-12 in the N-terminal domain (NTD), three sites in the GH1 (K-82, K-102, and K108), at K-155 in CTD (Table 2, Fig. 4 or supplementary Fig. 1). Analysis of the fraction containing H1.2, H1.3, and H1.4 reveled four methylation sites common for each variant (K-52, K-64, K-97 and K-106, numbered in H1.2), one site unique to H1.2 (K-168), and two sites exclusive for H1.4 (K-26, and K-148) (Table 2, Fig. 4 or supplementary Fig. 1). In addition we found a methylation at K-119, a site shared by H1.2 and H1.4 (Table 2, supplementary Fig. 1). In the variant H1.5 methylation at K-27, a site resembles K-26 at H1.4, was identified (Table 2, Fig. 4). Monomethylated lysine residues K-26 and K-75 in H1.4 were also found in their dimethylated forms (Fig. 5).

In our analysis we used 50% methanol/H_2O acidic solution as the spray buffer, for its ionization-supportive property and low background. Since recent work (25) has demonstrated that glutamic acid can partially be methylated in a buffer containing methanol, we analyzed carefully every methylation site. We found that several peptides occurred in two forms with either methylated lysine or with methyl glutamate. Both peptide forms have exact the same precursor ion mass, as shown in the MS/MS spectra of Fig.3A and E, supplementary Fig.1B. For this reason we do not consider several methylated peptides with adjacent glutamic acid and lysine

(including SETAPAAPAAAPPAEK (2-21, H1.2), SETAPAAPAAPAPAEK (2-17, H1.4), SETAPAETATPAPVEK (2-17, H1.5), YSDMIVAAIQAK (28-40, H1.0), ALAAAGYDVEK (65-75, H1.2-4) , ALAAGGYDVEK (68-78, H1.5)) because we missed clearly assigned y_1 ions in the MS/MS spectra. In these cases it was not possible to differentiate between esterified glutamic acid and methylated lysine solely by *b*-ions which easily loosing methyl moiety from glutamic acid ester during the fragmentation process. For peptides with glutamic acid and lysine in non-consecutive position, we kept the identifications without quantitation values, since we were not able to measure the relative stoichiometric abundance of two methylated forms.

Quantitative analysis showed that methylation sites in NTD and GH1 are much more abundant than those in CTD (Table 2, Fig. 6). In NTD we found that up to 25% of H1.4 is methylated at K-26. Another prominent methylation site was mapped to K-27 in H1.5. In H1.0 five less abundant (<1% of H1.0) were identified (Table 2, Fig. 6).

Since we observed high variation in levels of methylation at individual sites between different samples (Table 2, Fig. 6), we decided to analyze patient matched pairs of grade 2 cancer and adjacent normal breast. Analysis of the methylation extents revealed high variation between three studied cases. There was no correlation between individual cases and the levels of methylation at different sites (Fig. 7).

DISCUSSIONS

This study provides first insights in posttranslational modifications of linker histone variants in human tissues. We show that in both normal and in breast cancer tissue lysine methylation is the second, after phosphorylation, most abundant posttranslational modification of linker histones. Our results emphasize differences between tissue and cultured cells reported previously (10). Whereas in cultured breast and cervical cancer cells lysine multiple sites of lysine acetylation were found and only little lysine methylation was observed. In contrast, in linker histones isolated from mouse tissues methylation appeared to be more frequent. Two

methylation sites were identified in NTDs of mouse H1.1, H1.2, and H1.3, and six sites in the GH1 were mapped in different H1 variants. In this study, using improved method of analysis, we identified most of these sites in human tissue as well as identified additional methylation sites. Resembling methylation of mouse linker histones, human H1 from breast tissue are mainly methylated in NTD and GH1. Despite extending sequence coverage for the C-terminal portion of H1 we identified only few sites with methylated lysine in CTD.

In the NTDs of the H1 two distinct sites of methylation sites were identified, one at the from the N-terminus first lysine of H1.0 and the second at the previously reported K-26 in H1.4 (16), as well as methylation of K-27 of H1.5 resembles modification of K-26 of H1.4. Modification at the N-terminus first lysine of other histone 1 variants was uncertain because of the glutamic acid artifact methylation. Since function of the NTD is unknown it is difficult to speculate on the role of these modifications, however it has been demonstrated that methylated K-26 is specifically recognized by chromo domain of HP1 (26).

The mechanism and substrate specificity of enzymes involved in methylation of linker histone are not identified yet. Two major features of methylation of H1 can be seen from our results. First, extends of the methylation correlate with the abundance of individual variants (Fig. 1, chromatogram A). H1.4 which is the most abundant human variant of H1 is more methylated than less abundant H1.5 and H1.2. The methylation extents are lowest in the H1.0 which occurs at lowest levels when compared to the other analyzed variants. Since it is rather less probable that for example H1.4 is a better substrate for methyltransferases than other H1 variants the differences in H1 methylation can be explained in terms of enzyme kinetics. In such case the effective concentrations of individual H1 variants lie in the range of Km (acceptor) of the methyltransferase and therefore moderate differences in protein concentration result in higher changes in the methyl transfer.

The second feature of H1 methylation is its N-terminal *polarization.* The most abundant methylation site is located in NTD, the moderately methylated lysines are in GH1, and little methylation occurs in CTD. The distribution of the methylation in H1 inversely reflects charge distribution of linker histones with highest positive charge density in CTD and the lowest in NTD. The major function of CTD is nucleosome binding and stabilization of chromatin folding (27), thus it is possible that lysines in this domain are almost inaccessible to methyltrasferases. GH1 binds dynamically to nucleosomes (28,29) whereas the mainly non-polar NTD may be most accessible to the modifying enzymes.

For several reasons proteomic analysis of solid human tissues is not an easy task. These include genetic or age related variation between individuals and heterogeneity of tissue sample which is usually composed of different types of cells. In single cases cancer cells are often mixtures of cells that are classified to different grades (30). Laser dissection is normally introduced to select single type of cells. However, due to the laborious procedure, very limited amount of sample is available by this method and which is normally not enough for post-translational modifications analysis. Tissue storage and extraction procedures can additionally contribute to variability of the results. Our analysis clearly demonstrates how difficult it is to analyze and problematic it can be to draw conclusions from studies on human tissue. Quantitative mapping of methylation sites revealed high level variability between individual samples. Since all samples were stored, processed, and analyzed in the same way, we believe that observed variation reflects solely the nature of individual samples and that it reflects activities of endogenous methyltransferases.

Despite high differences in the methylation extent between samples of different origin our results show that in cancer at majority of sites levels of methylation are augmented. In particular an increase in monomethylation was observed at K-26 and K-97 of H1.4 The observed elevated extent of methylation correlates well with the previous reports showing that methyltransferase EZH2 is highly expressed in various

malign tumors including metastatic prostate cancer, lymphomas and breast (31,32). Another study demonstrated an ability of EZH2 to modify H1 (33), but the methylated lysines were not mapped. Further work should provide more detailed map of methyl-acceptor-lysines in H1 upon action of this methyltransferase.

In contrast to acetylation, which is usually related to transcriptional activation, methylation of histones results in either activation or repression of transcription. The effect of methylation depends on the modified residue and simultaneous posttranslational modification of other residues of a histone. So far more than 10 different histone methyltransferases were characterized in humans (34). Moreover, enzymatic lysine demethylation process has been described (35,36). Thus, by analogy to core histones, properties of linker histones appear to be regulated by methylation of lysine residues.

Our present and the previous study on posttranslational modification of linker histones (10) provide a relative complex picture. In addition to multiple phosphorylation sites that have been extensively studied in the past, we identified multiple sites of lysine acetylation and methylation. Furthermore, lysine residues in H1 are frequently formylated (24). In human cultured cells acetylation appear to be the more abundant than methylation, whereas in mouse and human tissue methylation appears to be more prominent (10). Our data suggest that elevated methylation in tissues reflects higher extent of transcriptional silencing in tissue, including tumor in comparison to cultured cells. However, we are far from elucidation of the role of posttranslational modifications in H1. Moreover, our studies on linker histones may raise a question of relevance of cultured cells as model system for studying cancer (10).

Acknowledgements: The authors would like to thank Dr. Michael Lund Nielsen for the helpful advice and discussion. This work was supported by the Max-Planck Society for the Advancement of Science and by HEROIC, an Integrated Project funded by the European Union under the 6th Framework Program (LSHG-CT-2005-018883).

References

1. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693-705.
2. Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, **293**, 1074-1080.
3. Thoma, F., Koller, T. and Klug, A. (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol*, **83**, 403-427.
4. Panyim, S. and Chalkley, R. (1971) The molecular weights of vertebrate histones exploiting a modified sodium dodecyl sulfate electrophoretic method. *The Journal of biological chemistry*, **246**, 7557-7560.
5. Lennox, R.W. and Cohen, L.H. (1983) The histone H1 complements of dividing and nondividing cells of the mouse. *The Journal of biological chemistry*, **258**, 262-268.
6. Albig, W., Drabent, B., Kunz, J., Kalff-Suske, M., Grzeschik, K.H. and Doenecke, D. (1993) All known human H1 histone genes except the H1(0) gene are clustered on chromosome 6. *Genomics*, **16**, 649-654.
7. Rall, S.C. and Cole, R.D. (1971) Amino acid sequence and sequence variability of the amino-terminal regions of lysine-rich histones. *The Journal of biological chemistry*, **246**, 7175-7190.
8. Yamamoto, T. and Horikoshi, M. (1996) Cloning of the cDNA encoding a novel subtype of histone H1. *Gene*, **173**, 281-285.
9. Happel, N., Schulze, E. and Doenecke, D. (2005) Characterisation of human histone H1x. *Biol Chem*, **386**, 541-551.
10. Wiśniewski, J.R., Zougman, A., Kruger, S. and Mann, M. (2007) Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol Cell Proteomics*, **6**, 72-87.
11. Zougman, A. and Wiśniewski, J.R. (2006) Beyond linker histones and high mobility group proteins: global profiling of perchloric acid soluble proteins. *J Proteome Res*, **5**, 925-934.
12. Kostova, N.N., Srebreva, L.N., Milev, A.D., Bogdanova, O.G., Rundquist, I., Lindner, H.H. and Markov, D.V. (2005) Immunohistochemical demonstration of histone H1(0) in human breast carcinoma. *Histochem Cell Biol*, **124**, 435-443.
13. Izzo, A., Kamieniarz, K. and Schneider, R. (2008) The histone H1 family: specific members, specific functions? *Biol Chem*.
14. Gurley, L.R., D'Anna, J.A., Barham, S.S., Deaven, L.L. and Tobey, R.A. (1978) Histone phosphorylation and chromatin structure during mitosis in Chinese hamster cells. *Eur J Biochem*, **84**, 1-15.
15. Hohmann, P., Tobey, R.A. and Gurley, L.R. (1976) Phosphorylation of distinct regions of f1 histone. Relationship to the cell cycle. *The Journal of biological chemistry*, **251**, 3685-3692.
16. Garcia, B.A., Busby, S.A., Barber, C.M., Shabanowitz, J., Allis, C.D. and Hunt, D.F. (2004) Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry. *J Proteome Res*, **3**, 1219-1227.
17. Garcia, B.A., Joshi, S., Thomas, C.E., Chitta, R.K., Diaz, R.L., Busby, S.A., Andrews, P.C., Ogorzalek Loo, R.R., Shabanowitz, J., Kelleher, N.L. *et al.*

(2006) Comprehensive phosphoprotein analysis of linker histone H1 from Tetrahymena thermophila. *Mol Cell Proteomics*, **5**, 1593-1609.

18. 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 LCMS on an orbitrap mass spectrometer. *Int. J. Mass Spetrometry*, **268**, 158-167.
19. Renner, U., Ghidelli, S., Schafer, M.A. and Wiśniewski, J.R. (2000) Alterations in titer and distribution of high mobility group proteins during embryonic development of Drosophila melanogaster. *Biochim Biophys Acta*, **1475**, 99-108.
20. Olsen, J.V., de Godoy, L.M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S. and Mann, M. (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.
21. Cox, J. and Mann, M. (2007) Is proteomics the new genomics? *Cell*, **130**, 395-398.
22. Perkins, D.N., Pappin, D.J., Creasy, D.M. and Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, **20**, 3551-3567.
23. Gurley, L.R., Valdez, J.G. and Buchanan, J.S. (1995) Characterization of the mitotic specific phosphorylation site of histone H1. Absence of a consensus sequence for the p34cdc2/cyclin B kinase. *The Journal of biological chemistry*, **270**, 27653-27660.
24. Wiśniewski, J.R., Zougman, A. and Mann, M. (2008) Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. *Nucleic Acids Res*, **36**, 570-577.
25. Jung, S.Y., Li, Y., Wang, Y., Chen, Y., Zhao, Y. and Qin, J. (2008) Complications in the Assignment of 14 and 28 Da Mass Shift Detected by Mass Spectrometry as in Vivo Methylation from Endogenous Proteins. *Analytical chemistry*, **80**, 1721-1729.
26. Daujat, S., Zeissler, U., Waldmann, T., Happel, N. and Schneider, R. (2005) HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *The Journal of biological chemistry*, **280**, 38090-38095.
27. Lu, X. and Hansen, J.C. (2003) Revisiting the structure and functions of the linker histone C-terminal tail domain. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, **81**, 173-176.
28. Brown, D.T., Izard, T. and Misteli, T. (2006) Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. *Nature structural & molecular biology*, **13**, 250-255.
29. Misteli, T., Gunjan, A., Hock, R., Bustin, M. and Brown, D.T. (2000) Dynamic binding of histone H1 to chromatin in living cells. *Nature*, **408**, 877-881.
30. Chapman, J.A., Miller, N.A., Lickley, H.L., Qian, J., Christens-Barry, W.A., Fu, Y., Yuan, Y. and Axelrod, D.E. (2007) Ductal carcinoma in situ of the breast (DCIS) with heterogeneity of nuclear grade: prognostic effects of quantitative nuclear assessment. *BMC cancer*, **7**, 174.

31. Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G., Otte, A.P. *et al.* (2002) The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*, **419**, 624-629.
32. Kleer, C.G., Cao, Q., Varambally, S., Shen, R., Ota, I., Tomlins, S.A., Ghosh, D., Sewalt, R.G., Otte, A.P., Hayes, D.F. *et al.* (2003) EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A*, **100**, 11606-11611.
33. Kuzmichev, A., Jenuwein, T., Tempst, P. and Reinberg, D. (2004) Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Molecular cell*, **14**, 183-193.
34. Santos-Rosa, H. and Caldas, C. (2005) Chromatin modifier enzymes, the histone code and cancer. *Eur J Cancer*, **41**, 2381-2402.
35. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A. and Shi, Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, **119**, 941-953.
36. Shi, Y.J., Matson, C., Lan, F., Iwase, S., Baba, T. and Shi, Y. (2005) Regulation of LSD1 histone demethylase activity by its associated factors. *Molecular cell*, **19**, 857-864.

Figure legends

Figure1. Overview of the method used for mapping of posttranslational modifications of linker histones. Frozen tissues were homogenized in 5% $HClO_4$. $HClO_4$ soluble proteins were precipitated with Cl_3CCOOH and chromatographed on a C_{18} reverse phase column. Proteins were eluted with CH_3CN gradient in 0.1% F_3CCOOH. Three fractions containing H1.0, H1.5 and H1.2-H1.4, respectively, were collected and vacuum-dried and digested with trypsin. Tryptic peptides were separated into 11 fractions on a c18 reverse-phase column using CH_3CN gradient in 0.25% F_3CCOOH. Peptide fractions were vacuum-dried and analyzed using a peptide on-chip implementation of nanoelectrospray (TriVersa), coupled to a linear ion trap-Orbitrap hybrid instrument. The chromatograms show separation of linker histones and peptide of tryptic peptides of the H1.2-H1.4 fraction. The histones were extracted from 1g of ductal cancer sample.

Figure 2. **Comparison of the sequence coverage of linker histone variants observed in Triversa- (A, C) and LC-MS/MS (B, D) analysis.** n, the number of experiments. NTD, N-terminal domain; GH1, globular domain of H1, CTD, C-terminal domain.

Figure 3. Linear range and error of the quantitation method. The quantitation method was analyzed for linearity and reproducibility using synthetic peptides mKAAGAGAAK and KAAGAGAAK that were electro-sprayed either individually or as a mixture over a wide range of their concentrations. (A) Correlation of intensities obtained from MS based and XIC based measurements. B) Estimation of the linearity range of the quantitation method. Mass peak intensities for peptide concentration ranging from 5x 10 M to 5 x 10-11 M was considered as linear. (C) Variability of peptide intensities measured over a wide range of peptide concentrations.

Figure 4. The MS/MS fragmentation spectra of K-methylated peptides in H1 from human breast. MS/MS spectra acquired in linear ion trap were shown with 2 decimal, and the spectra acquired in Orbitrap were shown with 4 decimal.

Figure 5. The MS/MS fragmentation spectra of K-dimethylated peptides in H1.4 variant.

Figure 6. Summary of identified methylation sites in linker histone variants H1.4 including sites matched by the sequences of H1.2 or/and H1.3 (A), H1.2 specific (B), H1.5 (C) and H1.0 (D). NTD, N-terminal domain; GH1, globular domain of H1, CTD, C-terminal domain. ***Black*** and ***grey bars*** represent values measured in normal tissue

and cancer, respectively. Closed circles represent the sites could not be quantified due to artifact methylation on glutamic acid.

Figure 7. Extent of methylation of lysine residues in matched pairs of normal tissue and cancer. Cases A, B, and C represent grade 2 of ductal invasive breast cancer in 73, 79, and 51 years old woman, respectively. ***Black*** and ***grey bars*** represent values measured in normal tissue and cancer, respectively.

Table 1. Relative ionization efficiency of synthetic methylated and non-methylated peptides of H1

H1 variant	Modified site	Residues	Peptide pairs	HPLC ratio	MS ratio	Ionization efficiency ratio
H1.4	K-26	26-32	**mK**SAGAAK / KSAGAAK	0.97	1.30	1.34
H1.5	K-27	27-35	**mK**AAGAGAAK / KAAGAGAAK	0.58	1.15	2.00
H1.2-H1.4	K-75	65-75	ALAAAGYDVE**mK** / ALAAAGYDVEK	0.97	1.25	1.29

Table 2. Relative abundance of linker histone H 1 methylation in cancer and normal breast tissue

H1 variant	Modified site	Residues	Peptide sequence	Methylation extent (Average ± Std)	
				Normal breast	Breast cancer
H1.0	K12	2-12	TENSTSAPAA**mK**	---	---
	K82	75-82	LVTTGVL**mK**	0.15% (1/4)	0.10% ± 0.13% (3/7)*
	K102	98-102	SDEP**mK**	---	---
	K108	104-108	SVAF**mK**	0.02% (1/4)	0.21% ± 0.21% (3/7)
	K155	150-155	LAATP**mK**	0.02%±0.02% (4/4)	0.05% ± 0.05% (6/7)
H1.2	K168	160-168	KPAAATVT**mK**	0.003%±0.003%(3/4)	0.26% ± 0.54% (5/6)
H1.2;	K52	47-52	AVAAS**mK**	0.22% ± 0.13% (4/4)	0.12% ± 0.08% (6/6)
H1.3; H1.4	K64	55-64	SGVSLAALK**mK**	0.96% (1/4)	0.70% ± 0.23% (2/6)
	K97	91-97	GTLVQT**mK**	0.02% ± 0.03%(3/4)	0.65% ± 0.56% (2/6)
	K106	98-106	GTGASGSF**mK**	0.021% (1/4)	0.09% ± 0.06% (2/6)
H1.2; H1.4	K119	111-119	AASGEAKP**mK**	---	---
H1.4	K26	26-32	**mK**SAGAAK	11.2% ± 2.87% (4/4)	20.4% ± 13.4% (6/6)
	K148	140-148	KATGAATP**mK**	0.003%±0.001%(3/4)	0.03% ± 0.05% (5/6)
H1.5	K27	27-35	**mK**AAGAGAAK	0.65 % ± 0.43% (3/3)	2.22% ± 3.56% (6/6)

*in brackets are given the number of identifications / number of samples.

