

TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Proteomik und Bioanalytik

Towards Comprehensive Identification of Proteins from MALDI imaging

Stefan Karl Maier

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. M. Klingenspor

Prüfer der Dissertation: 1. Univ.-Prof. Dr. B. Küster

2. apl. Prof. Dr. A. Walch

(Albert-Ludwigs-Universität Freiburg)

Die Dissertation wurde am 13.03.2014 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 25.06.2014 angenommen.

Je mehr ich sehe,
desto weniger weiß ich sicher.

John Lennon

Content

Abstract		iii
Zusammenfassung		v
Chapter 1	General Introduction	1
Chapter 2	Bottom up Characterization of the MALDI Matrix Proteome	27
Chapter 3	Proteoform Identification from MALDI Imaging Specimens using High Resolution Mass Spectrometry	41
Chapter 4	Proteoform Identification from Human Tumor Xenografts in Mice	51
Chapter 5	PAS-cal: Calibration Standards for MALDI Mass Spectrometry	61
Chapter 6	Final Conclusion	77
Appendix	Abbreviations and Acronyms	79
	References	81
	Publications	101
	Danksagung	103
	Curriculum vitae	105

Abstract

MALDI imaging mass spectrometry (IMS) is a powerful tool for the visualization of protein distribution in tissues. But as the identity of the detected proteins in most cases remains unknown the full potential -beyond diagnostic and prognostic purposes- cannot be unlocked.

To address this important challenge of the field and move towards the comprehensive protein identification from MALDI imaging studies a method for the extraction of the protein containing MALDI matrix layer was developed. The established procedure allowed the bottom-up identification of about 1,400 small, soluble and abundant proteins that constitute the set of proteins most likely detectable by MALDI imaging experiments of human tissues. The validity of this assumption was underlined by the conjuncture that over 90% of all proteins reported in MALDI imaging studies are included in the 'human MALDI matrix proteome'.

The potential of MALDI imaging is founded in the fact that the detected masses correspond to proteins including all their modifications (proteoforms). These proteoforms can therefore be seen as molecular probes for all processes the proteins underwent. As the usage of proteases corrupts the integrity of the intact mass bottom-up approaches cannot be utilized for the identification of such species. Therefore a top-down proteomics approach using high resolution LC-ESI-MS was employed and resulted in the identification of over 500 proteoforms derived from two MALDI imaging specimens. The analytical depth of this strategy could be increased by employing an iterative exclusion list strategy which prevented the repeated sequencing of high abundant proteoforms. Most of the identified species arose from N- and C-terminal fragmented proteins, indicating considerable protein processing activity in tissues.

The top-down proteomic approach was further applied to specimens obtained from human tumors xenotransplanted into mice. Owing to the comparatively long sequences of the 550 identified species, it was possible to distinguish whether the proteoform was derived from the host organism or from the tumor itself.

An important requirement for the concatenation of identified proteoforms to the imaging experiment is the accurate calibration of the MALDI MS spectra. Based on the PAS-sequence repeat a calibration standard was designed covering the typical mass range of imaging experiments. Based on the same sequence motive, a multi-peptide calibration standard tailored for bottom-up proteomics was synthesized and assessed for the calibration of MALDI MS and MS/MS spectra.

Zusammenfassung

MALDI Imaging Massenspektromie (IMS) ist eine leistungsfähige Technologie zur Abbildung von Proteinen in Geweben. Da die Identität der detektierten Proteine jedoch in den meisten Fällen ungeklärt bleibt, kann das volle Potential -über diagnostische und prognostische Anwendungen hinaus- nicht ausgeschöpft werden.

Um diese bedeutende Herausforderung des Feldes zu lösen und alle Proteine aus MALDI Imaging Studies zu identifizieren zu können, wurde zunächst eine Methode zur Extraktion der Protein-enthaltenden MALDI Matrix Schicht entwickelt. Die gezielte Extraktion erlaubte es, 1.400 kleine, leicht lösliche und abundante Proteine mittels eines bottom-up Ansatzes zu identifizieren. Diese Proteine stellen mit hoher Wahrscheinlichkeit die Gruppe der Proteine dar, welche in MALDI Imaging Experimenten von humanen Geweben detektierbar sind. Diese Vermutung konnte durch die Tatsache untermauert werden, dass sich 90% aller in MALDI Imaging Studien identifizierten Proteine im hier dargestellten "humanen MALDI Matrix Proteom" wiederfinden.

Das Potential von MALDI imaging liegt vor allem darin begründet, dass die detektierten Massen Proteinen einschließlich aller Modifikationen (Proteoformen) entsprechen. Proteoformen können daher als Sonde für alle Prozesse angesehen werden, die jene durchlaufen haben. Da der Einsatz von Proteasen die Integrität der Proteoformen zerstört, ist ein bottom-up Ansatz nicht für deren Sequenzaufklärung geeignet. Aus diesem Grund wurde zur Identifizierung von Proteoformen aus Matrixextrakten ein top-down Ansatz gewählt. Aus Proben zweier MALDI Imaging Proben konnten über 500 Proteoformen identifiziert werden. Die meisten dieser Spezies stammten von C- oder N-terminalen Proteinfragmenten. Dies lässt auf umfangreiche Proteinprozessierung im Geweben schließen. Ferner wurde der top-down Ansatz auf Proben von xenotransplantierte humanen Tumoren angewandt. Die vergleichsweise langen Sequenzen der 550 identifizierten Spezies erlaubten es hierbei, die Proteoform dem Tumor beziehungsweise dem Wirtsorganismus zuzuordnen.