--- means the site could not be truly quantified due to artifact glutamic acid methylation with the same nominal precursor ion mass

Figure 1

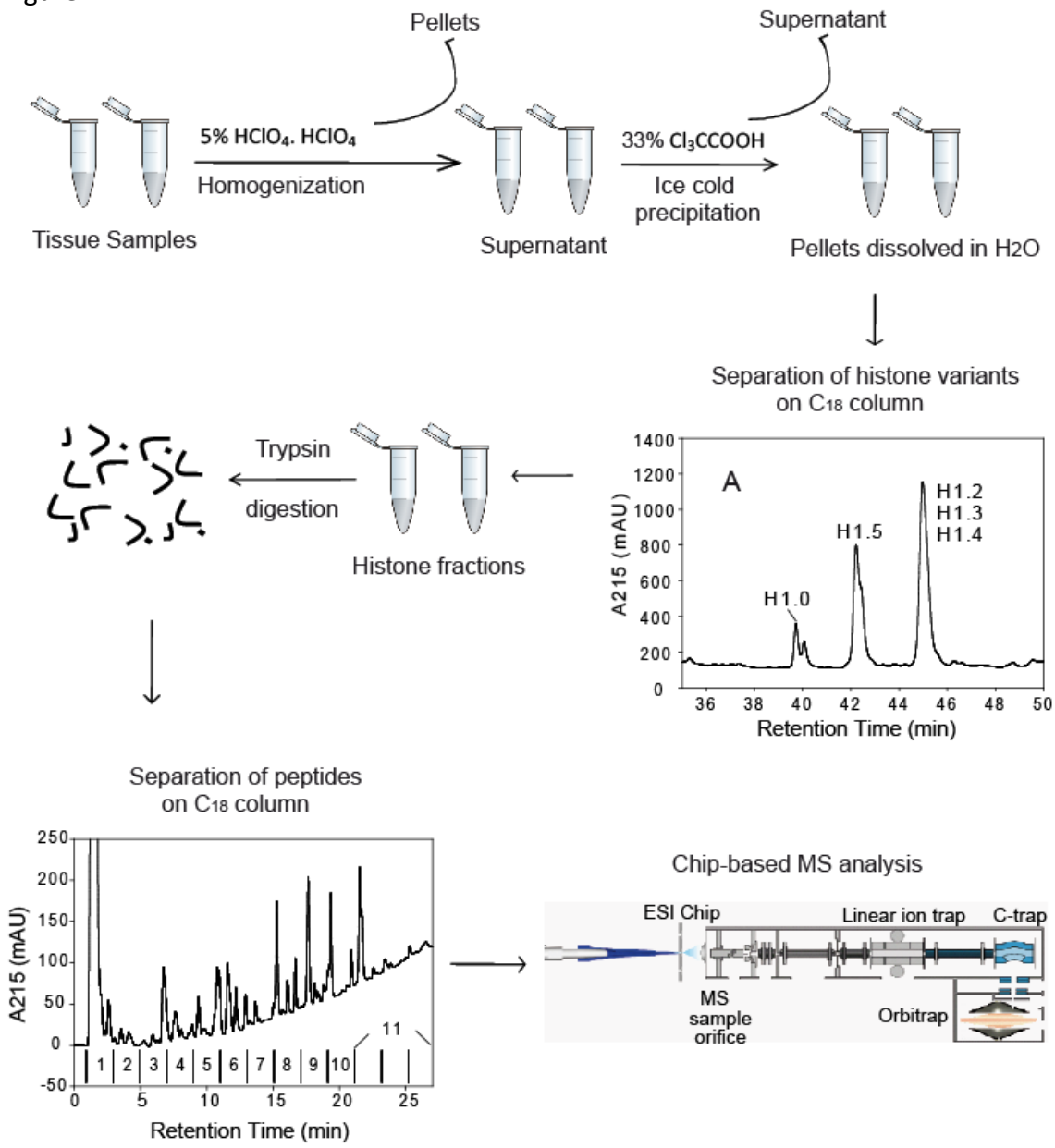

Pellets
Supernatant
5% HClO4. HClO4
Homogenization
33% Cl3CCOOH
Ice cold precipitation
Tissue Samples
Supernatant
Pellets dissolved in H2O
Separation of histone variants on C18 column
Trypsin
digestion
Histone fractions
A
H1.0
H1.5
H1.2
H1.3
H1.4
A215 (mAU)
Retention Time (min)
Separation of peptides on C18 column
A215 (mAU)
Retention Time (min)
Chip-based MS analysis
ESI Chip
Linear ion trap
C-trap
MS sample orifice
Orbitrap

Figure 2

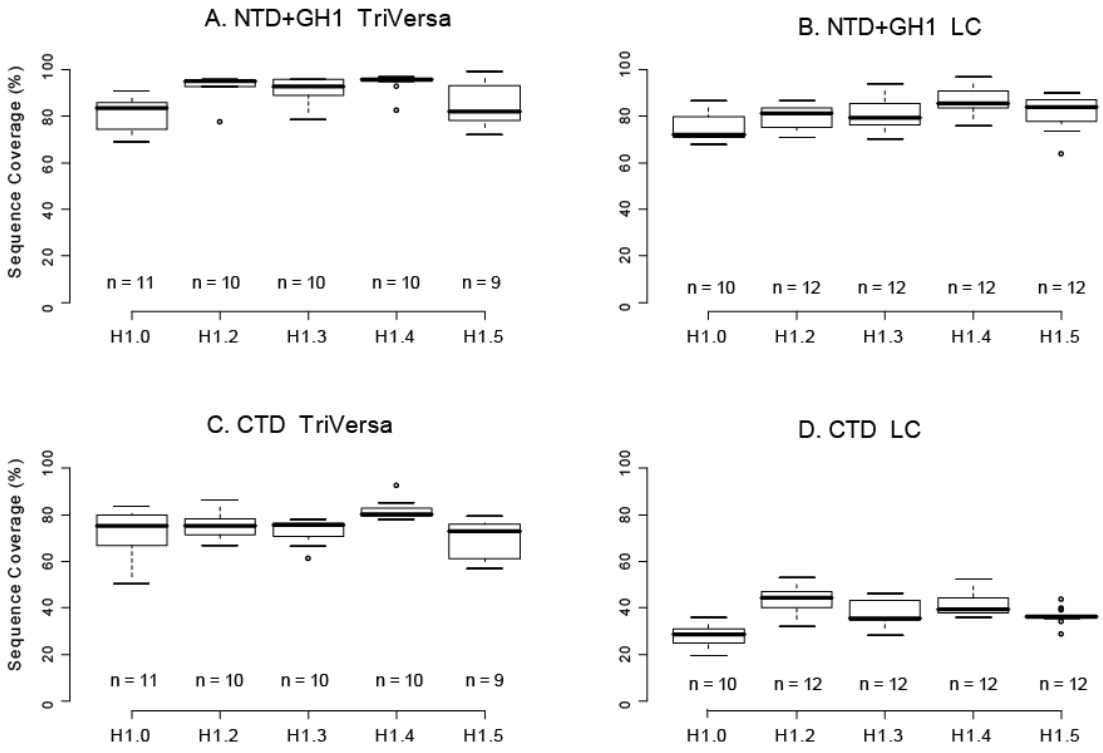

A. NTD+GH1 TriVersa
B. NTD+GH1 LC
C. CTD TriVersa
D. CTD LC
Sequence Coverage (%)
0
20
40
60
80
100
n = 11
n = 10
n = 10
n = 10
n = 9
n = 10
n = 12
n = 12
n = 12
n = 12
H1.0
H1.2
H1.3
H1.4
H1.5

Figure 3

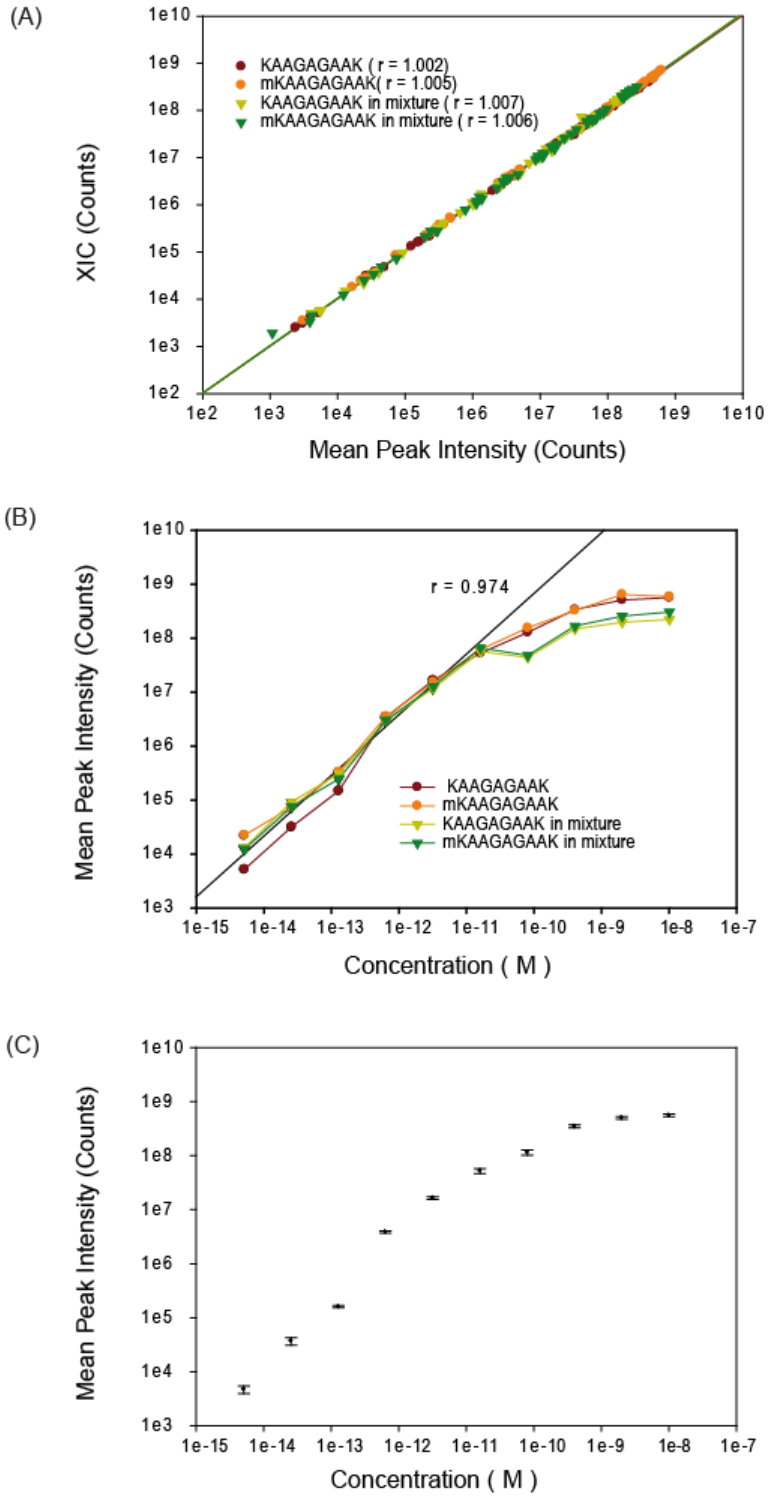

(A)
XIC (Counts)
Mean Peak Intensity (Counts)
KAAGAGAAK (r = 1.002)
mKAAGAGAAK(r = 1.005)
KAAGAGAAK in mixture (r = 1.007)
mKAAGAGAAK in mixture (r = 1.006)
(B)
Mean Peak Intensity (Counts)
Concentration (M)
r = 0.974
KAAGAGAAK
mKAAGAGAAK
KAAGAGAAK in mixture
mKAAGAGAAK in mixture
(C)
Mean Peak Intensity (Counts)
Concentration (M)

Figure 4

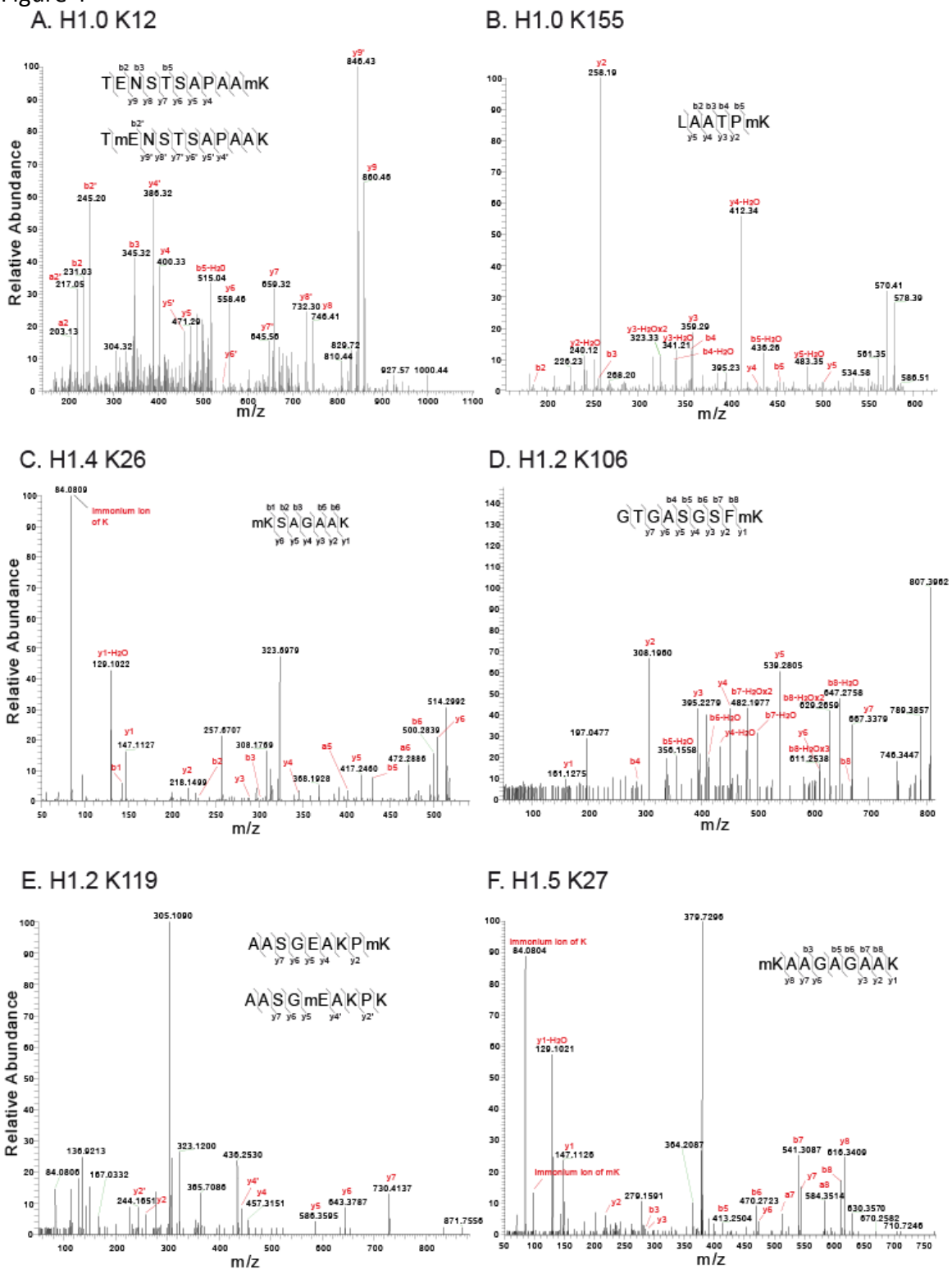
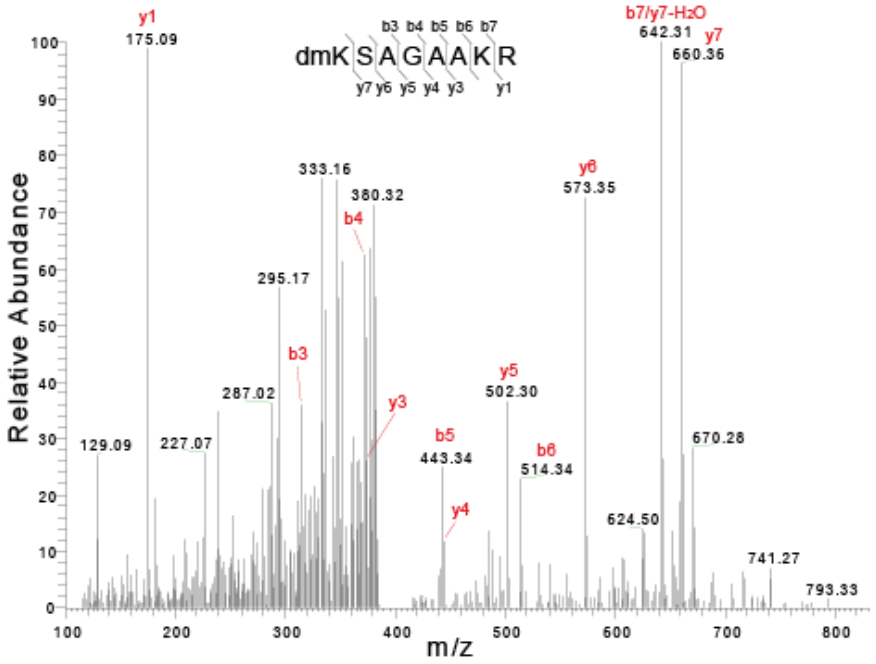

A. H1.0 K12
TENSTSAPAAmK
TmENSTSAPAAK
B. H1.0 K155
LAATPmK
C. H1.4 K26
mKSAGAAK
Immonium ion of K
D. H1.2 K106
GTGASGSFmK
E. H1.2 K119
AASGEAKPmK
AASGmEAKPK
F. H1.5 K27
mKAAGAGAAK
Immonium ion of K
Relative Abundance
m/z