Eine wichtige Voraussetzung, um die identifizierten Proteoformen mit der Masse aus dem Imaging Experiment vergleichen zu können, ist eine genaue Kalibrierung der MALDI Spektren. Basierend auf dem PAS Sequence Motiv wurde ein Standard etabliert, der den typischen MALDI Imaging Massenbereich abdeckt. Beruhend auf demselben Sequenz Motiv wurde ein Peptid-Kalibrations-Standard synthetisiert und hinsichtlich seiner Eignung als Kalibrant für MALDI MS und MS/MS charakterisiert.

Chapter 1:

General Introduction

Mass Spectrometry

General Principle

Mass spectrometry enables determining the mass-to-charge ratio of atoms and molecules. Since the pioneering work of J.J. Thomson [1] the general set-up of such an instrument has not changed and consist of three main parts (Figure 1-1)

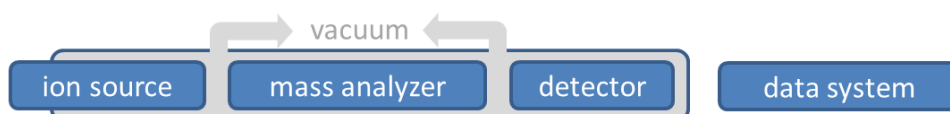


Figure 1-1: The three principal components of a mass spectrometer: The ion source, the mass analyzer and the detector. To minimize interactions of the gas-phase ions with gas particles a high vacuum has to be maintained in the complete system.

In the first part, the ion source, analyte molecules are charged and get transferred into the gas phase. The charge carrying analytes can easily be manipulated by means of electric or magnetic fields. This is exploited in the second component, the mass analyzer, where ions are separated depending on their mass-to-charge ratio (m/z). In the third part, the detector, the abundance of the separated ion species is determined. The data system creates the mass spectrum, a graphical representation of the m/z signals versus their abundance. As the instrument determines the mass-to-charge ratio (m/z) the actual molecular mass can be calculated if the charge state of the ion is known. The atomic mass unit (u or Da for Dalton) is defined by the convention that the ^{12}C isotope of carbon has a mass of exactly 12 u. As the mass directly depends on the (sub-) atomic composition, it is an inherent characteristic of atoms and

molecules and therefore one fundamental measure to determine their identity. For this reason mass spectrometry is used in many different fields for a huge number of analyte classes. Different ionization methods and analyzer types have been developed to address this variety. The mass-spectrometric characterization of proteins and peptides is a challenging analytical task as the molecular weight of proteins can range from a few kDa to MDa. This limits the number of applicable ionization methods and analyzers.

Ionization of proteins and peptides

Many of the available ionization methods are quite harsh and lead to fragmentation of the analyte molecules. Electrospray ionization (ESI) [2] and matrix-assisted laser desorption / ionization (MALDI) [3] allow 'soft' (meaning destruction free) generation of gas phase ions. The discovery of MALDI was influenced by the development of secondary ion mass spectrometry (SIMS) [4,5] and fast atom bombardment (FAB) [6] which allow for the ionization of smaller peptides but are mainly used for elemental analysis. Except for FAB all those ion sources can be used for imaging experiments.

Similar to fast atom bombardment, ionization in the SIMS process is achieved by a bombardment of the analytes with a primary high-energy particle beam penetrating into the sample. The penetrating atoms (FAB) or ions (SIMS) induce a collisional cascade which leads to desorption of some secondary ions into the gas phase. Two modes of SIMS that mainly differ in the intensity of the applied primary beam are available. Dynamic SIMS uses higher intensities and therefore penetrates deeper into the sample which enables profiling depths of nanometers to some hundred micrometers [7]. In contrast static SIMS is performed at lower beam intensities and mainly samples the surface region, leaving its structure intact [8]. Even in static mode the energy of primary beam particles is relatively high. Therefore, larger molecules tend to get fragmented during the process of ablation and ionization. Surface treatment approaches like metal-assisted (Meta-SIMS) or matrix enhanced SIMS (ME-SIMS) [9] allow expanding the accessible mass range from about 1000 Dalton to 3000 Dalton.