Figure 5

A. H1.4 K26 DiMethylation



B. H1.4 K75 DiMethylation

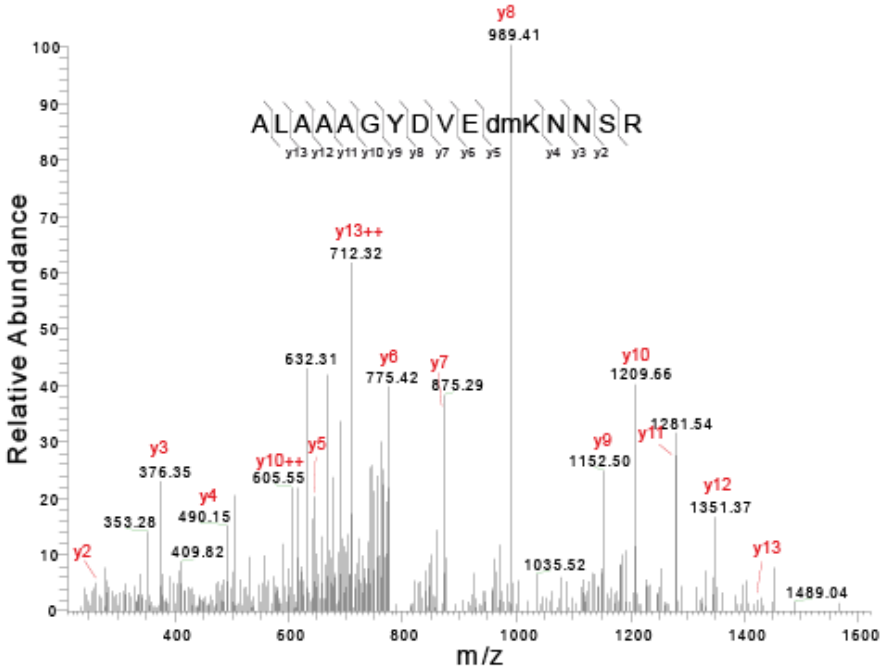
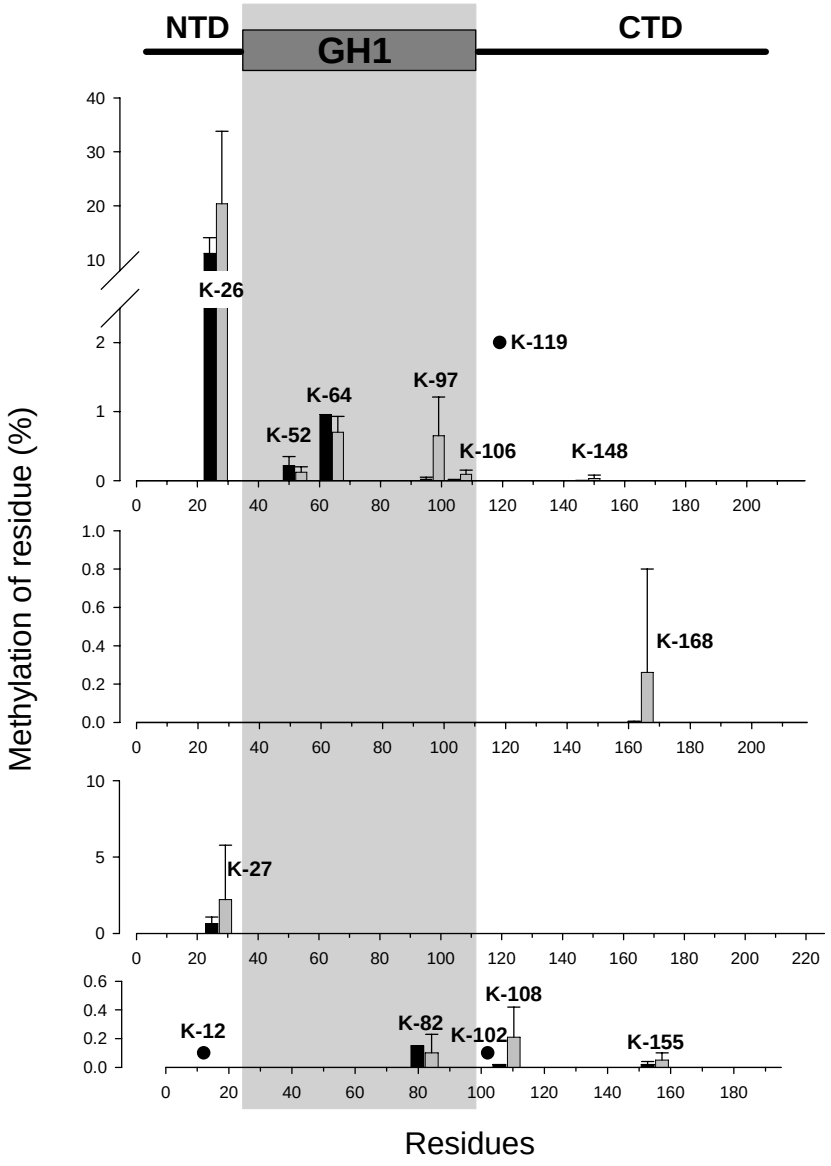


Figure 6



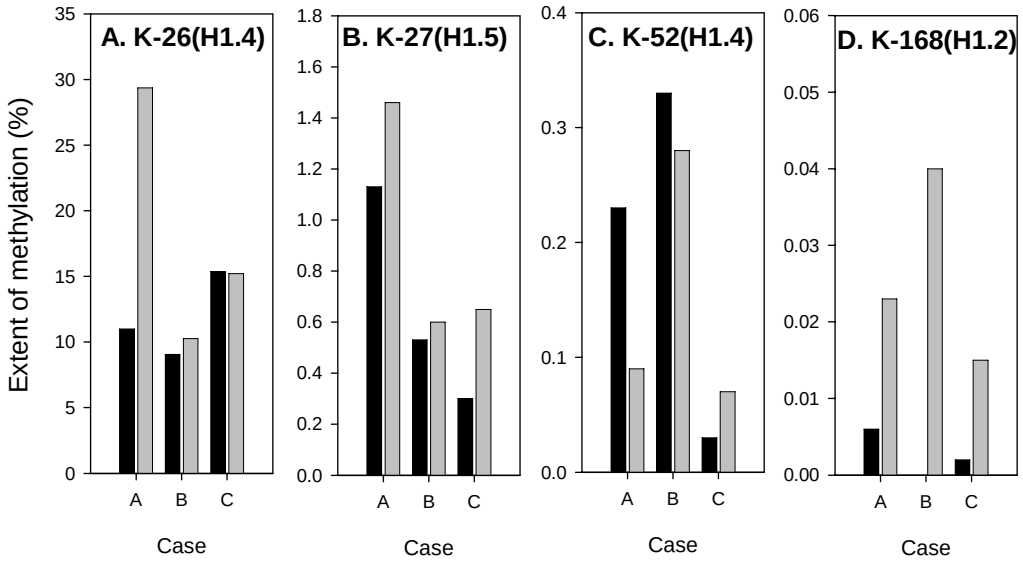
A. Histone H1.4 (H1.2 and H1.3)

B. Histone H1.2 (unique sites)

C. Histone H1.5

D. Histone H1.0

Figure 7


A. K-26(H1.4)
B. K-27(H1.5)
C. K-52(H1.4)
D. K-168(H1.2)
Extent of methylation (%)
Case
A
B
C

6 Comparative membrane proteomics from three rat nerve tissues

6.1 Publication: Comparative Proteomic Profiling of Membrane Proteins in Rat Cerebellum, Spinal Cord, and Sciatic Nerve

This manuscript contains the results from the project of membrane protein identification from three rat nerve tissue. The following pages contain the draft of the manuscript, which is being submitted to Journal of Proteome Research.

Comparative Proteomic Profiling of Membrane Proteins in Rat Cerebellum, Spinal Cord, and Sciatic Nerve

Aiping Lu, Jacek R. Wiśniewski*, and Matthias Mann*

Department of Proteomics and Signal Transduction, Max-Planck-Institute for Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

Running *title:* **Nerve tissue proteomics**

*To whom correspondence should be addressed:

Tel: +49 89 8578 2557

Fax: +49 89 8578 2219

Email: jwisniew@biochem.mpg.de or mmann@biochem.mpg.de

Keywords: proteomics • neurotransmitter receptor • ion channels • cerebellum • spinal cord • sciatic nerve

Abstract

Proteomics is an increasingly powerful technology that can provide in-depth insights into entire proteomes and their variation upon disease. Large scale proteomics today enables identification and measurement of changes of thousands of proteins from minute amount of tissues. Here we provide a comparative proteomic profile of three distinct parts of the murine nerve system: cerebellum, spinal cord, and sciatic nerve. Since membrane proteins are the key regulators of neural transmission and memory, our analysis is focused on this group of proteins. Rat tissues were homogenized and extracted to remove non membrane proteins and the resulting membranes were solubilized with detergents. Proteins were fractionated by size exclusion chromatography, depleted for detergents, digested and analyzed by LC-MS/MS using LTQ-Orbitrap Instrument. Applying stringent identification criteria in total 4,124 proteins were identified. Of these proteins 3,528, 3,290 and 1,649 were mapped to cerebellum, spinal cord and sciatic nerve, respectively. Our analysis allowed an in-depth mapping of neurotransmitter receptors, ion channels, and transporter proteins. This work is the most in-depth proteomic analysis of nerve tissues to date and provides the first unbiased insights into the proteomes of anatomically and functionally distinct parts of the membrane proteome of the central and peripheral nerve systems. The methods applied here can be directly applied to studying nerve systems and their disorders.

Introduction

Whereas traditional biochemical methods allow studying limited numbers of proteins, proteomics potentially can provide insights into the entire proteome. This modern approach has the potential to uncover differences between diseased and normal tissues in terms of single proteins, their posttranslational modifications, and protein-protein interactions as well as on a global scale such as analysis of signaling pathways and biological processes. In particular proteomic methods have frequently been applied to study brain, brain compartments, and other nerve tissue in respect to neural diseases[1]. Unfortunately, due to use of premature or inappropriate proteomic approaches the vast majority of these attempts provided only very incomplete pictures with limited value for understanding neural disorders. Technological limitations of the past were mainly of preparative origin including protein extraction, separation, and enzymatic digestion for mass spectrometric analysis. A case in point is the use of two-dimensional electrophoresis for protein separation and quantification. It has become apparent the 2D gels are particularly inadequate technique for studying tissues such as brain, providing biased datasets almost exclusively restricted to housekeeping proteins and with a very low content of membrane proteins such as ion channels and receptors [reviewed in [2]]. Owing to their unique properties, membrane proteins are likely to play an important role in etiology of diseases of nerve. Therefore techniques allowing detection and characterization of membrane proteins are a prerequisite for proteomic studies on neuronal disorders.

In parallel to 2DE based approaches, gel free approaches for proteomic analysis of membrane proteins, including their relative and absolute quantification, have been developed in recent years. Since many membrane proteins are low abundant, enrichment of membranes or isolation of specific type of membranes such as the plasma membrane is essential for an extensive analysis of these proteins. For this purpose proteomics specific extraction techniques were

developed for purification of total membranes[3, 4] and for plasma membrane enrichment[5, 6]. To quantify proteins, stable isotope tags such as ICAT and HysTag were developed[7, 8] and successfully applied for labeling of membrane proteins[8-10]. As an alternative way for relative quantification of mouse brain plasma membrane proteins label-free approaches were developed[11, 12]. Using a novel extraction and purification method and using "diagonal" reverse-phase peptide fractionation method[13] we have previously identified 1,700 proteins in mouse hippocampus[5]. Combining the plasma membrane purification method with the HysTag-labeling technique we have quantified 555 plasma membrane proteins between cortex, cerebellum and hippocampus of mouse brain. In another study membrane proteins from mouse fore- and hindbrain were analyzed using label-free technology, resulting in relative quantification of 967 proteins[14].

The nervous system of vertebrates consists of the central (CNS) and the peripheral nervous systems (PNS). Brain and spinal cord are the components of CNS. Whereas the vast majority of proteomics studies have been focused on profiling components of the entire brain or its parts such as hippocampus, cerebellum, and cortex, less attention has been paid to spinal cord[15-20] and studies attempting to analyze proteome of the PNS are scarce[21].

In this work we compared membrane proteomes of three distinct parts on the vertebrate nerve system: the cerebellum, spinal cord, and sciatic nerve. Our study provides the most comprehensive protein profile of nervous tissues to date. Applying stringent identification criteria, in total, 3,528, 3,290 and 1,649 proteins were identified from rat cerebellum, spinal cord and sciatic nerve, respectively. More than 70% of these proteins are annotated as membrane proteins. The in-house developed MaxQuant algorithms (Cox and Mann, submitted to Nature Biotechnology) allowed us to compare the abundance of the individual proteins between the three distinct parts of the nerve system. Our results provide an unprecedented depth of coverage of the nervous systems proteome. It allows comparative proteomic assessment, enabling insights into the membrane

proteomes of anatomically and functionally distinct parts of the central and peripheral nerve systems.

Materials and Methods

Membrane Proteins Extraction. The protocol applied here was slightly modified from the one published earlier in our group[5]. Briefly, 40-50 mg of frozen rat cerebellum, spinal cord, or sciatic nerve (Pel-freez Biologicals®, Rogers, AR) were blended in 1 ml of high salt buffer (2M NaCl, 10 mM HEPES-NaOH, pH 7.4, 1 mM EDTA) using IKA Ultra Turbax blender (IKA-Ultra Turrax, Staufen, Germany) at maximum speed for 20 seconds. The suspensions were centrifuged at 16,100 x g for 15 min in an Eppendorf 5145R centrifuge. The pellets were re-extracted twice in 1 ml of carbonate buffer (0.1 M Na_2CO_3, 1 mM EDTA). Subsequently the pellets were washed with 1 ml of urea buffer (4 M urea, 100 mM NaCl, 10 mM HEPES-NaOH, pH 7.4, and 1mM EDTA). After centrifugation the pellets were suspended in 100 µl 100mM CHAPS in PBS and were gently stirred at 20°C for 18h. After centrifugation, the pellets were re-extracted with 2% SDS in 0.1 M Na-phosphate buffer, pH 7.0 at 20°C for 18 h.

Size exclusion chromatography. Detergent dissolved proteins were chromatographed on Superdex 200 10/300GL column (GE bioscience) using Shimadzu UFLC chromatographic system (Kyoto, Japan). PBS containing 0.2% SDS was used as the mobile phase. Proteins were eluted at a flow rate of 250 µl/min and 400 µl fractions were collected.

Ethanol precipitation. One ml absolute ethanol was added to every 400 µl fraction. After 8h incubation at -20 °C precipitated proteins were collected by centrifugation at 16,000 x g at 4 °C. The pellets were washed 3 times with 70% ethanol and dried.

In solution digestion and peptides desalting. Detergent-free protein fractions were dissolved in 50 µl 8M urea and digested with 0.5 µg of Lys-C at room temperature for 4h. Subsequently, 150 µl of 50mM NH_4HCO_3 were added and

the digestion was continued room temperature in the presence 0.5 µg of trypsin (Promega, WI) for 4h. Digestion was terminated by adding 1% TFA and the tryptic peptides were desalted on RP-C_{18} StageTip columns[22].

Nano-HPLC-MS/MS. All Nano-HPLC-MS/MS experiments were performed using an Agilent 1100 nanoflow system (Agilent Technologies, Palo Alto, CA) connected to a linear ion trap-orbitrap hybrid mass spectrometer (LTQ-Orbitrap, Thermo Fisher Scientific, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark) as described before[23]. The reversed phase capillary emitter column was 15 cm length, 75 µm inner diameter, and in-house packed with 3 µm ReproSil-Pur C18-AQ media (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). The sample was separated by a 120 min gradient from 2% acetonitrile to 80% acetonitrile. Data acquisition on LTQ-orbitrap was operated in the data-dependent mode. Survey MS scans were acquired in the orbitrap at a resolution of 60,000. Up to the 7 most intense ions per scan were fragmented and detected in the linear ion trap. The online lock-mass option was utilized in order to improve mass accuracy[24].

Database searching and validation. The MS raw data was analyzed using the software MaxQuant[25] version 1.0.9.3, and then searched with the MASCOT engine against decoy IPI-rat database version 3.39. No fixed modification was applied, whereas protein N-terminal acetylation and methionine oxidation were set as variable modifications. Up to 2 missed cleavages were allowed for Lys-C digestion. The precursor ions were matched with an initial 7 ppm maximum mass deviation (MMD[26]) and the fragment ions had to match within an MMD of 0.4Da. The Mascot data were again imported into MaxQuant and parsed using the following criteria: (a) Peptide lenght of at least six amino acids and (b) the false discovery rate at both protein and peptides level lower than 0.01. “Match between runs” was ticked on during parsing, which means the precursor ions identification was transferred to unsequenced peptides based on elution time (±0.5 min) and accurate mass (within 2 X STD of the mass identified in other

runs) by comparing the same ion identified in other runs in which MS/MS fragmentation was performed. After the parsing, the proteins identified with at least two peptides, of which at least one was unique to the protein (or protein group) were considered to be genuine hits. The complete lists of identified proteins and peptides are in supplementary table 1 and 2, respectively.

Bioinformatics analysis. Protein localization and membrane and transmembrane domain annotation analyses were performed using Protein Center platform Version 1.3.3 (Proxeon Biosystems, Odense, Denmark) in which GO annotation and TMAP (http://emboss.sourceforge.net/apps/tmap.html) for transmembrane domain prediction are integrated. The KEGG overrepresentation analysis was performed by hypergeometric test using cytoscape[27] plugin BiNGO 2.0[28]. Briefly, a custom KEGG ontology containing mapping from rat EntrezGene identifiers to KEGG pathway identifiers was created for all the rat proteins, which served as the reference set. Subsequently, our identified proteome i.e. the test set was mapped to EntrezGene identifiers and analyzed for overrepresented pathways with respect to the reference set. Pathways with p-value < 0.01 were considered to be significant.

Results

There are many factors affecting the quality and the depth of a proteomic analysis. The main ones are the sample preparation, the mass spectrometric setup, and the data analysis. For the comparative proteomic study on cerebellum, spinal cord and sciatic nerve, we have applied a high resolution liquid chromatography tandem mass spectrometry system (LC-MS/MS) and the acquired high accuracy data were analyzed using the identification and quantification software MaxQuant developed in our group. As a reduction of sample complexity is a prerequisite for identification of proteins occurring at low abundance, we have used a three-step enrichment and fractionation protocol of membrane proteins. This protocol included purification of membranes, differential detergent-based extraction, and size exclusion chromatography of membrane proteins. The results provide first in-depth insights into the proteomes of three anatomically distinct parts of the central and peripheral nervous systems (CNC and PNS).

More than 4,000 proteins were identified in three tissues. The entire procedure for analysis of the nerve tissues is summarized in Fig.1. Crude membrane fractions were prepared by successive depletion of non-membrane proteins with high salt, sodium carbonate, and 4M urea[4]. Proteins were subsequently extracted from the membranes using CHAPS and SDS to achieve a partial separation of proteins according their solubility in the detergents. Each of the extracts was fractionated by means of size exclusion chromatography into 10 factions. After removal of detergent, proteins were digested with endoproteinase Lys-C and trypsin, and analyzed by LC-MS/MS.

For protein identifications we have applied very stringent identification criteria (see Methods). The accuracy of mass determination is exemplified in Fig. 2 for peptides originating from the cerebellum. The average mass deviation for all the peptides identified in cerebellum was -0.0688ppm, and the standard deviation

was 1.098 ppm. Similar values were observed for spinal cord and sciatic nerve peptides.

Using these criteria in total 4,124 proteins were identified. Of these 3,528, 3,290, and 1,649 were mapped to cerebellum, spinal cord, and rat sciatic nerve (Table 1, Fig. 3). A total of 70-75% of the proteins were annotated in GO-database as membrane proteins and 52-58% of the proteins have at least one predicted transmembrane domain (Table 1, Fig. 3). According to GO cellular compartment analysis proteins from plasma membrane, ER, and nucleus were most abundantly represented (up to 20%) (Table 1, Fig. 4).

GO molecular function and biological process analysis. GO molecular function and GO biological process distribution pattern of identified proteins were similar for each analyzed nerve tissues (Table 2). Figure 5A shows an example of the proteins identified from rat cerebellum. Approximately 22% of all identified proteins have catalytic activity function, followed by protein binding (21%), nucleotide binding (10%) and metal ion binding activity (9%). More interestingly, around 8% and 4% of proteins were annotated as transporters and receptors, respectively. The nature of these proteins is discussed in more detail in the sections below. GO biological process annotation analysis revealed that 14% and 12% proteins are involved in transport and cell communications, respectively (Figure 5B).

More than 100 genes were identified in at least two splice forms. For protein identification, only the proteins (group) with at least one unique peptide were accepted. Proteins containing the same identified peptides but no unique ones to differentiate each other were grouped together as a single identification. Using MaxQuant software we were able to differentiate at least two different protein isoforms from 129 genes (Supplementary Table 3, in total 276 proteins). Out of 260 annotated splice isoforms, 211 (81.2%) proteins were annotated as membrane proteins.

Neurotransmitter receptors. Neurotransmitter receptors can be grouped into two broad classes: the ligand-gated ion channels (ionotropic) and the G-protein coupled (metabotropic) receptors. Whereas ligand-gated ion channels mediate rapid postsynaptic responses, G protein-coupled receptors mediate slow postsynaptic responses. In our study, we identified in total 21 glutamate receptors, 13 γ-aminobutyric acid (GABA) receptors and 3 glycine receptors (Table 3).

Glutamate Receptors. L-Glutamate is the most abundant excitatory neurotransmitter in the mammalian nervous system. Ionotropic glutamate receptors appear to be tetrameric [29] or pentameric[30], and consist of four groups: N-methyl-D-aspartate (NMDA), α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA), kainate receptor, and glutamate receptor δ1 and 2. The metabotropic glutamate receptor (mGlur) family consists of eight different types of subunits: mGluR1 to mGluR8 that can be subdivided into groups I, II, and III based on receptor structure and physiological activity.