Instead of using fast ions or atoms also energy-rich electromagnetic beams can be employed for desorption and ionization. But only if assisted by the so called matrix, laser desorption ionization (LDI) facilitates the destruction free ionization of large molecules. The matrix is typically a derivate of an organic acid (Figure 1-2 A). A solution of it is mixed at huge molar excess with the analyte and 'co-crystallizes' after solvent evaporation. Common to all

substances used as MALDI matrix is a delocalized π -electron system which absorbs the energy of the employed laser allowing for an efficient, controllable and fast energy transfer by resonant absorption of the laser energy [10] without affecting the integrity of analyte molecules. The absorption damages the crystal lattice and leads to a rapid solid-to-gas phase transition, resulting in a plume (Figure 1-2 B) [11]. Even though introduced in 1988 [3,12] the mechanism of MALDI is still not completely understood. Two rivaling mechanisms are proposed. The lucky survivor model postulates that pre-charged analyte ions together with their counter ions are already embedded in the crystal before actual laser irradiation. Free ions are generated by dissociation of neutral clusters followed by counter-ion separation or by dissociation of charged clusters not followed by ion neutralization.

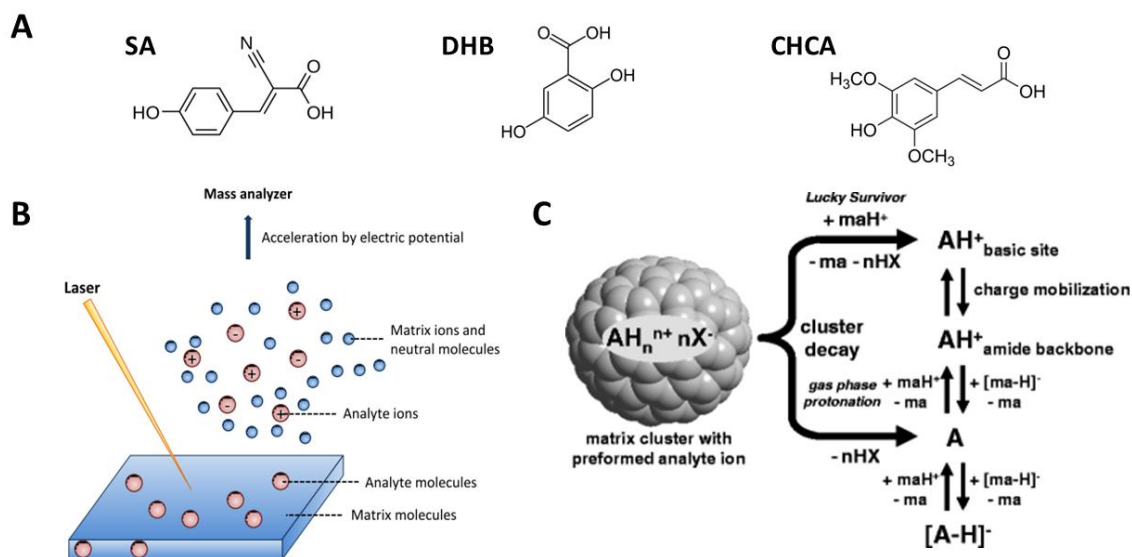


Figure 1-2: Matrix assisted laser desorption ionization.

A] MALDI matrix molecules with delocalized π -electron systems absorbing UV laser energy sorted from 'cold' to 'hot' (SA: sinapinic acid, DHB: 2,5-dihydroxybenzoic acid, CHCA or HCCA: alpha-cyano-4-hydroxycinnamic acid) **B]** Scheme of the MALDI process. [13] **C]** The unified protonation mechanism [13] postulates that both lucky survivor and gas phase protonation are parallel processes.

The gas phase ionization model on the opposite predicts that neutral analytes get only charged in the plume by collisions with (de)protonated species in the gas phase, e.g. matrix molecules. Both models seem to take place more or less in parallel but to what extent is influenced by a multitude of different parameters. Therefore the 'unified MALDI analyte protonation

mechanism' [14] which combines both models has been proposed (Figure 1-2 C). Even though the details may be subject of further research, the development of MALDI enables the destruction free or "soft" ionization of large organic molecules like intact proteins.

Another soft ionization method is electro spray ionization (ESI, Figure 1-3). It enables the destruction free transfer of a large variety of substances from the solvated state into gas phase [2] with almost no upper mass limitation. Due to the possibility to interface liquid chromatography with mass spectrometry by an ESI source, it became the most widely used ionization technique [15]. The first step in the ionization process is the dispersion of the liquid. This is accomplished by the application of a potential to a liquid filled nozzle. Two opposing forces affect the shape of the liquid. The surface tension tries to minimize the surface area and the electrical Coulomb attraction pulls the liquid into the direction of the counter electrode. The resulting elliptical shape inverts at a certain voltage into a pointed cone emitting a solvent jet. This jet becomes unstable and the spray consisting of small highly charged droplets is formed. Due to solvent evaporation those droplets shrink and the field density on the surface increases until the so-called Rayleigh limit is reached. At this point a droplet-based Taylor cone is formed resulting again in the formation of a spray of smaller droplets. This process is repeated by every droplet until the remaining charge is insufficient to reach the field density limit. Finally drops containing only one analyte molecule are produced and by further solvent evaporation a 'naked' ion resides (charge residue model [16]).

An alternative mechanism is proposed by the ion evaporation model [17]. Here, the field strength at the surface of the main droplet is rescued by expelling solvated ions. Experimental data suggests that both models are true but the charge residue model is in better agreement for larger molecules. Contrary to MALDI, ESI often generates multiple charged ions. Minimization of the flow rate and miniaturization of columns and needles led to the development of nanoESI. Due to flow rates in the range of a few nano liters per minute, the initial droplet size is very small, which leads to nearly 100% ionization efficiency [15].

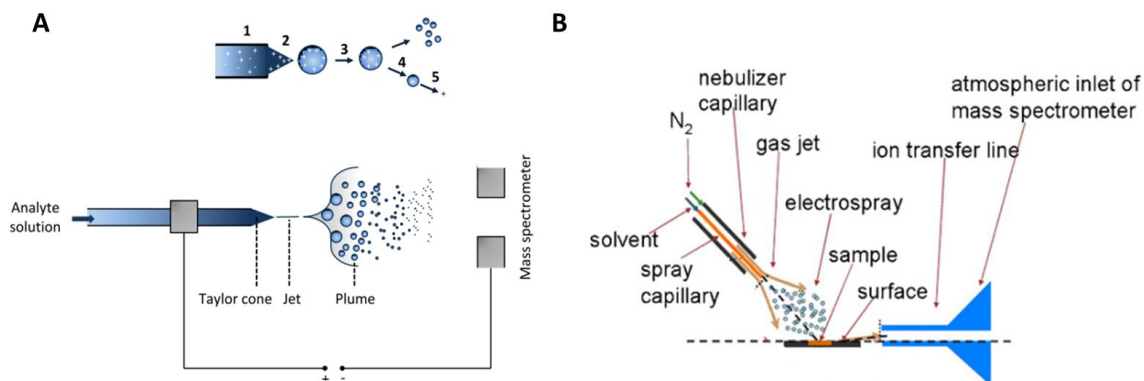


Figure 1-3: The electro spray process.

A] Formation of analyte ions with ESI. (1) Electrophoretic separation of anions and cations; (2) Taylor cone formation; (3) solvent evaporation from droplets; (4) Coulomb explosion and formation of single solvated analyte ions via charge residue or ion evaporation; (5) declustering of solvated ions. (adapted from [13]) **B]** Desorption ESI (DESI) ion source enabling surface sampling (adapted from [18]).

Mass analysis

After the ionization process the charged analytes are separated according to their mass-to-charge ratio. Different concepts are used for this purpose but all are based on the manipulation of the gas-phase ions by means of electrostatic or magnetic fields.

Conceptually the simplest principle is employed in the time-of-flight analyzer (TOF) (Figure 1-4 A) where ions are accelerated by an electrical potential into the direction of a detector. After leaving the electrical field all ions possess the same kinetic energy but have different m/z dependent velocities. Based on their velocity ions need different times to pass through the field-free drift zone. By recording the time of impact on the detector, the flight times are measured and m/z values can be calculated (Figure 1-4 B).

A quadrupole mass analyzer is constructed from four rods which are arranged in parallel around a common axis (Figure 1-4 C). A radio frequency voltage is applied to the opposing electrically paired rods. The rapidly changing electrical field induces an m/z dependent oscillating motion of the ion moving along the axis of the rods. Additionally to the radio frequency a direct current is applied to the paired rods which induces a m/z dependent deflection of traveling ions. This

leads to three scenarios: i) The amplitude of the radio frequency induced oscillation is too large and the ion is depleted. ii) The influence of the deflection leads to a collision with the rods. iii) The influence of both forces allows a stable trajectory and the ion passes through the quadrupole. Depending on the voltage settings a quadrupole can be used as an ion guide allowing a wide m/z range to pass or as a mass filter which only permits the passage of ions in a narrow m/z range.

A linear ion trap basically consists of a quadrupole with two additional electrodes on both ends imposing a potential barrier. This allows trapping of ions on stable trajectories inside of the trap. Mass analysis is realized by varying the applied voltages which leads to an m/z dependent ejection followed by the ion detection. Furthermore it is possible to use the analyzer for the isolation of certain ions. To achieve this, m/z ranges above and below the targeted range are selectively ejected. Three-dimensional traps follow the same principle but the four rods are replaced by a ring electrode (Figure 1-4 D).

The onion shaped orbitrap analyzer consists of an inner spindle electrode and an outer electrode which is divided into two parts (Figure 1-4 F). Ions oscillate in a static electrical field between both electrodes on stable trajectories. Of the three possible oscillations, only the frequency of the movement along the spindle axis is dependent on the m/z value. This frequency is detected as an image current induced between the two parts of the outer electrode. The m/z value of the oscillating ion is deduced using Fourier transformation. Even though the principle of the orbitrap was known for a long time (Kingdon trap [19]) it could not be utilized because the controlled injection of ion packages into the orbitrap was not possible. Only with the introduction of a c-shaped trap this problem was solved [20].