In rodents, two groups of NMDA receptor subunits were defined as NR1 (ζ1, *gene: Grin1*) and NR2A-D (ε1-4, *gene: Grin2a-d*). The NR1 protein can form homomeric channels whereas the NR2 subunits form functioning channel only as heterodimers with the NR1 subunit[31, 32]. In our study, we identified two splice variants of NR1, NR2A and NR2C in the rat cerebellum. The two splice variants of NR1 shares 11 identified peptides with one differentiating them from each other. In spinal cord, only one isoform of NR1 was identified. We failed to identify this group of proteins in the sciatic nerve, which may indicate relatively low expression levels in the nerve.

AMPA receptors are composed of four types of subunits, designated as GluR1 (*Gria1*), GluR2 (*Gria2*), GluR3 (*Gria3*), and GluR4 (*Gria4*). Most AMPA receptors are heterotetrameric, consisting of symmetric 'dimer of dimers' of GluR2 and either GluR1, GluR3 or GluR4 (ref[33]). Not only all four subunits but also two

splice isoforms for GluR1, GluR3 and GluR4 in were identified in rat cerebellum. In spinal cord, GluR2 and two splice forms of both GluR1 and GluR3 were identified. In sciatic nerve, GluR2, one splice form of GluR1 and two splice forms of GluR3 were identified.

There are five types of known kainate receptor subunits, GluR5 (*Grik1*), GluR6 (*Grik2*), GluR7 (*Grik3*), KA1 (*Grik4*) and KA2 (*Grik5*). In our study, GluR6 was identified in all three tissues, whereas GluR7 was identified in cerebellum and spinal cord.

The glutamate receptor, iontropic, δ1 (*Grid1*) and δ2 (*Grid2*) are another group of members of the family of ionotropic glutamate receptors. GluR-δ2 has been found to be expressed selectively in cerebellar Purkinje cells [34]. We identified the GluR-δ2 protein in all three tissues as well as GluR-δ2 interacting protein 1 (*Grid2ip*). GluR-δ1 was identified in the cerebellum and spinal cord.

Of the eight types of metabotropic glutamate receptors, we identified mGluR1, mGluR2, mGluR3 and mGluR 4 in all three tissues, mGluR7 in cerebellum and spinal cord, and mGluR5 in spinal cord only.

GABA receptors. GABA is the key inhibitory neurotransmitter in the vertebrate central nervous system. It acts through three classes of receptors: two types of ligand-gated $GABA_A$ and $GABA_C$ chloride channels and $GABA_B$ receptors that are coupled to separate K^+ or Ca^{2+} channels via G-proteins. Most $GABA_A$ receptors are assembled from subunits to form a hetero-pentameric structure, and in the rodent brain, they appear to be quite heterogeneous and comprise six different α-subunits (α1 - α6), three β-subunits (β1 - β3), three γ-subunits (γ1 - γ3), one δ-subunit and two more candidate ε-subunit and θ-subunit[35]. We identified α1, α2, α3, α5 and α6 in rat spinal cord, α1, α2 and α6 in cerebellum, and α1 and α6 in sciatic nerve. β1 - β3 were identified in cerebellum, whereas β1 and β3 were found in spinal cord, and β2 and β3 in sciatic nerve. γ1 and γ2 were

identified in both cerebellum and spinal cord, whereas only γ2 was identified in the sciatic nerve. The δ-subunit was also identified in the rat cerebellum and sciatic nerve. $GABA_B$ subunit-1 and 2 were identified in the three tissues.

Glycine receptor. Glycine receptors are widely distributed inhibitory ligand-gated ion channels in the central nervous system. Four α-subunits (α1 - α4) and a single β-subunit are known. In our study, α3and β subunits were identified in both the cerebellum and spinal cord, whereas α1 was only identified in the spinal cord.

Ion channels. Ion channels are pore-forming proteins that help to establish and control the voltage gradient across the plasma membrane of all living cells by allowing the flow of ions down their electrochemical gradient. Classified by the nature of their gating, ion channels can be divided into voltage-gated ion channels and ligand-gated ion channels. As the name indicats, voltage-gated ion channels activate/inactivate depending on the voltage gradient across the plasma membrane, whereas ligand-gated ion channels activate/inactivate depending on binding of ligands to the channel. In total, we identified 14 voltage-gated calcium ion channels, 11 voltage-gated potassium ion channels, 9 voltage-gated sodium ion channels and 10 chloride channels (Table 4). Furthermore, 12 other potassium ion channels belong to other subfamilies (supplementary table 1).

Voltage-dependent calcium channels. Voltage-dependent calcium channels are formed as a complex of several different subunits: α1, α2δ, β1-4, and γ. The α1 subunit has 24 putative transmembrane segments and forms the ion conducting pore with four homologous repeated domains while the other three subunits have several auxiliary functions including modulation of gating[36]. In our study, by using the nomenclature suggested by Ertel EA *et al*[37], out of 10 cloned α1 subunits, we identified $Ca_v2.1$ (*Cacna1a*, 2 isoforms), $Ca_v2.2$ (*Cacna1b*), $Ca_v2.3$ (*Cacna1e*) and $Ca_v3.1$ (*Cacna1g*) in rat cerebellum, $Ca_v2.1$ (1 isoform) and $Ca_v2.3$ in spinal cord and $Ca_v2.1$ (1 isoform) in sciatic nerve. Subunits α2δ1-3 (*Cacna2d1-3*) were

clearly identified in cerebellum and spinal cord, while only α2δ1-2 were clearly assigned to the sciatic nerve. Subunits β2 (*Cacnb2*) and β3 (*Cacnb3*) were only identified in rat cerebellum. The subunits β4 (*Cacnb4*) was identified in both cerebellum and spinal cord but we could not differentiate it from β1 subunits due to their high sequence similarity. Five peptides specifically different from other subunits but shared by them were assigned. In the last group of γ subunits, we identified γ2 (*Cacng2*), γ5 (*Cacng5*), γ8 (*Cacng8*) in the rat cerebellum. γ2 and γ8 subunits were identified in both spinal cord and sciatic nerve.

Voltage-dependent potassium channels. Potassium-selective channels are the largest and most diverse group of ion channels. Functional voltage-gated K_v channels are formed by either homotetramer or heterotetramer of K_v1, K_v7, and K_v10 families. K_v5, K_v6, K_v8, and K_v9 families encode subunits that act as modifiers, and β subunits, KCHIP1 (K_v4), calmodulin (K_v10), and minK (K_v11) serve as the accessory proteins associate with K_v tetramers and modify their properties [38]. We identified $K_v1.1$ (*Kcna1*), $K_v1.2$ (*Kcna2*), $K_v1.3$ (*Kcna3*), $K_v1.4$ (*Kcna4*), $K_v1.6$ (*Kcna6*), $K_v7.2$ (*Kcnq2*), $K_v10.1$ (*Kcnh1*) or $K_v10.2$ (*Kcnh5*) as the core channels to form the "pore" in rat cerebellum. No K_v modifier was identified in the study, but as accessory proteins we identified two isoforms of β2 protein, $K_v4.2$ (*Kcnd2*) and $K_v4.3$ (*Kcnd3*) in rat cerebellum. Subunits $K_v10.1$ (*Kcnh1*) and $K_v10.2$ (*Kcnh5*) cannot be clearly differentiated in cerebellum with unique peptides. All subunits were identified in cerebellum except $K_v10.1$ or $K_v10.2$, which were identified in spinal cord. Only two subunits Kv1.1 and Kv1.2 were found in the sciatic nerve.

Voltage-dependent sodium channels. The voltage-gated sodium channel in mammalian neuron is a large, multimeric complex, composed of a 260 kDa α subunit and one or more 33-36 kDa β subunits[39]. The α subunit is predicted to fold into four similar domains (I-IV) and each contains six α-helical transmembrane segments (S1-S6). The ion-conducting aqueous pore is contained entirely within the α subunit. The essential elements of sodium-channel

function, e.g. channel opening, ion selectivity and rapid inactivation can be demonstrated when α subunits are expressed alone in heterologous cells [40]. Coexpression of the β subunit modifies the kinetics and voltage-dependence of the gating of the channel and is required for full reconstitution of the properties of native sodium channels [40]. Five α subunits (Na_vα1.1(*Scn1a*), Na_vα1.2(*Scn2a1*), Na_vα1.4(*Scn4a*) or Na_vα1.5(*Scn5a*), Na_vα1.6(*Scn8a*) and an unknown Na_x sub-family (*Scn7a*)) and three β subunits (Na_vβ1(*Scn1b*), Na_vβ2(*Scn2b*), Na_vβ4(*Scn4b*)) were identified in rat cerebellum. In rat spinal cord, three α subunits (Na_vα1.1, Na_vα1.6 and Na_x) and four β subunits (Na_vβ1, Na_vβ2, Na_vβ3 (*Scn3b*) and Na_vβ4) were identified. In the rat sciatic nerve, one α subunit (Na_x) and two β subunits (Na_vβ2 and Na_vβ4) were found.

Chloride channels. Chloride channels consist of approximately 13 members and are important for setting cell resting membrane potential and maintaining proper cell volume. Unlike the specific ion transmission of the cation channel described before, these channels conduct Cl^- as well as other anions such as HCO^{3-}, I^-, SCN^-, and NO^{3-}. The chloride channel proteins identified in cerebellum include chloride channel protein 2 (*Clcn2*), 3 (*Clcn3*), 4 (*Clcn4-2*), 5 (*Clcn5*), 6 (*Clcn6*), 7 (*Clcn7*), chloride intracellular channel protein 1 (*Clic1*), 4 (*Clic4*) and chloride channel CLIC-like protein 1 (*Clcc1*). Besides the proteins listed above, one long isoform encoding a form of the *Clcn3* gene was also identified in the spinal cord. In the sciatic nerve, only chloride channel protein 6 and chloride intracellular channel protein 1 and 4 were identified.

Multidrug resistance proteins (MRPs) and ATP-binding cassette transporters (ABC). The ABC genes represent the largest family of transmembrane proteins. These proteins are classified as ABC transporters based on the sequence and organization of their ATP-binding domains, also known as nucleotide-binding folds NBFs[41]. Since the first discovery in 1976 (ref[42]) that it was possible for one of the ABC transporter P-glycoprotein (P-gp/MDR1, *Abcb1*) to confer resistance to a relatively large number of structurally diverse

drugs with different mechanisms of action, 11 more ABC transporters were found to be involved in drug resistance, including MPR1-6 (*Abcc1-6*), MRPs 7-9 (*Abcc 10-12*), MRP10 (*Abcc13*) and ABCG2/BCRP (*Abcg2*)[41,43]. Out of 27 identified ABC transporters in our study (Table 5), 5 multidrug resistance proteins, including P-gp/MDR1 (*Abcb1*), MDR2 (A*bcb4*), MRP1 (*Abcc1*), MRP4 (*Abcc4*) and ABCG2/BCRP (*Abcg2*) were identified. All five proteins were found in spinal cord, and all except MDR2 were found in cerebellum. Only MRP4 was identified in sciatic nerve.

Solute carrier. The SoLute Carrier (SLC) group includes over 300 members organized into 47 families[44]. In our study, we identified in total 112 solute carrier subunits/isoforms covering 33 families (supplementary table 4).

Long-term potentiation and Long-term depression. Besides analysis on different protein families as describe above, we used Bingo 2.0 to determine KEGG pathways that are overrepresented in all tissue samples by a significance level p-value less than 0.01 (Table 5) . In total 24 pathways are overrepresented in the identified protein cluster, in which some could be artificially overrepresented because of the biased membrane fraction enrichment during sample preparation, e.g. ECM-receptor interaction, gap junction and tight junction, while some are nerve tissue specific overrepresentation, such as long-term potentiation [45] and long-term depression. Long-term depression is the opposite process of long-term potentiation, and both are considered the major cellular mechanisms underlying learning and memory[46]. In the rat IPI database v3.39, 62 proteins/subunits are annotated to be involved in long-term potentiation and 71 in long-term depression. We identified 46 and 41 proteins involved in two pathways, respectively, in rat cerebellum (Table 6), which already coveres the majority of the pathway diagram (Fig.6).

Discussions

Membrane proteins are key regulators of vital processes including the passage of information and substances between cells and mediating activities such as hormone action and nerve impulses. More than 30% of the mammalian genome encodes membrane proteins. These proteins are challenging to study, but pivotal to understand, as they represent two thirds of drug targets[47]. Studies on these proteins can potentially lead to novel and improved pharmaceutical treatments for broad range of diseases including disorders of the nerve system. Mapping of membrane proteomes and identification of drug targets requires methods for tissue extraction, protein processing, and mass spectrometric analysis. Taking advantage of our recently developed methods for large scale identification of membrane proteins[4, 5, 48] and our proteomic platform allowing high accuracy and confidential protein identification and relative quantification[25] we have performed an in-depth comparative analysis of membrane proteomes of rat cerebellum, spinal cord, and sciatic nerve.

This work provides information on in total 4,124 proteins, of which 3,528, 3,290 and 1,649 proteins were mapped to cerebellum, spinal cord and sciatic nerve, respectively. From three tissues, 2,223, 2,071 and 1,009 proteins were annotated as membrane proteins, and 2,040, 1,891 and 856 proteins have at least one transmembrane domain. Around 20% of the proteins from each tissue were GO localized to plasma membrane. GO cellular compartment analysis, GO molecular function analysis and GO biological process analysis show quite similar protein expression patterns in the three nerve tissues. In our study, we identified 22 glutamate receptors, 13 GABA receptors and 3 glycine receptors. Furthermore, we identified 56 different subunits involved in the organization of calcium, potassium, sodium ion, and chloride ion channels.

Solute carrier (SLC) and ATP-binding cassette transporter (ABC) are the major gene superfamilies that play essential roles in the transport of material including

drugs across plasma and other biological membranes. This study revealed occurrence of 112 SLC and 27 ABC members in the nerve system.

ABC transporter related drug resistance is a challenging obstacle in the treatment of cancer. So far, substrates for 7 multidrug resistance proteins were identified in various cell systems[41]. Of these, four were identified in this study as integral components of nerve tissue.

The depth of our study can also be judged by KEGG ontology and Bingo analysis that resulted in the identification of 24 pathways that are overrepresented. A high number of proteins implicated in individual pathways has been identified. These include protein clusters involved in long-term potentiation and depression. Thus, our approach can be useful for studying processes of learning and memory at the protein level.

To our knowledge, this work provides the most comprehensive proteomic analysis of nerve tissues published to date. For the first time proteomes of anatomically and functionally distinct parts of the nerve system, including its central and peripheral components were compared. Our approach can be taken as a starting point for future proteomic study of the nerve system, its abnormalities and illnesses.

Acknowledgements: We would like to thank Mr. Chanchal Kumar for helpful advices in the bioinformatics analysis of the data.

Supporting information available: Four supplemental tables that support the data presented are provided. These include: Supplemental Table 1, list of all proteins identified in cerebellum, spinal cord and sciatic nerve; Supplemental Table 2, list of all peptides identified for each protein; Supplemental Table 3, list of gene for which at least two splicing isoforms are identified; Supplemental Table 4, list of solute carriers identified in three tissues. These material is freely available via the Internet at http://pubs.acs.org.

References

1. Vercauteren, F. G.; Bergeron, J. J.; Vandesande, F.; Arckens, L.; Quirion, R., Proteomic approaches in brain research and neuropharmacology. *Eur J Pharmacol* **2004,** 500, (1-3), 385-98.
2. Grant, K. J.; Wu, C. C., Advances in neuromembrane proteomics: efforts towards a comprehensive analysis of membrane proteins in the brain. *Brief Funct Genomic Proteomic* **2007,** 6, (1), 59-69.
3. Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R., 3rd, A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* **2003,** 21, (5), 532-8.
4. Nagaraj, N., Lu, A., Mann, M., Wiśniewski, J.R., Detergent-based but Gel-free Method Allows Identification of Several Hundred Membrane Proteins in Single MS Runs. *J. Proteome. Res.* **2008**.
5. Nielsen, P. A.; Olsen, J. V.; Podtelejnikov, A. V.; Andersen, J. R.; Mann, M.; Wiśniewski, J. R., Proteomic mapping of brain plasma membrane proteins. *Mol Cell Proteomics* **2005,** 4, (4), 402-8.
6. Schindler, J.; Lewandrowski, U.; Sickmann, A.; Friauf, E.; Nothwang, H. G., Proteomic analysis of brain plasma membranes isolated by affinity two-phase partitioning. *Mol Cell Proteomics* **2006,** 5, (2), 390-400.
7. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999,** 17, (10), 994-9.
8. Olsen, J. V.; Andersen, J. R.; Nielsen, P. A.; Nielsen, M. L.; Figeys, D.; Mann, M.; Wiśniewski, J. R., HysTag--a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. *Mol Cell Proteomics* **2004,** 3, (1), 82-92.
9. Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R., Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* **2001,** 19, (10), 946-51.
10. Olsen, J. V.; Nielsen, P. A.; Andersen, J. R.; Mann, M.; Wiśniewski, J. R., Quantitative proteomic profiling of membrane proteins from the mouse brain cortex, hippocampus, and cerebellum using the HysTag reagent: mapping of neurotransmitter receptors and ion channels. *Brain Res* **2007,** 1134, (1), 95-106.
11. Stewart, II; Zhao, L.; Le Bihan, T.; Larsen, B.; Scozzaro, S.; Figeys, D.; Mao, G. D.; Ornatsky, O.; Dharsee, M.; Orsi, C.; Ewing, R.; Goh, T., The reproducible acquisition of comparative liquid chromatography/tandem mass spectrometry data from complex biological samples. *Rapid Commun Mass Spectrom* **2004,** 18, (15), 1697-710.
12. Bukhman, Y. V.; Dharsee, M.; Ewing, R.; Chu, P.; Topaloglou, T.; T, L. E. B.; Goh, T.; Duewel, H.; Stewart, II; Wiśniewski, J. R.; Ng, N. F., Design and analysis of quantitative differential proteomics investigations using lc-ms technology. *J Bioinform Comput Biol* **2008,** 6, (1), 107-23.
13. Kristensen, D. B.; Brond, J. C.; Nielsen, P. A.; Andersen, J. R.; Sorensen, O. T.; Jorgensen, V.; Budin, K.; Matthiesen, J.; Veno, P.; Jespersen, H. M.; Ahrens, C. H.; Schandorff, S.; Ruhoff, P. T.; Wisniewski, J. R.; Bennett, K. L.; Podtelejnikov, A. V., Experimental Peptide Identification Repository (EPIR): an integrated peptide-centric

platform for validation and mining of tandem mass spectrometry data. *Mol Cell Proteomics* **2004,** 3, (10), 1023-38.
14. Le Bihan, T.; Goh, T.; Stewart, II; Salter, A. M.; Bukhman, Y. V.; Dharsee, M.; Ewing, R.; Wiśniewski, J. R., Differential analysis of membrane proteins in mouse fore- and hindbrain using a label-free approach. *J Proteome Res* **2006,** 5, (10), 2701-10.
15. Butt, R. H.; Pfeifer, T. A.; Delaney, A.; Grigliatti, T. A.; Tetzlaff, W. G.; Coorssen, J. R., Enabling coupled quantitative genomics and proteomics analyses from rat spinal cord samples. *Mol Cell Proteomics* **2007,** 6, (9), 1574-88.
16. Dreger, M.; Mika, J.; Bieller, A.; Jahnel, R.; Gillen, C.; Schaefer, M. K.; Weihe, E.; Hucho, F., Analysis of the dorsal spinal cord synaptic architecture by combined proteome analysis and in situ hybridization. *J Proteome Res* **2005,** 4, (2), 238-49.
17. Ekegren, T.; Hanrieder, J.; Aquilonius, S. M.; Bergquist, J., Focused proteomics in post-mortem human spinal cord. *J Proteome Res* **2006,** 5, (9), 2364-71.
18. Kunz, S.; Tegeder, I.; Coste, O.; Marian, C.; Pfenninger, A.; Corvey, C.; Karas, M.; Geisslinger, G.; Niederberger, E., Comparative proteomic analysis of the rat spinal cord in inflammatory and neuropathic pain models. *Neurosci Lett* **2005,** 381, (3), 289-93.
19. Yu, L. R.; Conrads, T. P.; Uo, T.; Kinoshita, Y.; Morrison, R. S.; Lucas, D. A.; Chan, K. C.; Blonder, J.; Issaq, H. J.; Veenstra, T. D., Global analysis of the cortical neuron proteome. *Mol Cell Proteomics* **2004,** 3, (9), 896-907.
20. Wang, H.; Qian, W. J.; Chin, M. H.; Petyuk, V. A.; Barry, R. C.; Liu, T.; Gritsenko, M. A.; Mottaz, H. M.; Moore, R. J.; Camp Ii, D. G.; Khan, A. H.; Smith, D. J.; Smith, R. D., Characterization of the mouse brain proteome using global proteomic analysis complemented with cysteinyl-peptide enrichment. *J Proteome Res* **2006,** 5, (2), 361-9.
21. Jimenez, C. R.; Stam, F. J.; Li, K. W.; Gouwenberg, Y.; Hornshaw, M. P.; De Winter, F.; Verhaagen, J.; Smit, A. B., Proteomics of the injured rat sciatic nerve reveals protein expression dynamics during regeneration. *Mol Cell Proteomics* **2005,** 4, (2), 120-32.
22. Rappsilber, J.; Ishihama, Y.; Mann, M., Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **2003,** 75, (3), 663-70.
23. Olsen, J. V.; Mann, M., Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci U S A* **2004,** 101, (37), 13417-22.
24. Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M., Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **2005,** 4, (12), 2010-21.
25. Cox, J.; Mann, M., Is proteomics the new genomics? *Cell* **2007,** 130, (3), 395-8.
26. Zubarev, R.; Mann, M., On the proper use of mass accuracy in proteomics. *Mol Cell Proteomics* **2007,** 6, (3), 377-81.
27. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003,** 13, (11), 2498-504.

28. Maere, S.; Heymans, K.; Kuiper, M., BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **2005,** 21, (16), 3448-9.
29. Laube, B.; Kuhse, J.; Betz, H., Evidence for a tetrameric structure of recombinant NMDA receptors. *J Neurosci* **1998,** 18, (8), 2954-61.
30. Ferrer-Montiel, A. V.; Montal, M., Pentameric subunit stoichiometry of a neuronal glutamate receptor. *Proc Natl Acad Sci U S A* **1996,** 93, (7), 2741-4.
31. Kutsuwada, T.; Kashiwabuchi, N.; Mori, H.; Sakimura, K.; Kushiya, E.; Araki, K.; Meguro, H.; Masaki, H.; Kumanishi, T.; Arakawa, M.; et al., Molecular diversity of the NMDA receptor channel. *Nature* **1992,** 358, (6381), 36-41.
32. Monyer, H.; Sprengel, R.; Schoepfer, R.; Herb, A.; Higuchi, M.; Lomeli, H.; Burnashev, N.; Sakmann, B.; Seeburg, P. H., Heteromeric NMDA receptors: molecular and functional distinction of subtypes. *Science* **1992,** 256, (5060), 1217-21.
33. Greger, I. H.; Ziff, E. B.; Penn, A. C., Molecular determinants of AMPA receptor subunit assembly. *Trends Neurosci* **2007,** 30, (8), 407-16.
34. Landsend, A. S.; Amiry-Moghaddam, M.; Matsubara, A.; Bergersen, L.; Usami, S.; Wenthold, R. J.; Ottersen, O. P., Differential localization of delta glutamate receptors in the rat cerebellum: coexpression with AMPA receptors in parallel fiber-spine synapses and absence from climbing fiber-spine synapses. *J Neurosci* **1997,** 17, (2), 834-42.
35. Pirker, S.; Schwarzer, C.; Wieselthaler, A.; Sieghart, W.; Sperk, G., GABA(A) receptors: immunocytochemical distribution of 13 subunits in the adult rat brain. *Neuroscience* **2000,** 101, (4), 815-50.
36. Dolphin, A. C., A short history of voltage-gated calcium channels. *Br J Pharmacol* **2006,** 147 Suppl 1, S56-62.
37. Ertel, E. A.; Campbell, K. P.; Harpold, M. M.; Hofmann, F.; Mori, Y.; Perez-Reyes, E.; Schwartz, A.; Snutch, T. P.; Tanabe, T.; Birnbaumer, L.; Tsien, R. W.; Catterall, W. A., Nomenclature of voltage-gated calcium channels. *Neuron* **2000,** 25, (3), 533-5.
38. Gutman, G. A.; Chandy, K. G.; Grissmer, S.; Lazdunski, M.; McKinnon, D.; Pardo, L. A.; Robertson, G. A.; Rudy, B.; Sanguinetti, M. C.; Stuhmer, W.; Wang, X., International Union of Pharmacology. LIII. Nomenclature and molecular relationships of voltage-gated potassium channels. *Pharmacol Rev* **2005,** 57, (4), 473-508.
39. Catterall, W. A., From ionic currents to molecular mechanisms: the structure and function of voltage-gated sodium channels. *Neuron* **2000,** 26, (1), 13-25.
40. Yu, F. H.; Catterall, W. A., Overview of the voltage-gated sodium channel family. *Genome Biol* **2003,** 4, (3), 207.
41. Dean, M.; Rzhetsky, A.; Allikmets, R., The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res* **2001,** 11, (7), 1156-66.
42. Juliano, R. L.; Ling, V., A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochim Biophys Acta* **1976,** 455, (1), 152-62.
43. Deeley, R. G.; Westlake, C.; Cole, S. P., Transmembrane transport of endo- and xenobiotics by mammalian ATP-binding cassette multidrug resistance proteins. *Physiol Rev* **2006,** 86, (3), 849-99.
44. Hediger, M. A.; Romero, M. F.; Peng, J. B.; Rolfs, A.; Takanaga, H.; Bruford, E. A., The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction. *Pflugers Arch* **2004,** 447, (5), 465-8.

45. Lomo, T., The discovery of long-term potentiation. *Philos Trans R Soc Lond B Biol Sci* **2003,** 358, (1432), 617-20.
46. Cooke, S. F.; Bliss, T. V., Plasticity in the human central nervous system. *Brain* **2006,** 129, (Pt 7), 1659-73.
47. Drews, J., Drug discovery: a historical perspective. *Science* **2000,** 287, (5460), 1960-4.
48. Wiśniewski, J. R., Protocol to enrich and analyze plasma membrane proteins from frozen tissues. *Methods Mol Biol* **2008,** 432, 175-83.

Figure legends

Figure 1. Flow chart of membrane fraction preparation and MS analysis. Tissue samples were first homogenized in high salt buffer, the pellets were then washed twice with high pH Na_2CO_3 buffer and finally once with 4M Urea buffer. The crude membrane pellets were first dissolved overnight in 100 µl 100 mM CHAPS (Supernatant 1), and then dissolved again overnight in 100 µl 2% SDS buffer (Supernatant 2). Both supernatants were separated on size exclusion chromatography column into 10 fractions. The 20 fractions for each tissue sample were precipitated with 70% ethanol to remove detergent SDS. Pure membrane proteins were then redissolved in 50 mM NH_4HCO_3 buffer for overnight trypsin digestion. After desalting with C_{18} StageTips, peptides were loaded for online HPLC for mass spectrometry analysis.

Figure 2. Mass Error (ppm) of the precursor ions for identified peptides from cerebellum. Proteins were identified with accurate precursor ions with the cutoff of maximum 7ppm mass deviation for the Mascot search. The average mass deviation for all the peptides identified in cerebellum was -0.0688 ppm, and the standard deviation was 1.098 ppm. (The average and STD for spinal cord were -0.0061 ppm and 0.971 ppm respectively and the values for sciatic nerve were -0.0956ppm and 1.418ppm). The plot was performed separately for cerebellum only since the total identified peptides number is too large.

Figure 3. Proteins identified in different tissues. Three bars of each tissue from left to right indicate the number of all identified proteins, membrane proteins, and proteins with at least one transmembrane domain.

Figure 4. Subcellular localization of all annotated proteins from rat cerebellum (A), spinal cord (B) and sciatic nerve (C). *PM*, plasma membrane; *VES*, cytoplasmic membrane-bound vesicles; *ER*, endoplasmatic reticulum; *MIT*, mitochondria; *GOL*, Golgi apparatus; *NUC*, nucleus.

Figure 5. Molecular function and biological process GO annotation of identified proteins from rat cerebellum.

Figure 6. Pathway diagrams of long-term potentiation and long-term depression (retrieved from http://www.genome.jp/kegg/tool/search_pathway.html). Proteins in green-filled box indicate the presence of genes in the rat genome, and the ones with red box outline and red font name were identified in rat cerebellum. The genes for the proteins in grey boxes were not identified in the rat genome.

Table 1. Summary of the identification of proteins and peptides in the nerve tissues[a].

Protein category	Number of identifications		
	Cerebellum	Spinal Cord	Sciatic Nerve
Non-redundant peptides	29,597	25,199	13,992
Non-redundant proteins	3,528	3,290	1,649
Annotated proteins	2,977	2,780	1,449
Membrane proteins	2,223	2,071	1,009
Proteins ≥1 TM domain	2,040	1,891	856
Cellular localization			
Plasma membrane	476	474	295
Cytoplasmic vesicle	195	198	113
Endoplasmic reticulum	280	272	144
Mitochondrium	497	476	270
Golgi apparatus	227	221	105
Nucleus	516	453	248

[a]The entire lists of the identified proteins and peptides are in the Supplementary Tables 1 and 2).

Table 2. GO molecular function and biological process.

	Cerebellum		Spinal Cord		Sciatic Nerve	
	Protein Number	% of all identified proteins	Protein Number	% of all identified proteins	Protein Number	% of all identified proteins
Molecular Function						
Unannotated	429	12.2	392	11.9	144	8.7
catalytic activity	1,738	49.3	1,614	49.1	825	50
enzyme regulator activity	229	6.5	198	6	99	6
metal ion binding	716	20.3	668	20.3	346	21
motor activity	120	3.4	114	3.5	85	5.2
nucleic acid binding	421	11.9	368	11.2	206	12.5
nucleotide binding	769	21.8	703	21.4	398	24.1
protein binding	1,723	48.8	1,605	48.8	866	52.5
receptor activity	313	8.9	294	8.9	121	7.3
signal transducer activity	478	13.5	444	13.5	205	12.4
structural molecule activity	349	9.9	341	10.4	268	16.3
transcription regulator activity	124	3.5	122	3.7	58	3.5
translation regulator activity	36	1	31	0.9	16	1
transporter activity	637	18.1	600	18.2	308	18.7
Biological Process						
Unannotated	623	17.7	568	17.3	235	14.3
behavior	108	3.1	102	3.1	48	2.9
cell communication	958	27.2	861	26.2	422	25.6
cell death	218	6.2	214	6.5	119	7.2
cell division	33	0.9	26	0.8	19	1.2
cell growth	25	0.7	25	0.8	13	0.8

cell homeostasis	121	3.4	111	3.4	69	4.2
cell motility	157	4.5	155	4.7	87	5.3
cell organization and biogenesis	925	26.2	887	27	493	29.9
cell proliferation	162	4.6	159	4.8	93	5.6
coagulation	27	0.8	25	0.8	27	1.6
defense response	125	3.5	126	3.8	75	4.5
development	719	20.4	684	20.8	396	24
metabolism	1,900	53.9	1,785	54.3	959	58.2
regulation of biological process	878	24.9	819	24.9	433	26.3
sensory perception	66	1.9	64	1.9	36	2.2
transport	1,109	31.4	1,056	32.1	570	34.6

Table 3. Neurotransmitter receptors identified in different tissues

Gene Name	Accessory Number	TM[a]	Protein Name	Number of Peptides					
				cerebellum		Spinal Cord		Sciatic Nerve	
				All	Unique[b]	All	Unique	All	Unique
A. Glutamate receptors									
	Ionotropic NMDA receptors								
Grin1	IPI00198625	4	Isoform A/B/C/D/F of Glutamate [NMDA] receptor subunit zeta-1	11	1	5	0	0	0
Grin1	IPI00231257	4	Isoform E/G of Glutamate [NMDA] receptor subunit zeta-1	11	1	6	1	0	0
Grin2a	IPI00326054	7	Glutamate [NMDA] receptor subunit epsilon-1	6	6	3	3	0	0
Grin2c	IPI00201739	5	N-methyl-D-aspartate receptor NMDAR2C subunit	4	4	0	0	0	0
	Ionotropic AMPA receptors								
Gria1	IPI00231012	4	Isoform Flip of Glutamate receptor 1	23	2	7	1	7	2
Gria1	IPI00324555	4	Isoform Flop of Glutamate receptor 1	22	1	8	1	6	0
Gria2	IPI00231061	3	Isoform Flip of Glutamate receptor 2	29	1	18	1	7	1
Gria3	IPI00195443	3	Isoform Flop of Glutamate receptor 3	22	1	13	1	4	1
Gria3	IPI00231095	3	Isoform Flip of Glutamate receptor 3	22	2	13	2	4	2
Gria4	IPI00231131	3	Isoform 2 of Glutamate receptor 4	23	2	9	0	6	0
Gria4	IPI00231132	3	Isoform 3 of Glutamate receptor 4	22	1	10	0	7	0
	Ionotropic Kainate receptors								
Grik2	IPI00324708	3	Glutamate receptor, ionotropic kainate 2	7	7	3	3	1	1
Grik3	IPI00230977	4	Isoform GluR7A/B of Glutamate receptor, ionotropic kainate 3	3	3	2	2	0	0
	Other ionotropic receptors								
Grid1	IPI00207091	4	Glutamate receptor delta-1 subunit	8	7	5	4	1	0

Grid2	IPI00206854	3	Glutamate receptor delta-2 subunit	40	39	18	17	23	22
Grid2ip	IPI00870890	0	glutamate receptor, ionotropic, delta 2 (Grid2) interacting protein 1	22	22	1	1	7	7
	Metabotropic receptors								
Grm1	IPI00210260	6	Isoform 1A/B/C of Metabotropic glutamate receptor 1	25	25	5	5	7	7
Grm2	IPI00212618	6	Metabotropic glutamate receptor 2	12	9	3	2	4	3
Grm3	IPI00769125	9	Metabotropic glutamate receptor 3	8	6	6	5	2	1
Grm4	IPI00327692	7	Metabotropic glutamate receptor 4	12	11	2	2	1	1
Grm5	IPI00212621	6	Isoform 1/2 of Metabotropic glutamate receptor 5	0	0	2	2	0	0
Grm7	IPI00198587	8	Metabotropic glutamate receptor 7	2	2	2	2	0	0
B. GABA receptors									
	GABA A receptors								
Gabra1	IPI00192642	2	Gamma-aminobutyric acid receptor subunit alpha-1	13	6	9	4	4	2
Gabra2	IPI00679252	2	Gamma-aminobutyric acid receptor subunit alpha-2	7	2	6	2	2	0
Gabra3	IPI00197343	2	Gamma-aminobutyric acid receptor subunit alpha-3	4	0	6	3	2	0
Gabra5	IPI00325359	3	Gamma-aminobutyric acid receptor subunit alpha-5	4	0	4	1	1	0
Gabra6	IPI00206049	4	Gamma-aminobutyric acid receptor subunit alpha-6	12	12	1	1	7	7
Gabrb1	IPI00209268	3	Gamma-aminobutyric acid receptor subunit beta-1	8	2	7	2	4	0
Gabrb2	IPI00209269	2	Gamma-aminobutyric acid receptor subunit beta-2	16	8	5	0	5	1
Gabrb3	IPI00327083	2	Gamma-aminobutyric acid receptor subunit beta-3	11	5	7	3	5	2
Gabrd	IPI00192644	3	Gamma-aminobutyric acid receptor subunit	8	8	0	0	3	3

			delta						
Gabrg1	IPI00211960	2	Gamma-aminobutyric acid receptor subunit gamma-1	3	2	3	2	1	0
Gabrg2	IPI00192646	2	Gamma-aminobutyric acid receptor subunit gamma-2	4	3	3	2	2	1
GABA B receptors									
Gabbr1	IPI00208182	7	Gamma-aminobutyric acid type B receptor subunit 1	9	9	6	6	3	3
Gabbr2	IPI00331966	7	Gamma-aminobutyric acid type B receptor subunit 2	13	13	4	4	2	2
C. Glycine receptors									
Glra1	IPI00206855	4	Isoform a/b of Glycine receptor subunit alpha-1	0	0	7	6	0	0
Glra3	IPI00392219	4	Glycine receptor alpha 3	1	1	2	1	0	0
Glrb	IPI00202491	5	Glycine receptor subunit beta	1	1	10	10	0	0

[a]**TM:** number of predicted transmembrane domain by TMAP algorithm; [b] unique peptide is the peptide with a sequence specific to the protein but not shared with other proteins

Table 4. Ion channels identified in different tissues

Gene Name	Accessory Number	TM[a]	Protein Name	Number of Peptides					
				Cerebellum		Spinal Cord		Sciatic Nerve	
				All	Unique[b]	All	Unique	All	Unique
Voltage-gated Ca^{2+} channels									
Cacna1a	IPI00197594	4	Class A calcium channel variant riA-I (Fragment)	6	1	0	0	1	0
Cacna1a	IPI00211870	23	calcium channel, voltage-dependent, P/Q type, alpha 1A subunit	19	13	3	2	3	1
Cacna1b	IPI00200639	24	Voltage-dependent N-type calcium channel subunit alpha-1B	4	4	0	0	0	0
Cacna1e	IPI00198751	23	Voltage-dependent R-type calcium channel subunit alpha-1E	7	6	2	1	1	0
Cacna1g	IPI00196758	22	Voltage-dependent calcium channel T type alpha 1G subunit	6	6	0	0	0	0
Cacna2d1	IPI00391769	3	Voltage-dependent calcium channel subunit alpha-2/delta-1	35	34	30	30	3	3
Cacna2d2	IPI00191089	3	Voltage-dependent calcium channel subunit alpha-2/delta-2	25	24	17	17	4	4
Cacna2d3	IPI00191088	1	Voltage-dependent calcium channel subunit alpha-2/delta-3	10	10	9	9	0	0
Cacnb2	IPI00421663	0	Voltage-dependent L-type calcium channel subunit beta-2	4	1	2	0	1	0
Cacnb3	IPI00211876	1	Voltage-dependent L-type calcium channel subunit beta-3	3	1	2	0	1	0
Cacnb1 or Cacnb4	IPI00211872 IPI00768297	0	Voltage-dependent L-type calcium channel subunit beta-1/beta-4	8	5	5	3	1	0
Cacng2	IPI00201313	3	Voltage-dependent calcium channel gamma-2 subunit	8	8	7	7	2	2
Cacng5	IPI00207430	3	Voltage-dependent calcium channel gamma-5 subunit	2	2	0	0	0	0
Cacng8	IPI00207426	4	Voltage-dependent calcium channel	3	3	4	4	2	2

			gamma-8 subunit						
Voltage-gated K^+ channels (Kv)									
Kcna1	IPI00190644	7	Potassium voltage-gated channel subfamily A member 1	13	7	12	6	8	4
Kcna2	IPI00208365	7	Potassium voltage-gated channel subfamily A member 2	8	2	8	2	6	2
Kcna3	IPI00208359	7	Potassium voltage-gated channel subfamily A member 3	7	1	7	1	4	0
Kcna4	IPI00208362	7	Potassium voltage-gated channel subfamily A member 4	2	1	4	3	1	0
Kcna6	IPI00190053	7	Potassium voltage-gated channel subfamily A member 6	4	3	5	4	1	0
Kcnab2	IPI00211012	0	Voltage-gated potassium channel subunit beta-2	7	2	10	2	3	0
Kcnab2	IPI00780996	0	47 kDa protein	6	1	9	1	3	0
Kcnd2	IPI00394218	6	Potassium voltage-gated channel subfamily D member 2	10	9	4	3	1	0
Kcnd3	IPI00389372	5	Potassium voltage-gated channel subfamily D member 3	5	4	3	2	1	0
Kcnh1 or Kcnh5	IPI00339126 IPI00189118	4	Potassium voltage-gated channel subfamily H member 1 or 5	3	3	0	0	0	0
Kcnq2	IPI00214278	6	Potassium voltage-gated channel subfamily KQT member 2	2	2	2	2	0	0
Voltage-gated Na^+ channels									
Scn1a	IPI00198841	23	Sodium channel protein type 1 subunit alpha	23	7	12	5	4	0
Scn2a1	IPI00400699	23	Sodium channel protein type 2 subunit alpha	46	1	9	0	8	0
Scn4a or Scn5a	IPI00339065	26	Sodium channel protein type 4/5 subunit alpha	8	1	4	0	3	0