Based on the separation principle, all analyzer types have different performance specifications (Figure 1-5) which make them advantageous for different applications. Besides all other parameters the resolution is an important criterion for the ability of a mass analyzer to resolve different molecular species with similar but distinct masses [21]. Low resolution instruments only allow determining the peak apex of the isotopic envelope and therefore the calculation of the average m/z value. High resolution analyzers are capable to separate the single isotopic peaks and therefore facilitate to deduce the charge and monoisotopic mass (the mass of a molecule containing just the most abundant isotopes for each element).

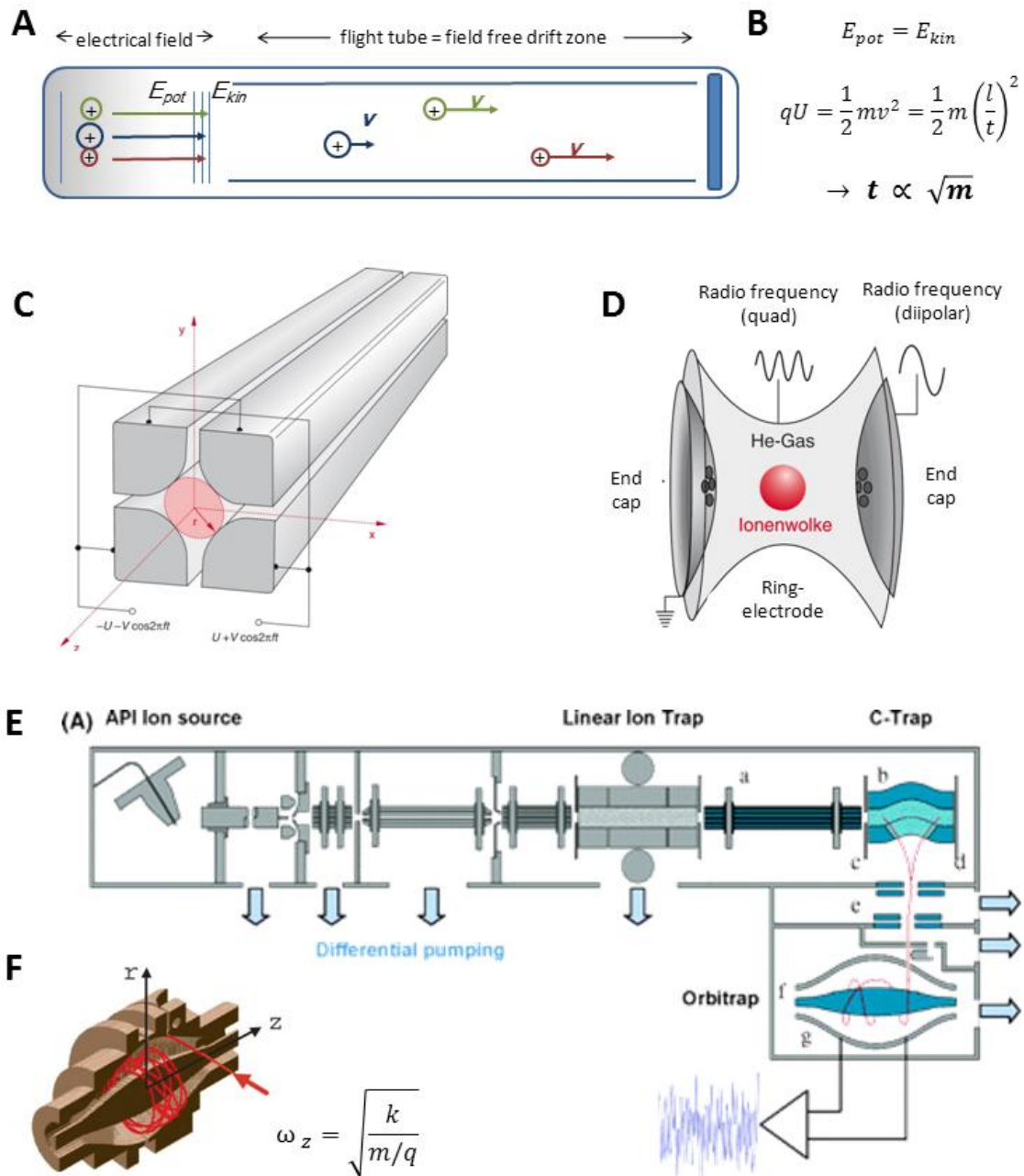


Figure 1-4 Different mass analyzers.

A] Schematic representation of a time-of-flight analyzer: Ions are accelerated by an electrical field into the flight tube which has a defined length. All ions possess the same potential energy but m/z dependent velocities. The velocity results in different impact times on the detector.

B] The 'TOF' equation: showing the proportionality of the flight time and the square root of the mass.

C] A Quadrupole analyzer consists of four parallel rods imposing a quadrupolar field. The latter is generated by applying an oscillating electrostatic field to the two opposing rod pairs.

(adapted from [22] **D**) In a 3-dimensional ion trap a quadrupolar field is imposed by a ring electrode. Ions can be trapped by the potential barrier imposed by the end caps (adapted from [22]. **E**) Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. The linear ion trap can be used as a mass analyzer or as a device for ion accumulation and selection. The accumulated ions are transferred to the c-trap which focuses the ions and introduces them into the Orbitrap analyzer. (adapted from [20]). **F**) cutaway view of the orbitrap mass analyzer (adapted from [23]).