Scn7a	IPI00326646	24	Sodium channel;Glial voltage-gated sodium channel	2	2	5	5	8	8
Scn8a	IPI00213193	25	Voltage-gated sodium channel variant rPN4a	11	2	6	2	1	0
Scn1b	IPI00204294	1	Sodium channel subunit beta-1	8	8	5	5	0	0
Scn2b	IPI00189882	1	Sodium channel subunit beta-2	7	7	3	3	1	1
Scn3b	IPI00201894	1	Sodium channel subunit beta-3	0	0	2	2	0	0
Scn4b	IPI00369414	1	Sodium channel subunit beta-4	5	5	5	5	1	1
Voltage-gated chloride channels (CLC)									
Clcn2	IPI00199560	11	Chloride channel protein 2	6	6	6	6	0	0
Clcn3	IPI00476274	10	Chloride channel protein 3	7	1	4	1	0	0
Clcn3	IPI00566593	10	Chloride channel protein 3 long form (Fragment)	6	0	4	1	0	0
Clcn4-2	IPI00205423	11	chloride channel 4	6	5	5	4	0	0
Clcn5	IPI00326242	11	Chloride channel protein 5	2	1	4	3	0	0
Clcn6	IPI00369827	10	chloride channel 6	11	11	10	10	2	2
Clcn7	IPI00205428	10	Chloride channel protein 7	1	1	2	2	0	0
Chloride intracellular channels									
Clic1	IPI00421995	1	Chloride intracellular channel 1	1	1	2	2	4	4
Clic4	IPI00208249	1	Chloride intracellular channel protein 4	2	2	3	3	1	1
Clcc1	IPI00203434	3	Chloride channel CLIC-like protein 1	1	1	2	2	0	0

[a]**TM:** number of predicted transmembrane domain by TMAP algorithm; [b] unique peptide is the peptide with a sequence specific to the protein but not shared with other proteins

Table 5. ABC transporters

Gene Name	Accessory Number	TM[a]	Protein Name	Number of Peptides					
				Cerebellum		Spinal Cord		Sciatic Nerve	
				All	Unique[b]	All	Unique	All	Unique
Multidrup resistance proteins									
Abcb1a/b	IPI00470287	10	Multidrug resistance protein 1a/b	29	1	22	1	4	0
Abcb4	IPI00198519	10	Multidrug resistance protein 2	6	0	5	1	0	0
Abcc1	IPI00331756	17	Multidrug resistance-associated protein 1	4	4	1	1	0	0
Abcc4	IPI00421457	11	ATP-binding cassette protein C4	6	3	7	3	1	1
Abcg2	IPI00327093	5	ATP-binding cassette sub-family G member 2	7	7	6	6	0	0
Other ABC transporters									
Abca1	IPI00287199	11	ATP-binding cassette 1	4	4	2	2	0	0
Abca2	IPI00192286	14	ATP-binding cassette sub-family A member 2	14	14	10	10	0	0
Abca3	IPI00368700	12	ATP-binding cassette sub-family A member 3	4	4	1	1	0	0
Abca5	IPI00190668	12	ATP-binding cassette sub-family A member 5	7	7	4	4	1	1
Abca7	IPI00382157	12	ATP-binding cassette sub-family A member 7	3	3	1	1	0	0
Abca8a	IPI00361512	13	ATP-binding cassette, sub-family A (ABC1), member 8a	1	1	1	1	2	2
Abca9 or Abca8b	IPI00765350	13	ATP-binding cassette transporter sub-family A member 9/8b	2	1	1	0	2	1
Abcb6	IPI00199586	10	Mitochondrial ATP-binding cassette sub-family B member 6	3	3	3	3	0	0
Abcb7	IPI00421940	6	ATP-binding cassette sub-family B member 7, mitochondrial	20	20	16	16	5	5

Abcb8	IPI00200542	5	ABC transporter 8	13	10	10	7	2	1
Abcb8	IPI00781425	0	14 kDa protein	4	1	4	1	2	1
Abcb9	IPI00214174	9	ATP-binding cassette sub-family B member 9	3	3	4	4	1	1
Abcb10	IPI00189064	5	ATP-binding cassette, sub-family B (MDR/TAP), member 10	10	10	11	11	2	2
Abcc8	IPI00204427	14	ATP-binding cassette transporter sub-family C member 8	6	6	1	1	0	0
Abcd1	IPI00559296	6	ATP-binding cassette, sub-family D (ALD), member 1	1	1	4	4	0	0
Abcd2	IPI00213550	6	ATP-binding cassette sub-family D member 2	6	6	8	8	0	0
Abcd3	IPI00231860	5	ATP-binding cassette sub-family D member 3	9	9	11	11	5	5
Abce1	IPI00193816	0	ATP-binding cassette, sub-family E (OABP), member 1	0	0	2	2	0	0
Abcf2	IPI00213162	0	Abcf2_predicted protein	5	5	6	6	0	0
Abcf3	IPI00370458	0	ATP-binding cassette sub-family F member 3	1	1	2	2	0	0
Abcg1	IPI00189036	6	ABC transporter, white homologue	4	4	2	2	0	0
Abcg3	IPI00470299	5	ATP-binding cassette, sub-family G (WHITE), member 3	3	1	4	1	1	1

[a]**TM:** number of predicted transmembrane domain by TMAP algorithm; [b] unique peptide is the peptide with a sequence specific to the protein but not shared with other proteins

Table 6. KEGG pathways (http://www.genome.jp/kegg/) overrepresented in our identified dataset with a p-value lower than 0.01 are shown.

KEGG pathway	Description	Total involved rat proteins	Cerebellum		Spinal Cord		Sciatic Nerve	
			Proteins Identified	p-value*	Proteins Identified	p-value	Proteins Identified	p-value
5030	Amyotrophic lateral sclerosis (ALS)	30	14	5.52E-04	13	1.97E-03	11	1.07E-03
4730	Long-term depression	71	41	6.06E-06	38	7.56E-05	29	1.65E-04
4720	Long-term potentiation	62	46	8.43E-12	46	3.39E-12	30	2.14E-06
4540	Gap junction	89	48	1.32E-05	47	1.72E-05	32	1.15E-03
4530	Tight junction	125	61	5.65E-05	62	1.17E-05	43	5.20E-04
4512	ECM-receptor interaction	68	36	2.62E-04	36	1.57E-04	31	7.02E-06
4510	Focal adhesion	176	77	5.89E-04	74	1.42E-03	60	5.65E-05
4130	SNARE interactions in vesicular transport	36	30	2.64E-10	30	1.40E-10	21	1.58E-06
3050	Proteasome	25	19	7.85E-06	19	5.37E-06	13	7.60E-04
3010	Ribosome	68	44	2.57E-08	45	2.73E-09	41	3.12E-12
1430	Cell Communication	118	55	5.78E-04	51	3.90E-03	57	5.14E-11
930	Caprolactam degradation	8	8	1.11E-04	8	9.23E-05	6	1.85E-03
720	Reductive carboxylate cycle (CO_2 fixation)	8	8	1.11E-04	8	9.23E-05	7	1.38E-04
650	Butanoate metabolism	31	24	2.54E-07	22	6.29E-06	14	2.74E-03
640	Propanoate metabolism	23	19	8.01E-07	19	5.38E-07	13	2.58E-04
620	Pyruvate metabolism	28	21	3.65E-06	21	2.41E-06	18	1.22E-06
480	Glutathione metabolism	29	20	4.80E-05	20	3.30E-05	14	1.24E-03
280	Valine, leucine and isoleucine degradation	34	29	1.72E-10	28	1.02E-09	17	2.19E-04
190	Oxidative phosphorylation	102	73	1.25E-16	73	3.01E-17	67	2.32E-22
100	Biosynthesis of steroids	21	15	2.43E-04	17	3.81E-06	11	1.81E-03
71	Fatty acid metabolism	40	29	1.63E-07	29	9.27E-08	18	7.39E-04
62	Fatty acid elongation in	10	8	2.56E-03	8	2.17E-03	7	1.37E-03

	mitochondria							
20	Citrate cycle (TCA cycle)	21	19	3.89E-08	19	2.56E-08	16	1.29E-07
10	Glycolysis / Gluconeogenesis	44	23	4.15E-03	24	1.13E-03	22	2.61E-05

*The p-value was calculated by the hypergeometric test for identified rat proteins w.r.t the complete rat proteome.

Figure 1

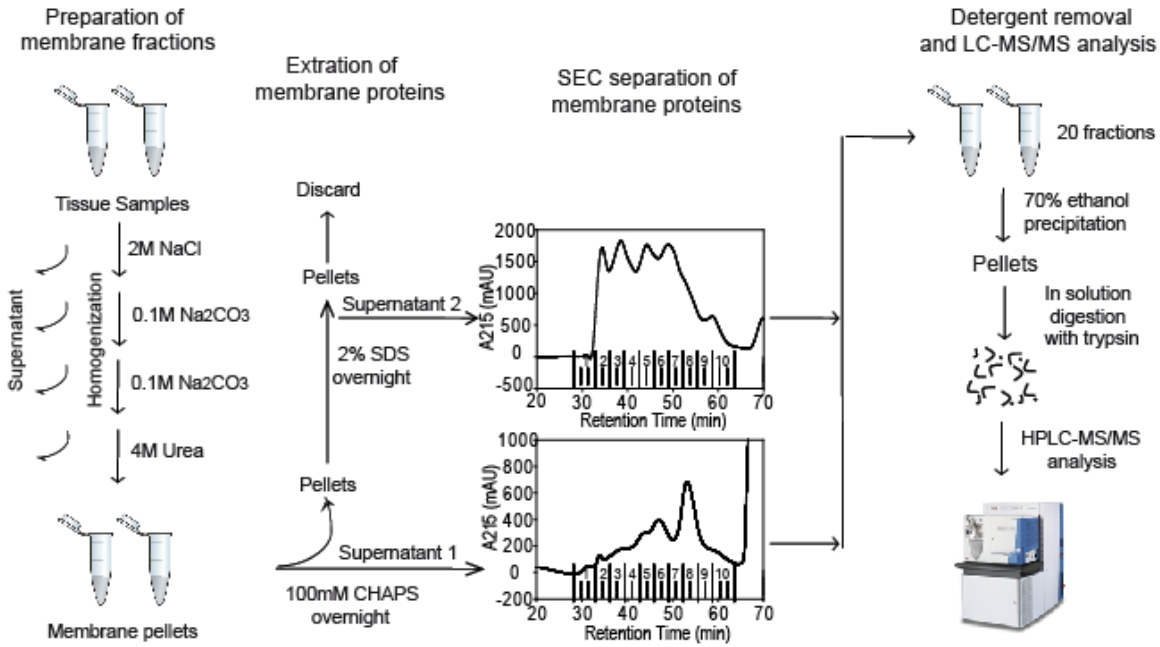

Preparation of membrane fractions
Tissue Samples
2M NaCl
0.1M Na2CO3
0.1M Na2CO3
4M Urea
Supernatant
Homogenization
Membrane pellets
Extration of membrane proteins
Discard
Pellets
Supernatant 2
2% SDS overnight
Pellets
Supernatant 1
100mM CHAPS overnight
SEC separation of membrane proteins
A215 (mAU)
Retention Time (min)
Detergent removal and LC-MS/MS analysis
20 fractions
70% ethanol precipitation
Pellets
In solution digestion with trypsin
HPLC-MS/MS analysis

Figure 2.

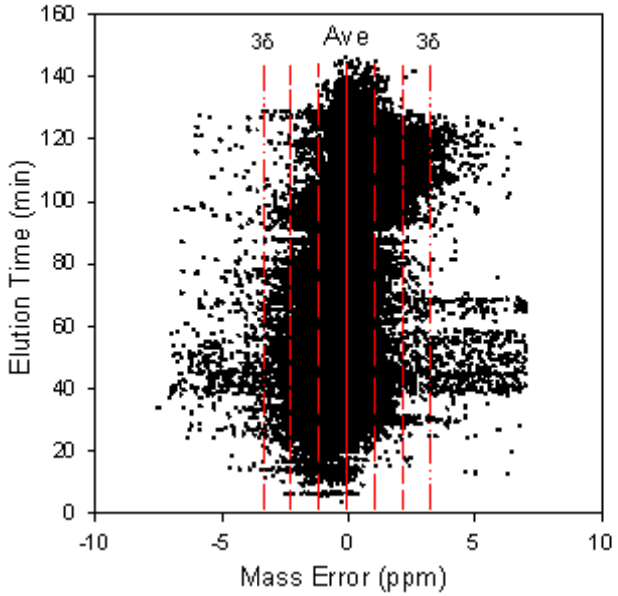

3δ
Ave
3δ
160
140
120
100
80
60
40
20
0
Elution Time (min)
-10
-5
0
5
10
Mass Error (ppm)

Figure 3

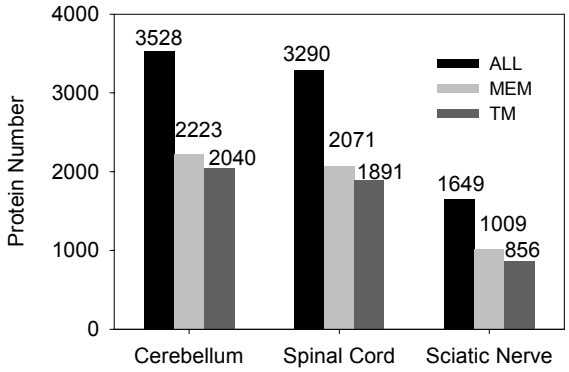


Figure 4

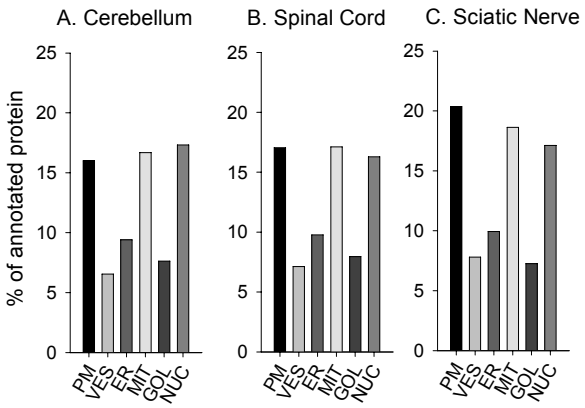


Figure 5

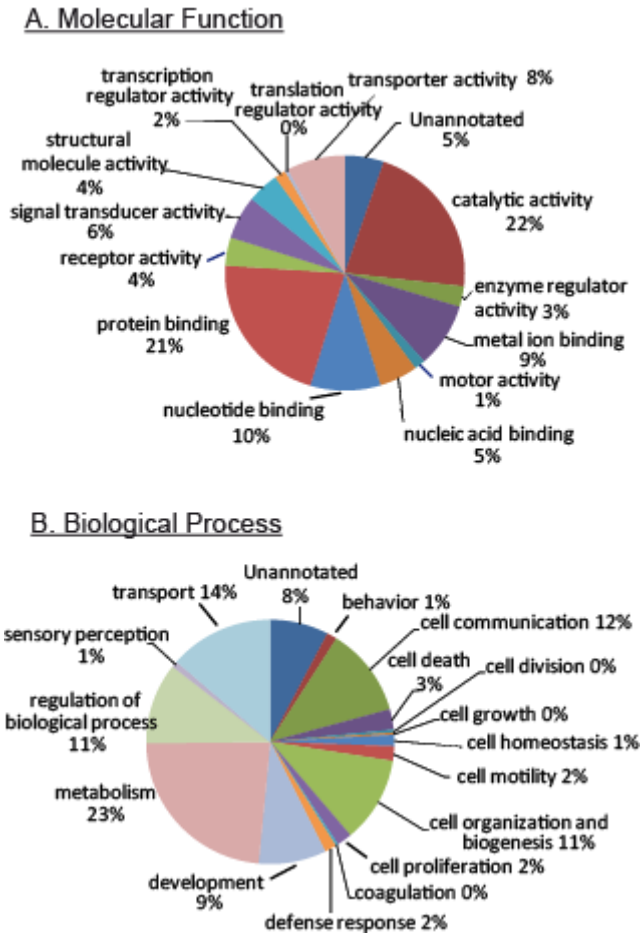
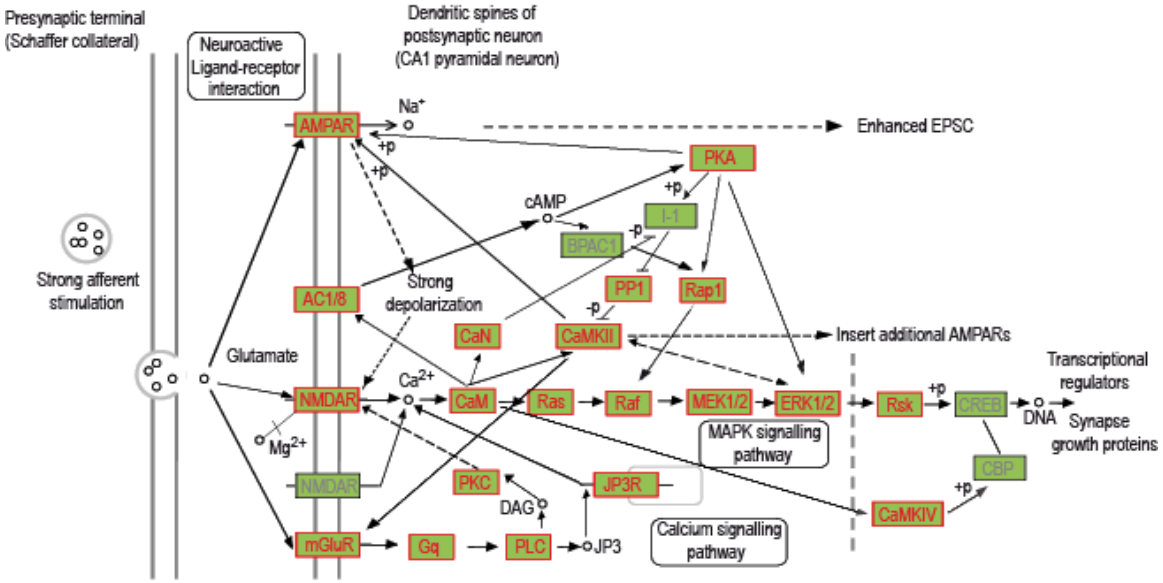
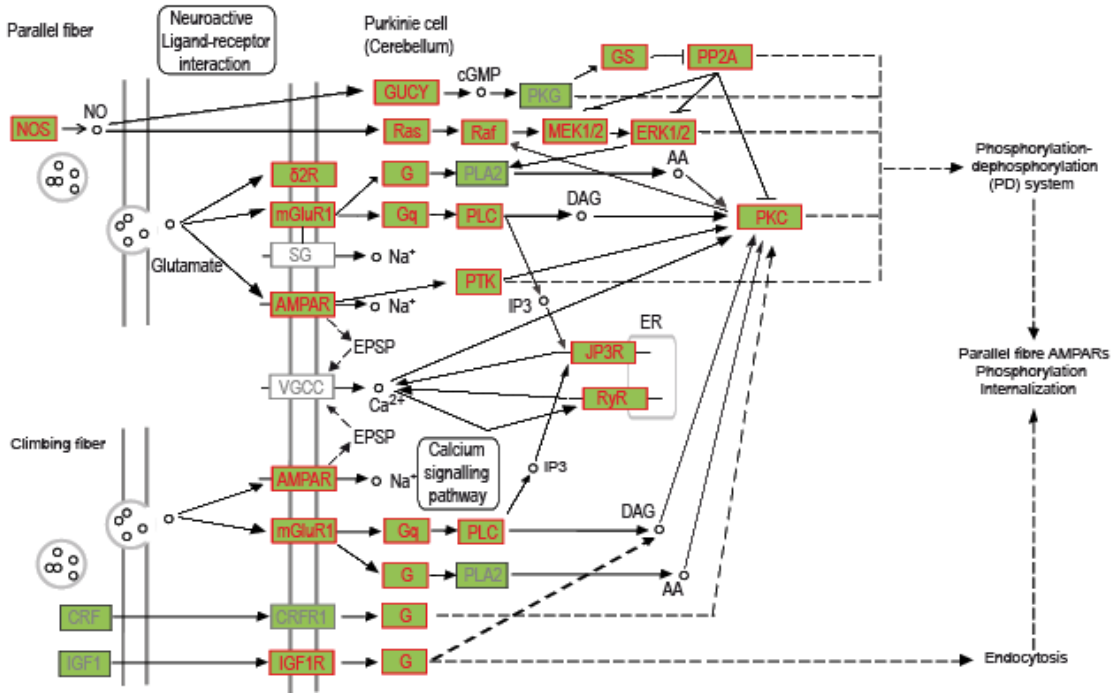

A. Molecular Function
transcription regulator activity 2%
translation regulator activity 0%
transporter activity 8%
Unannotated 5%
structural molecule activity 4%
signal transducer activity 6%
receptor activity 4%
catalytic activity 22%
enzyme regulator activity 3%
metal ion binding 9%
motor activity 1%
nucleic acid binding 5%
nucleotide binding 10%
protein binding 21%
B. Biological Process
Unannotated 8%
behavior 1%
cell communication 12%
cell death 3%
cell division 0%
cell growth 0%
cell homeostasis 1%
cell motility 2%
cell organization and biogenesis 11%
cell proliferation 2%
coagulation 0%
defense response 2%
development 9%
metabolism 23%
regulation of biological process 11%
sensory perception 1%
transport 14%

Figure 6

A. Long-term potentiation



B. Long-term depression



7 Summary and perspectives

The work presented in this thesis describe my effort to develop optimal MS-based proteomics techniques for high sequence mapping of relative pure protein samples as well as membrane protein identification from tissue samples.

Dynamic range and sensitivity are two critical issues related to MS-based proteomics. Real biological samples, even the peptide digested from the same protein, can display a dynamic range of several thousand to more than 10^{10}. Fractionation starting from separating high intensity proteins from the low proteins at the beginning of sample preparation to the segmentation of high intensity m/z peak to low intensity ones at the final data acquisition step, increases the dynamic range of MS identifications. In the peptide mapping approach, we introduce the concept of "composite spectra", which are spectra composed of a high number of segmented SIM scans. High abundant mass ranges and low abundant ranges are acquired separately. These composite spectra allow very high sensitivity, accuracy and dynamic range due to optimized fill times for each mass segment. The dynamic range of the approaches could be further increased by coupling off-line HPLC separation of the peptides before spray. By applying this approach to histone H1 samples from human breast cancer, on average around 80% sequence coverage was obtained for the K/R rich histone H1 variants, H1.0, H1.2-4, and H1.5. Another advantage of the peptide mapping approach is that it is possible to quantify the occupancy for different modifications in a single spectrum, since all peptides come out simultaneously at a stable spray rate.

Detergent is an indispensible chemical to dissolve membrane proteins, however, even a small amount of detergent is incompatible with MS acquisition of the peptides. Normally MS-based proteomics protocols starting with detergent dissolving removes the detergent in a gel based method. In our project, we were able to demonstrate that detergent can be completely removed by solvent exchange in a desalting column. Compared to the gel based approach, it is fast and more importantly, results in much higher recovery of high mass proteins, which may not enter the gel matrix. With the help of advanced LTQ-Orbitrap techniques, we are able to identify nearly 1000 membrane proteins from a single

MS analysis. To increase the identification number, we incorporated gel filtration chromatography, a method compatible with detergent, which increased the number to 3000 for one sample. Mining different family of membrane proteins, e.g. receptors, ion channels, our method demonstrates the possibility to identify most of them.

Although the techniques developed here are clearly powerful, they can still be further developed. For example, in the writing of the histone H1 methylation manuscript, we found a disadvantage of using methanol as the spray buffer: Glutamic acid shows strong artifactual methylation in buffers containing methanol. This necessitated discarding many potential methylation sites because they were indistinguishable from artifactual methylation. This includes a very interesting discovery of a possible *phospho-switch* in H1.4 at K17/T18. Due to the limitation of sample and time, we did not repeat the experiment with other organic solvent, e.g. ethanol or acetonitrile, but this would be interesting for further studies.

Techniques for sample preparation develop very fast, especially in our laboratory. Prof. Jacek R. Wiśniewski has recently developed a protocol call "FASP", which results in very clean protein sample. Combine with OFFGEL peptide separation, it is possible to identify more than 6000 proteins in one experiment and the number of identified membrane protein is more than 3000. This raises the question, when and where is necessary to purify membrane proteins. If it is only for the identification level, with so far developed sample preparation techniques and MS instrument, it might seem of less interest to separate membrane proteins from the soluble proteins. Although subfractionation increases dynamic range, extra procedure causes more sample loss, in particular the low abundant membrane proteins. However, when the cellular localization of protein is of interest, it is still worthwhile to perform membrane protein enrichment.

The algorithmic and software development of label free quantitation in our group will now make comparative proteomics much more valuable. However, for robust quantitation, several repeat runs are necessary.

In conclusion, the chip implementation of nanoelectrospray method provides an alternative way to HPLC-ESI-MS or MALDI-MS for individual protein sequence

mapping and PTM characterization. It is a method with high sensitivity and high dynamic range. From the application to linker histone H1, we can see that the method has the further advantage of identifying tryptic peptides of a length of five to six amino acids, which normally are missed in HPLC-MS based analysis. This thesis also reports, for the first time, large-scale membrane proteomics for nerve tissue, especially the peripheral nerve tissue. Most of the known receptors and ion channel proteins were identified. Our detergent-based gel-free protocol enable identification of nearly 1000 proteins within a single MS run, and the total time from tissue homogenization to start trypsin digestion is only about 4 h. This makes the method potentially well suited to practical application in clinical analysis.

8 References

1. Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **1997,** 389, (6648), 251-60.
2. Robinson, P. J.; Rhodes, D., Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol* **2006,** 16, (3), 336-43.
3. Cosgrove, M. S.; Wolberger, C., How does the histone code work? *Biochem Cell Biol* **2005,** 83, (4), 468-76.
4. Dhalluin, C.; Carlson, J. E.; Zeng, L.; He, C.; Aggarwal, A. K.; Zhou, M. M., Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **1999,** 399, (6735), 491-6.
5. Winston, F.; Allis, C. D., The bromodomain: a chromatin-targeting module? *Nat Struct Biol* **1999,** 6, (7), 601-4.
6. Owen, D. J.; Ornaghi, P.; Yang, J. C.; Lowe, N.; Evans, P. R.; Ballario, P.; Neuhaus, D.; Filetici, P.; Travers, A. A., The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *Embo J* **2000,** 19, (22), 6141-9.
7. Wei, Y.; Yu, L.; Bowen, J.; Gorovsky, M. A.; Allis, C. D., Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell* **1999,** 97, (1), 99-109.
8. Bannister, A. J.; Zegerman, P.; Partridge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T., Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **2001,** 410, (6824), 120-4.
9. Lachner, M.; O'Carroll, D.; Rea, S.; Mechtler, K.; Jenuwein, T., Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **2001,** 410, (6824), 116-20.
10. Strahl, B. D.; Allis, C. D., The language of covalent histone modifications. *Nature* **2000,** 403, (6765), 41-5.
11. Turner, B. M., Histone acetylation and an epigenetic code. *Bioessays* **2000,** 22, (9), 836-45.
12. Jenuwein, T.; Allis, C. D., Translating the histone code. *Science* **2001,** 293, (5532), 1074-80.
13. Fischle, W.; Wang, Y.; Allis, C. D., Binary switches and modification cassettes in histone biology and beyond. *Nature* **2003,** 425, (6957), 475-9.
14. Cocklin, R. R.; Wang, M., Identification of methylation and acetylation sites on mouse histone H3 using matrix-assisted laser desorption/ionization time-of-flight and nanoelectrospray ionization tandem mass spectrometry. *J Protein Chem* **2003,** 22, (4), 327-34.
15. Zhang, K.; Tang, H.; Huang, L.; Blankenship, J. W.; Jones, P. R.; Xiang, F.; Yau, P. M.; Burlingame, A. L., Identification of acetylation and methylation sites of histone H3 from chicken erythrocytes by high-accuracy matrix-assisted laser desorption ionization-time-of-flight, matrix-assisted laser desorption ionization-postsource decay, and nanoelectrospray ionization tandem mass spectrometry. *Anal Biochem* **2002,** 306, (2), 259-69.
16. Zhang, L.; Eugeni, E. E.; Parthun, M. R.; Freitas, M. A., Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosoma* **2003,** 112, (2), 77-86.
17. Freitas, M. A.; Sklenar, A. R.; Parthun, M. R., Application of mass spectrometry to the identification and quantification of histone post-translational modifications. *J Cell Biochem* **2004,** 92, (4), 691-700.
18. Cosgrove, M. S.; Boeke, J. D.; Wolberger, C., Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol* **2004,** 11, (11), 1037-43.

19. Izzo, A.; Kamieniarz, K.; Schneider, R., The histone H1 family: specific members, specific functions? *Biol Chem* **2008,** 389, (4), 333-43.
20. Panyim, S.; Chalkley, R., The molecular weights of vertebrate histones exploiting a modified sodium dodecyl sulfate electrophoretic method. *J Biol Chem* **1971,** 246, (24), 7557-60.
21. Lennox, R. W.; Cohen, L. H., The histone H1 complements of dividing and nondividing cells of the mouse. *J Biol Chem* **1983,** 258, (1), 262-8.
22. Albig, W.; Drabent, B.; Kunz, J.; Kalff-Suske, M.; Grzeschik, K. H.; Doenecke, D., All known human H1 histone genes except the H1(0) gene are clustered on chromosome 6. *Genomics* **1993,** 16, (3), 649-54.
23. Rall, S. C.; Cole, R. D., Amino acid sequence and sequence variability of the amino-terminal regions of lysine-rich histones. *J Biol Chem* **1971,** 246, (23), 7175-90.
24. Yamamoto, T.; Horikoshi, M., Cloning of the cDNA encoding a novel subtype of histone H1. *Gene* **1996,** 173, (2), 281-5.
25. Happel, N.; Schulze, E.; Doenecke, D., Characterisation of human histone H1x. *Biol Chem* **2005,** 386, (6), 541-51.
26. Meergans, T.; Albig, W.; Doenecke, D., Varied expression patterns of human H1 histone genes in different cell lines. *DNA Cell Biol* **1997,** 16, (9), 1041-9.
27. Chapman, G. E.; Hartman, P. G.; Bradbury, E. M., Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. The isolation of the globular and non-globular regions of the histone H1 molecule. *Eur J Biochem* **1976,** 61, (1), 69-75.
28. Patterton, H. G.; Landel, C. C.; Landsman, D.; Peterson, C. L.; Simpson, R. T., The biochemical and phenotypic characterization of Hho1p, the putative linker histone H1 of Saccharomyces cerevisiae. *J Biol Chem* **1998,** 273, (13), 7268-76.
29. Misteli, T.; Gunjan, A.; Hock, R.; Bustin, M.; Brown, D. T., Dynamic binding of histone H1 to chromatin in living cells. *Nature* **2000,** 408, (6814), 877-81.
30. Orrego, M.; Ponte, I.; Roque, A.; Buschati, N.; Mora, X.; Suau, P., Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin. *BMC Biol* **2007,** 5, 22.
31. Talasz, H.; Sapojnikova, N.; Helliger, W.; Lindner, H.; Puschendorf, B., In vitro binding of H1 histone subtypes to nucleosomal organized mouse mammary tumor virus long terminal repeat promotor. *J Biol Chem* **1998,** 273, (48), 32236-43.
32. Th'ng, J. P.; Sung, R.; Ye, M.; Hendzel, M. J., H1 family histones in the nucleus. Control of binding and localization by the C-terminal domain. *J Biol Chem* **2005,** 280, (30), 27809-14.
33. Brown, D. T.; Izard, T.; Misteli, T., Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. *Nat Struct Mol Biol* **2006,** 13, (3), 250-5.
34. Catez, F.; Ueda, T.; Bustin, M., Determinants of histone H1 mobility and chromatin binding in living cells. *Nat Struct Mol Biol* **2006,** 13, (4), 305-10.
35. Godde, J. S.; Ura, K., Cracking the enigmatic linker histone code. *J Biochem* **2008,** 143, (3), 287-93.
36. Fan, Y.; Nikitina, T.; Morin-Kensicki, E. M.; Zhao, J.; Magnuson, T. R.; Woodcock, C. L.; Skoultchi, A. I., H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. *Mol Cell Biol* **2003,** 23, (13), 4559-72.
37. Alami, R.; Fan, Y.; Pack, S.; Sonbuchner, T. M.; Besse, A.; Lin, Q.; Greally, J. M.; Skoultchi, A. I.; Bouhassira, E. E., Mammalian linker-histone subtypes differentially affect gene expression in vivo. *Proc Natl Acad Sci U S A* **2003,** 100, (10), 5920-5.
38. Fan, Y.; Nikitina, T.; Zhao, J.; Fleury, T. J.; Bhattacharyya, R.; Bouhassira, E. E.; Stein, A.; Woodcock, C. L.; Skoultchi, A. I., Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **2005,** 123, (7), 1199-212.

39. Wierzbicki, A. T.; Jerzmanowski, A., Suppression of histone H1 genes in Arabidopsis results in heritable developmental defects and stochastic changes in DNA methylation. *Genetics* **2005,** 169, (2), 997-1008.
40. Gunjan, A.; Alexander, B. T.; Sittman, D. B.; Brown, D. T., Effects of H1 histone variant overexpression on chromatin structure. *J Biol Chem* **1999,** 274, (53), 37950-6.
41. Banks, G. C.; Deterding, L. J.; Tomer, K. B.; Archer, T. K., Hormone-mediated dephosphorylation of specific histone H1 isoforms. *J Biol Chem* **2001,** 276, (39), 36467-73.
42. Belikov, S.; Astrand, C.; Wrange, O., Mechanism of histone H1-stimulated glucocorticoid receptor DNA binding in vivo. *Mol Cell Biol* **2007,** 27, (6), 2398-410.
43. Hohmann, P., Phosphorylation of H1 histones. *Mol Cell Biochem* **1983,** 57, (1), 81-92.
44. Deterding, L. J.; Banks, G. C.; Tomer, K. B.; Archer, T. K., Understanding global changes in histone H1 phosphorylation using mass spectrometry. *Methods* **2004,** 33, (1), 53-8.
45. Garcia, B. A.; Busby, S. A.; Barber, C. M.; Shabanowitz, J.; Allis, C. D.; Hunt, D. F., Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry. *J Proteome Res* **2004,** 3, (6), 1219-27.
46. Garcia, B. A.; Joshi, S.; Thomas, C. E.; Chitta, R. K.; Diaz, R. L.; Busby, S. A.; Andrews, P. C.; Ogorzalek Loo, R. R.; Shabanowitz, J.; Kelleher, N. L.; Mizzen, C. A.; Allis, C. D.; Hunt, D. F., Comprehensive phosphoprotein analysis of linker histone H1 from Tetrahymena thermophila. *Mol Cell Proteomics* **2006,** 5, (9), 1593-609.
47. Wisniewski, J. R.; Zougman, A.; Kruger, S.; Mann, M., Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol Cell Proteomics* **2007,** 6, (1), 72-87.
48. Stevens, T. J.; Arkin, I. T., Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins* **2000,** 39, (4), 417-20.
49. Hopkins, A. L.; Groom, C. R., The druggable genome. *Nat Rev Drug Discov* **2002,** 1, (9), 727-30.
50. Russ, A. P.; Lampel, S., The druggable genome: an update. *Drug Discov Today* **2005,** 10, (23-24), 1607-10.
51. Itier, V.; Bertrand, D., Neuronal nicotinic receptors: from protein structure to function. *FEBS Lett* **2001,** 504, (3), 118-25.
52. Malinow, R.; Malenka, R. C., AMPA receptor trafficking and synaptic plasticity. *Annu Rev Neurosci* **2002,** 25, 103-26.
53. Huettner, J. E., Kainate receptors and synaptic transmission. *Prog Neurobiol* **2003,** 70, (5), 387-407.
54. Stephenson, F. A., Structure and trafficking of NMDA and GABAA receptors. *Biochem Soc Trans* **2006,** 34, (Pt 5), 877-81.
55. Bormann, J., The 'ABC' of GABA receptors. *Trends Pharmacol Sci* **2000,** 21, (1), 16-9.
56. Luttrell, L. M., Reviews in molecular biology and biotechnology: transmembrane signaling by G protein-coupled receptors. *Mol Biotechnol* **2008,** 39, (3), 239-64.
57. Doherty, J.; Dingledine, R., The roles of metabotropic glutamate receptors in seizures and epilepsy. *Curr Drug Targets CNS Neurol Disord* **2002,** 1, (3), 251-60.
58. Meldrum, B. S.; Akbar, M. T.; Chapman, A. G., Glutamate receptors and transporters in genetic and acquired models of epilepsy. *Epilepsy Res* **1999,** 36, (2-3), 189-204.
59. Catterall, W. A., From ionic currents to molecular mechanisms: the structure and function of voltage-gated sodium channels. *Neuron* **2000,** 26, (1), 13-25.
60. Dolphin, A. C., A short history of voltage-gated calcium channels. *Br J Pharmacol* **2006,** 147 Suppl 1, S56-62.