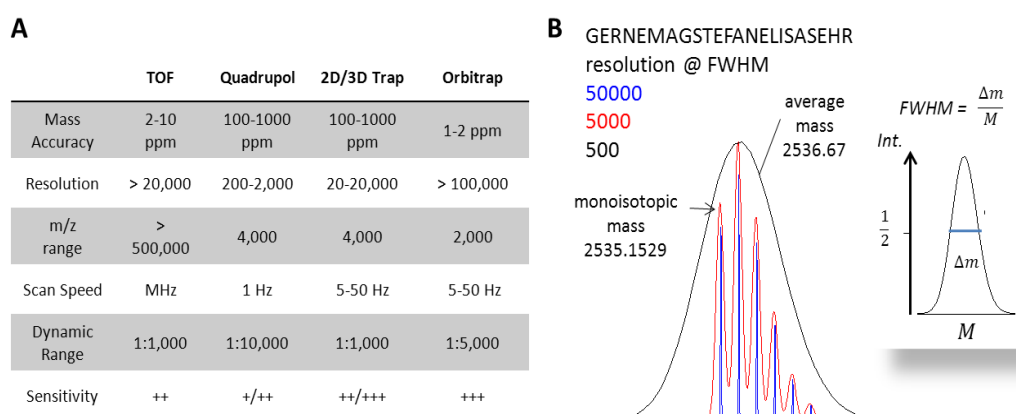


Figure 1-5: Specifications of different mass analyzers.

A] Standard values for different analyzer types. (Adapted from [24,25]). **B**] Calculated peak shapes for a peptide at different resolutions. The peak shape at 500 corresponds to the isotopic envelope. The insert shows the definition of FWHM (full width at half maximum) (simulated with Xcalibur).

As mentioned before, some analyzers can be used to selectively filter or isolate a certain m/z ranges. This is exploited for an experiment termed tandem MS or MS/MS where ions are selected, fragmented and detected subsequently. The information of the resulting fragment spectrum delivers additional information about the structure of the fragmented ion species. Those experiments can be carried out in time (e.g. both in an ion trap) or in space (isolation and fragmentation in different analyzers). The latter is performed on so called hybrid instruments (e.g. Figure 1-4 E). Popular combinations are Trap-Orbitrap or Quadrupole-TOF. Also TOF analyzers allow a flight time dependent selection of ions. For this purpose a charged grid capable of deflecting ions is placed in the flight tube. As all parameters (m/z value of ions to

pass, distance of the ion source to the grid and therefore the time at which the ions arrive at the grid) are known, the voltage of the deflecting ion gate is only switched off at that very moment a certain m/z range arrives, allowing these to pass. As the fragment ions, created after acceleration, obtain the same velocity as the precursor ion (and therefore arrive the grid at the same time) this technique can be used for the acquisition of MS/MS spectra in MALDI-TOF instruments.

Imaging Mass Spectrometry

Imaging techniques in tissue research

Image generating techniques influence the way we can see the world and often led to dramatic changes in its perception. The discovery of the X-ray for example not only enables visualizing fractures of bones but also allows elucidating the structure of crystalized proteins and DNA. The invention of the light microscope has basically revolutionized the concepts of biology and is still one of the fundamental tools in many areas like material science, microbiology and histology. Nearly as long as microscopy itself, dye-based approaches have been used to improve contrast and therefore the interpretability of microscopic preparations. The field of histochemistry uses the specific interactions of dyes with structures or substance classes. The most prominent example is the combination of Hematoxilin and Eosin, allowing a rough differentiation of nuclei and protein-rich structures. Immunohistochemistry takes advantage of the high specificity of antibodies against a target molecule structure. After an antibody has bound to its target, an enzymatic staining reaction or a fluorescent dye is used to unravel distribution and abundance of the targeted structure. Small analytes like drugs are difficult to be detected by antibodies. A method which allows investigation of small molecules is autoradiography. Here an atom of a molecule is exchanged with its radioactive counterpart allowing to visualize the distribution of the labeled substance in a biological system from whole body level down to microscopic dimensions [26]. Immunohistochemistry and autoradiography are very powerful tools and histochemistry is indispensable in the field of light microscopy. A shortcoming of all these techniques is that they only have a high chemical specificity or even need prior knowledge of the target. In contrast, mass spectrometric imaging approaches do not depend on *a priori* assumptions but enable a less biased discovery of correlations between the distribution of different molecular features and the morphology of the underlying tissue.

Imaging Mass spectrometry

Imaging mass spectrometry (IMS) or mass spectrometry imaging (MSI) allows the comprehensive analysis of the spatial intensity distribution profiles of different molecular species [27]. Mass spectra are acquired in a spatially resolved fashion from defined positions on the surface of a planar sample (Figure 1-6 A, B). Analogue to digital imaging each spectrum is linked to a position on a two-dimensional grid and represents a single pixel that can be reconstructed into a raster image (Figure 1-6 C).

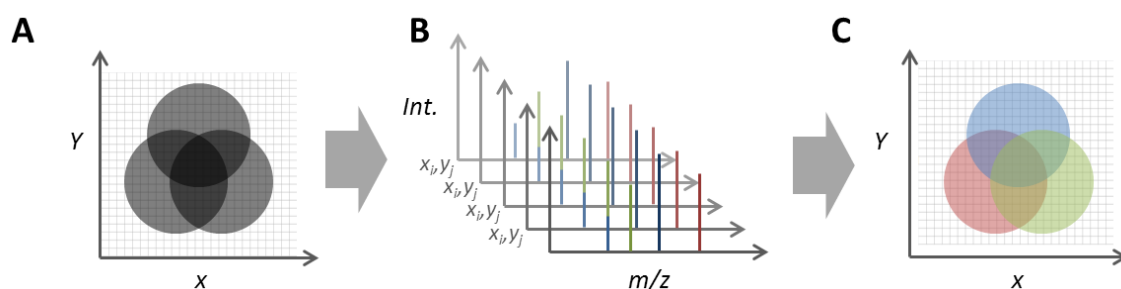


Figure 1-6: General principle of imaging mass spectrometry.

A] Imaging specimen placed on a 2-dimensional grid **B**] Acquisition of mass spectra for each coordinate x_i, y_j ('pixel') of the grid **C**] False color reconstruction of the distribution of the mass values.

For imaging MS three major ionization methods are used. Two of them are based on desorption and ionization of analytes from a solid surface via a beam of ions generating secondary ions (secondary ion mass spectrometry SIMS) or a beam of photons (laser desorption ionization LDI). The third one employs a modified set-up based on electro spray desorption and ionization (DESI, Figure 1-3 B). Secondary ion mass spectrometry was the first ionization method employed for imaging experiments [28,29] and is widely used in the field of material sciences [30]. Static SIMS leaves the sample intact and mainly samples the surface region. It offers a lateral resolution [8] below $1 \mu\text{m}$ which corresponds to the size of mitochondria [31] (Figure 1-7 A) and enables the visualization of subcellular structures. The achievable lateral resolution and the accessible mass range makes SIMS imaging a very attractive tool for the detailed visualization of a variety of different substance classes like inorganic ions [32], lipids [33,34] as well as drugs and their metabolites [26].

MALDI allows overcoming the mass range limitations of SIMS and enables to visualize the distribution of a wide variety of different molecules. Which analyte class is preferentially ionized and therefore imageable is to a huge extent determined by the matrix. A number of different approaches has been reported over the last 17 years including imaging experiments focusing on proteins and peptides [35] or the spatial expression of N-linked glycans [36], lipids [37] or sulfatides [38]. Also combinatorial imaging approaches -targeting different analytes- are possible like for example the investigation of the distribution of drugs and their metabolites [39] or the co-localization of proteins and drugs in whole body animal sections [40] were reported. The achievable lateral resolution is on the one side determined by the laser wavelength and on the other side by the minimal surface area that needs to be irradiated in order to ablate and ionize a sufficient number of ions to reach the limit of detection. Furthermore the sample preparation and its influence on matrix crystal size are critical factors which define the regularly achievable lateral resolution of MALDI imaging experiments to the range of 50-200 μm . which is equal to one or some cells per pixel. Only in recent years some studies reported lateral resolution down to 20 μm [41] or even subcellular resolution [42] .

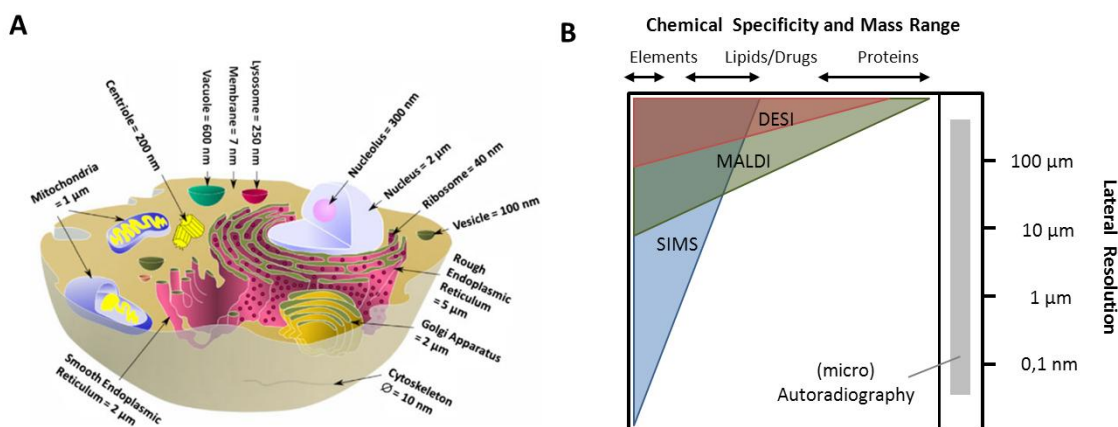


Figure 1-7: Dimensions of Biological systems and capabilities of ionization methods to resolve such structures . A| Schematic picture of an eukaryotic animal cell and size ranges of the organelles [31]. **B|** Achievable lateral resolution and chemical selectivity for different ionization techniques used in MSI [43] compared to micro autography (added from [44]).