61. Gutman, G. A.; Chandy, K. G.; Grissmer, S.; Lazdunski, M.; McKinnon, D.; Pardo, L. A.; Robertson, G. A.; Rudy, B.; Sanguinetti, M. C.; Stuhmer, W.; Wang, X., International Union of Pharmacology. LIII. Nomenclature and molecular relationships of voltage-gated potassium channels. *Pharmacol Rev* **2005,** 57, (4), 473-508.
62. Decoursey, T. E., Voltage-gated proton channels and other proton transfer pathways. *Physiol Rev* **2003,** 83, (2), 475-579.
63. Hediger, M. A.; Romero, M. F.; Peng, J. B.; Rolfs, A.; Takanaga, H.; Bruford, E. A., The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction. *Pflugers Arch* **2004,** 447, (5), 465-8.
64. Dean, M.; Rzhetsky, A.; Allikmets, R., The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res* **2001,** 11, (7), 1156-66.
65. Juliano, R. L.; Ling, V., A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochim Biophys Acta* **1976,** 455, (1), 152-62.
66. Santoni, V.; Molloy, M.; Rabilloud, T., Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **2000,** 21, (6), 1054-70.
67. Lu, B.; McClatchy, D. B.; Kim, J. Y.; Yates, J. R., 3rd, Strategies for shotgun identification of integral membrane proteins by tandem mass spectrometry. *Proteomics* **2008,** 8, (19), 3947-55.
68. Speers, A. E.; Wu, C. C., Proteomics of integral membrane proteins--theory and application. *Chem Rev* **2007,** 107, (8), 3687-714.
69. Mirza, S. P.; Halligan, B. D.; Greene, A. S.; Olivier, M., Improved method for the analysis of membrane proteins by mass spectrometry. *Physiol Genomics* **2007,** 30, (1), 89-94.
70. Dormeyer, W.; van Hoof, D.; Braam, S. R.; Heck, A. J.; Mummery, C. L.; Krijgsveld, J., Plasma membrane proteomics of human embryonic stem cells and human embryonal carcinoma cells. *J Proteome Res* **2008,** 7, (7), 2936-51.
71. Wessel, D.; Flugge, U. I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **1984,** 138, (1), 141-3.
72. de Araujo, M. E.; Huber, L. A.; Stasyk, T., Isolation of endocitic organelles by density gradient centrifugation. *Methods Mol Biol* **2008,** 424, 317-31.
73. Rieder, S. E.; Emr, S. D., Overview of subcellular fractionation procedures for the yeast Saccharomyces cerevisiae. *Curr Protoc Cell Biol* **2001,** Chapter 3, Unit 3 7.
74. Huber, L. A.; Pfaller, K.; Vietor, I., Organelle proteomics: implications for subcellular fractionation in proteomics. *Circ Res* **2003,** 92, (9), 962-8.
75. Wehmhoner, D.; Dieterich, G.; Fischer, E.; Baumgartner, M.; Wehland, J.; Jansch, L., "LaneSpector", a tool for membrane proteome profiling based on sodium dodecyl sulfate-polyacrylamide gel electrophoresis/liquid chromatography-tandem mass spectrometry analysis: application to Listeria monocytogenes membrane proteins. *Electrophoresis* **2005,** 26, (12), 2450-60.
76. Blonder, J.; Terunuma, A.; Conrads, T. P.; Chan, K. C.; Yee, C.; Lucas, D. A.; Schaefer, C. F.; Yu, L. R.; Issaq, H. J.; Veenstra, T. D.; Vogel, J. C., A proteomic characterization of the plasma membrane of human epidermis by high-throughput mass spectrometry. *J Invest Dermatol* **2004,** 123, (4), 691-9.
77. Song, Y.; Hao, Y.; Sun, A.; Li, T.; Li, W.; Guo, L.; Yan, Y.; Geng, C.; Chen, N.; Zhong, F.; Wei, H.; Jiang, Y.; He, F., Sample preparation project for the subcellular proteome of mouse liver. *Proteomics* **2006,** 6, (19), 5269-77.
78. Cao, R.; Li, X.; Liu, Z.; Peng, X.; Hu, W.; Wang, X.; Chen, P.; Xie, J.; Liang, S., Integration of a two-phase partition method into proteomics research on rat liver plasma membrane proteins. *J Proteome Res* **2006,** 5, (3), 634-42.

79. Cao, R.; He, Q.; Zhou, J.; He, Q.; Liu, Z.; Wang, X.; Chen, P.; Xie, J.; Liang, S., High-throughput analysis of rat liver plasma membrane proteome by a nonelectrophoretic in-gel tryptic digestion coupled with mass spectrometry identification. *J Proteome Res* **2008,** 7, (2), 535-45.
80. Huang, F.; Parmryd, I.; Nilsson, F.; Persson, A. L.; Pakrasi, H. B.; Andersson, B.; Norling, B., Proteomics of Synechocystis sp. strain PCC 6803: identification of plasma membrane proteins. *Mol Cell Proteomics* **2002,** 1, (12), 956-66.
81. Khemiri, A.; Galland, A.; Vaudry, D.; Chan Tchi Song, P.; Vaudry, H.; Jouenne, T.; Cosette, P., Outer-membrane proteomic maps and surface-exposed proteins of Legionella pneumophila using cellular fractionation and fluorescent labelling. *Anal Bioanal Chem* **2008,** 390, (7), 1861-71.
82. Nielsen, P. A.; Olsen, J. V.; Podtelejnikov, A. V.; Andersen, J. R.; Mann, M.; Wisniewski, J. R., Proteomic mapping of brain plasma membrane proteins. *Mol Cell Proteomics* **2005,** 4, (4), 402-8.
83. Jethwaney, D.; Islam, M. R.; Leidal, K. G.; de Bernabe, D. B.; Campbell, K. P.; Nauseef, W. M.; Gibson, B. W., Proteomic analysis of plasma membrane and secretory vesicles from human neutrophils. *Proteome Sci* **2007,** 5, 12.
84. Schindler, J.; Lewandrowski, U.; Sickmann, A.; Friauf, E., Aqueous polymer two-phase systems for the proteomic analysis of plasma membranes from minute brain samples. *J Proteome Res* **2008,** 7, (1), 432-42.
85. Schindler, J.; Nothwang, H. G., Aqueous polymer two-phase systems: effective tools for plasma membrane proteomics. *Proteomics* **2006,** 6, (20), 5409-17.
86. Schindler, J.; Lewandrowski, U.; Sickmann, A.; Friauf, E.; Nothwang, H. G., Proteomic analysis of brain plasma membranes isolated by affinity two-phase partitioning. *Mol Cell Proteomics* **2006,** 5, (2), 390-400.
87. Natera, S. H.; Ford, K. L.; Cassin, A. M.; Patterson, J. H.; Newbigin, E. J.; Bacic, A., Analysis of the Oryza sativa plasma membrane proteome using combined protein and peptide fractionation approaches in conjunction with mass spectrometry. *J Proteome Res* **2008,** 7, (3), 1159-87.
88. Alexandersson, E.; Saalbach, G.; Larsson, C.; Kjellbom, P., Arabidopsis plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. *Plant Cell Physiol* **2004,** 45, (11), 1543-56.
89. Morel, J.; Claverol, S.; Mongrand, S.; Furt, F.; Fromentin, J.; Bessoule, J. J.; Blein, J. P.; Simon-Plas, F., Proteomics of plant detergent-resistant membranes. *Mol Cell Proteomics* **2006,** 5, (8), 1396-411.
90. Zhao, Y.; Zhang, W.; Kho, Y.; Zhao, Y., Proteomic analysis of integral plasma membrane proteins. *Anal Chem* **2004,** 76, (7), 1817-23.
91. Zhao, Y.; Zhang, W.; White, M. A.; Zhao, Y., Capillary high-performance liquid chromatography/mass spectrometric analysis of proteins from affinity-purified plasma membrane. *Anal Chem* **2003,** 75, (15), 3751-7.
92. Zhang, W.; Zhou, G.; Zhao, Y.; White, M. A.; Zhao, Y., Affinity enrichment of plasma membrane for proteomics analysis. *Electrophoresis* **2003,** 24, (16), 2855-63.
93. Chen, W. N.; Yu, L. R.; Strittmatter, E. F.; Thrall, B. D.; Camp, D. G., 2nd; Smith, R. D., Detection of in situ labeled cell surface proteins by mass spectrometry: application to the membrane subproteome of human mammary epithelial cells. *Proteomics* **2003,** 3, (8), 1647-51.
94. Blonder, J.; Goshe, M. B.; Xiao, W.; Camp, D. G., 2nd; Wingerd, M.; Davis, R. W.; Smith, R. D., Global analysis of the membrane subproteome of Pseudomonas aeruginosa using liquid chromatography-tandem mass spectrometry. *J Proteome Res* **2004,** 3, (3), 434-44.

95. Jang, J. H.; Hanash, S., Profiling of the cell surface proteome. *Proteomics* **2003,** 3, (10), 1947-54.
96. Springer, D. L.; Auberry, D. L.; Ahram, M.; Adkins, J. N.; Feldhaus, J. M.; Wahl, J. H.; Wunschel, D. S.; Rodland, K. D., Characterization of plasma membrane proteins from ovarian cancer cells using mass spectrometry. *Dis Markers* **2003,** 19, (4-5), 219-28.
97. Wu, C. C.; Yates, J. R., 3rd, The application of mass spectrometry to membrane proteomics. *Nat Biotechnol* **2003,** 21, (3), 262-7.
98. Furth, A. J., Removing unbound detergent from hydrophobic proteins. *Anal Biochem* **1980,** 109, (2), 207-15.
99. Goshe, M. B.; Blonder, J.; Smith, R. D., Affinity labeling of highly hydrophobic integral membrane proteins for proteome-wide analysis. *J Proteome Res* **2003,** 2, (2), 153-61.
100. Blonder, J.; Goshe, M. B.; Moore, R. J.; Pasa-Tolic, L.; Masselon, C. D.; Lipton, M. S.; Smith, R. D., Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography-tandem mass spectrometry. *J Proteome Res* **2002,** 1, (4), 351-60.
101. Blonder, J.; Conrads, T. P.; Yu, L. R.; Terunuma, A.; Janini, G. M.; Issaq, H. J.; Vogel, J. C.; Veenstra, T. D., A detergent- and cyanogen bromide-free method for integral membrane proteomics: application to Halobacterium purple membranes and the human epidermal membrane proteome. *Proteomics* **2004,** 4, (1), 31-45.
102. Zhang, N.; Chen, R.; Young, N.; Wishart, D.; Winter, P.; Weiner, J. H.; Li, L., Comparison of SDS- and methanol-assisted protein solubilization and digestion methods for Escherichia coli membrane proteome analysis by 2-D LC-MS/MS. *Proteomics* **2007,** 7, (4), 484-93.
103. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001,** 19, (3), 242-7.
104. Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R., 3rd, A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* **2003,** 21, (5), 532-8.
105. Cravatt, B. F.; Simon, G. M.; Yates, J. R., 3rd, The biological impact of mass-spectrometry-based proteomics. *Nature* **2007,** 450, (7172), 991-1000.
106. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** 246, (4926), 64-71.
107. Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F., *Matrix-Assisted Ultraviolet Laser Desorption of Non-Volatile Compounds*. International Journal of Mass Spectrometry and Ion Processes: 1987; Vol. 78, p 53-68.
108. Godoy, L. M. F. d.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Fröhlich, F.; Walther, T. C.; Mann, M., *Comprehensive mass spectrometry-based proteome quantitation of haploid versus diploid yeast*. submitted manuscript: 2008.
109. Siuti, N.; Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nat Methods* **2007,** 4, (10), 817-21.
110. Rappsilber, J.; Ishihama, Y.; Mann, M., Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **2003,** 75, (3), 663-70.
111. Mattauch, J.; Herzog, R., *Mass spectrograph*. z. Physik: 1934; Vol. 89, p 786.
112. Johnson, E. G.; Nier, A. O., *Angular aberrations in sector shaped electromagnetic lenses for focusing beams of charged particles*. Phys. Rev: 1953; Vol. 91, p 10.
113. Guilhaus, M., *Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts*. Journal of mass spectrometry: 1995; Vol. 30, p 1519.

114. Dawson, P. H., *Quadrupole Mass Spectrometry and its Applications*. Elsevier Scientific Publishing: New York, 1976; Vol. 349.
115. March, R., *An introduction to quadrupole ion trap mass sepctrometry*. Journal of Mass Spectrometry: 1997; Vol. 32, p 351.
116. Amster, I. J., *Fourier transform mass spectrometry*. Journal of Mass Spectrometry: 1996; Vol. 31, p 1325.
117. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R., The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **2005,** 40, (4), 430-43.
118. McLuckey, S. A.; Wells, J. M., Mass analysis at the advent of the 21st century. *Chem Rev* **2001,** 101, (2), 571-606.
119. Thomson, J. J., *Rays of positive electricity and their applicaiton to chemical analysis*. Longmans Green: London, 1913.
120. Wilm, M.; Mann, M., Analytical properties of the nanoelectrospray ion source. *Anal Chem* **1996,** 68, (1), 1-8.
121. Cech, N. B.; Enke, C. G., Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev* **2001,** 20, (6), 362-87.
122. Wisniewski, J. R., Mass spectrometry-based proteomics: principles, perspectives, and challenges. *Arch Pathol Lab Med* **2008,** 132, (10), 1566-9.
123. Cotter, R. J., *Laser mass spectrometry: an overview of techniques, instruments and applications*. Analytica Chimica Acta: 1987; p 45-59.
124. Hoffmann, E. d.; Stroobant, V., *Mass Spectrometry: Principles and Applications*.
125. Schwartz, J. C.; Senko, M. W.; Syka, J. E., A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* **2002,** 13, (6), 659-69.
126. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* **1998,** 17, (1), 1-35.
127. Amster, I. J., *Fourier Transform Mass Spectrometry*. Journal of Mass Spectrometry: 1996; Vol. 31, p 1325-1337.
128. Schaub, T. M.; Hendrickson, C. L.; Horning, S.; Quinn, J. P.; Senko, M. W.; Marshall, A. G., High-performance mass spectrometry: Fourier transform ion cyclotron resonance at 14.5 Tesla. *Anal Chem* **2008,** 80, (11), 3985-90.
129. Scigelova, M.; Makarov, A., Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **2006,** 6 Suppl 2, 16-21.
130. Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M., Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* **2005,** 4, (12), 2010-21.
131. Makarov, A., Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* **2000,** 72, (6), 1156-62.
132. Schrader, W.; Klein, H. W., Liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry (LC-FTICR MS): an early overview. *Anal Bioanal Chem* **2004,** 379, (7-8), 1013-24.
133. Roepstorff, P.; Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **1984,** 11, (11), 601.
134. Biemann, K., Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom* **1988,** 16, (1-12), 99-111.
135. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **2007,** 4, (9), 709-12.

136. Service, R. F., Proteomics. Proteomics ponders prime time. *Science* **2008,** 321, (5897), 1758-61.
137. Ong, S. E.; Foster, L. J.; Mann, M., Mass spectrometric-based approaches in quantitative proteomics. *Methods* **2003,** 29, (2), 124-30.
138. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **2007,** 389, (4), 1017-31.
139. Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T., Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* **1999,** 96, (12), 6591-6.
140. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002,** 1, (5), 376-86.
141. Dengjel, J.; Akimov, V.; Olsen, J. V.; Bunkenborg, J.; Mann, M.; Blagoev, B.; Andersen, J. S., Quantitative proteomic assessment of very early cellular signaling events. *Nat Biotechnol* **2007,** 25, (5), 566-8.
142. Mann, M., Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* **2006,** 7, (12), 952-8.
143. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999,** 17, (10), 994-9.
144. Olsen, J. V.; Andersen, J. R.; Nielsen, P. A.; Nielsen, M. L.; Figeys, D.; Mann, M.; Wisniewski, J. R., HysTag--a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. *Mol Cell Proteomics* **2004,** 3, (1), 82-92.
145. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004,** 3, (12), 1154-69.
146. Pierce, A.; Unwin, R. D.; Evans, C. A.; Griffiths, S.; Carney, L.; Zhang, L.; Jaworska, E.; Lee, C. F.; Blinco, D.; Okoniewski, M. J.; Miller, C. J.; Bitton, D. A.; Spooncer, E.; Whetton, A. D., Eight-channel iTRAQ enables comparison of the activity of six leukemogenic tyrosine kinases. *Mol Cell Proteomics* **2008,** 7, (5), 853-63.
147. Ong, S. E.; Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* **2005,** 1, (5), 252-62.
148. Le Bihan, T.; Goh, T.; Stewart, II; Salter, A. M.; Bukhman, Y. V.; Dharsee, M.; Ewing, R.; Wisniewski, J. R., Differential analysis of membrane proteins in mouse fore- and hindbrain using a label-free approach. *J Proteome Res* **2006,** 5, (10), 2701-10.
149. Ishihama, Y.; Sato, T.; Tabata, T.; Miyamoto, N.; Sagane, K.; Nagasu, T.; Oda, Y., Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. *Nat Biotechnol* **2005,** 23, (5), 617-21.
150. Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* **2003,** 100, (12), 6940-5.
151. Beynon, R. J.; Doherty, M. K.; Pratt, J. M.; Gaskell, S. J., Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods* **2005,** 2, (8), 587-9.

9 Acknowledgements

All the work was done in the department of Proteomics and Signal Transduction at Max-Planck Institute of Biochemistry. During my Ph.D study, I got much support and help from all people in our department, and I would like to say “thanks” to everyone. Special appreciation would like to give to:

Prof. Dr. Matthias Mann for the generous support and constructive supervision of the whole period of my study here, for giving the chance to learn and see the development of a super excellent MS-based proteomics group in the world, for his open leadership and more,

Prof. Dr. Jacek. R. Wiśniewski for his great help to be my co-supervisor, for his support for any kind, for his beautiful desktop pictures and more,

Prof. Dr. Matthias Mann and Prof. Dr. Jacek. R. Wiśniewski for proofreading of the thesis,

Leonie F. Waanders, Reinaldo Almeida, Nagarjuna Nagaraj, for their collaboration and discussion of projects,

Dr. Jesper V. Olsen, M. D. Alexandre Zougman, Dr. Boris Macek, Peter Bandilla, Dr. Marcus Krüger and Dr. Cuiping Pan, for their excellent and professional suggestions/feedback regarding MS and biology knowledge,

Guoqing Li, Dr. Yong Zhang, Chanchal Kumar, Dr. Jürgen Cox, Mario Oroshi, for their bioinformatic support,

Tine Klitmoeller, Theresa Schneider, for their enthusiastic help in different aspect to make my life here much more comfortable and happy,

all people organizing/participating parties/hiking/barbecue/festa/retreat together to make the life here more colorful,

Rong Shi, Cuiping Pan, Shubin Ren, Yong Zhang, Guoqing Li, Yanling Zhang, Lei Deng for sharing happiness and suffering abroad,

Jin’an, little Luoluo, my parents and little brother, for the endless love.

10 Curriculum Vitae

Name: Aiping Lu

Gender: Female

Nationality: P.R.China

Birthday: 30/08/1981

Family Status: Married

Email: scudshirley@hotmail.com

Tel: 0049 17696361225

Education Background

Master of Science, Biotechnology, August 2005

University of Southern Denmark, Odense, Denmark

Bachelor of Engineering, Biochemical Engineering, June 2003

Zhejiang University, Hangzhou, China

Research Experiences

2003 "*Characterization of proteins contaminating human epidermal growth factor (hEGF)secreted with Escherichia coli*", in the lab Prof. Shanjing Yao at Zhejiang University (bachelor project)

2004 "*optimization of procedures for enrichment of nucleic acid binding proteins using Heparin-sepharose*", in the group of Prof. Peter Roepstorff at University of Southern Denmark

2004-2005 "*Identification and Characterization of DNA-Binding Proteins involved in Diabetes Type 1 Using Affinity Purification and Mass Spectrometry*" in the group of Prof. Peter Roepstorff at University of Southern Denmark (master project)

2005-2008 Ph.D projects mainly involve MS/MS/MS spectra analysis, single protein sequence and modification mapping by using electrostatic spray robot Triversa-Nanomate, protein Neddylation with EGF stimulation by SILAC, membrane protein purification and comparative membrane protein analysis in three rat nerve tissue, in the lab of Prof. Dr. Matthias Mann at Max-Planck Institute of Biochemistry

Publications

- Aiping Lu, Leonie F. Waanders, Reinaldo Almeida, Guoqing Li, Mark Allen , Jürgen Cox, Jesper V. Olsen, Tiziana Bonaldi and Matthias Mann. "Nanoelectrospray peptide mapping

revisited: Composite survey spectra allow high dynamic range protein characterization without LCMS on an orbitrap mass spectrometer". International Journal of Mass Spectrometry. 268(2007):158-167.

- Nagarjuna Nagaraj#, Aiping Lu#, Matthias Mann, and Jacek R. Wiśniewski. "Detergent-based but Gel-free Method Allows Identification of Several Hundred Membrane Proteins in Single MS Runs". Journal of Proteome Research 2008. (#both authors contributed equally)
- Aiping Lu, Jacek R. Wiśniewski, and Matthias Mann. "Comparative Proteomic Profiling of membrane Proteins in Rat Cerebellum, Spinal Cord, and Sciatic Nerve". Submitted
- Aiping Lu, Alexandre Zougman, Marek Pudełko, Marek Bębenek, Piotr Ziółkowski, Matthias Mann, and Jacek R. Wiśniewski. "Mapping of Lysine Monomethylation of Linker Histones in Human Breast Cancer". Submitted

Honors and Awards

1998.11 Chemistry Competition of the National Senior Middle School students, Zhejiang Province, first class

1999.9-2000.9 scholarship of Zhejiang University, first class

2000.9-2001.9 scholarship of Zhejiang University, second class

2001.9-2002.9 scholarship of Zhejiang University, first class

2003.6 Graduated with excellent in Zhejiang University

2003.12 "Prins Joachim og Prinsesse Alexandras Internationale Uddannelsesfond" award (Denmark)

2004.5-2005.4 Novo scholarship Programme award (Denmark)