Other than in MALDI and SIMS which are by default equipped with a movable two-dimensional sample, ESI sources are usually static. One solution enabling ESI imaging is the liquid extraction

from defined spots on the surface of the sample, followed by direct infusion ESI-MS measurements. An automated sampling and spraying device is described [45] but due to dispersion of the liquid on the sample surface this technique only allows minor lateral resolution. A more sophisticated solution is desorption electrospray ionization (DESI) [46] (Figure 1-3 B). In this setup, the ESI spray is directed onto the sample surface and the charged droplets ablate analyte molecules. The 'loaded' droplets transport the analyte molecules in the direction of the counter-electrode (inlet of the MS) while the droplet based ESI process takes place. An advantage of the technique is that it is carried out under atmospheric conditions and no perfectly plane surface has to be prepared. The system has been applied for different samples and analyte classes, such as lipids of the human lens [47], or mouse brain [48] as well as for metabolites of bacteria [49] or chemical profiling of banknotes [50].

The different imaging techniques cover most of the analytes in biological systems. Furthermore, MS imaging workflows need little to no sample preparation while preserving tissue integrity. The latter allows the combination of different (traditional) imaging techniques on the basis of the very same sample. Imaging mass spectrometry is therefore a highly valuable addition to the toolkit of microscopy-based tissue research. Especially MALDI imaging is appealing as it enables the *in situ* analysis of proteins and peptides.

MALDI Imaging mass spectrometry of proteins and peptides

Proteins are polymers assembled from 20 different amino acids. The chain length can vary from some to hundreds of amino acids. Determined by their structure, they perform a variety of different tasks, such as substance transport and binding, intra- and extracellular signaling and catalyze metabolic reactions. The structure itself is determined by the amino acid sequence which is encoded in the genome. Today it is believed that humans possess about 20,000 protein coding genes [51]. Due to additional modifications during transcription and translation as well as posttranslational modifications on the protein level, the variety of different proteoforms Figure 1-8 [52] is tremendous. The proteome (the entirety of all proteoforms under a given condition [53]) varies from cell type to cell type and reflects the adaptation of a biological system to different environments and conditions. Elucidating these changes is the goal of proteomics as the large-scale study of the proteome [54]. MALDI imaging of proteins or better proteoforms is therefore very appealing as it enables taking spatially resolved snapshots of the proteome in the context of tissue morphology.

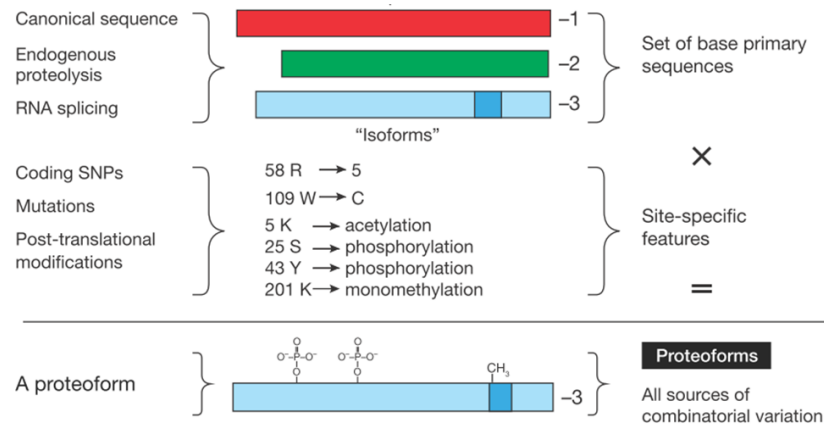


Figure 1-8: Possible modifications events from protein sequence to the final proteoform.

(from [52])

From tissue sections to false color proteoforms distribution portraits

The first step in every tissue centered research is the preservation of the collected specimens. Activity of endogenous proteases will not only change the morphology but will also produce artificial proteoforms. As it preserves the histological features and allows storage at room temperature for long time periods formalin fixation is the most frequently used preservation technique. Formaldehyde treatment leads to inactivation, stabilization and immobilization of proteins by its reaction with N-termini of amino acids as well as the side chains of arginine, cysteine, histidine and lysine. Furthermore intermolecular crosslinks by methylene bridges are formed [55]. Snap freezing of tissue specimens causes no additional modification and sample stability in liquid nitrogen is guaranteed over years [56]. But as storage of such samples is expensive the number of snap frozen specimens is low compared to the rich archives of formalin fixed preparations. An alternative is alcohol-based preservation as it unifies the benefits of the cryo- and FFPE preservation and allows storage at moderate temperatures without severe alteration in the chemistry or morphology [57]. Especially the crosslinks in FFPE samples make MALDI imaging of a formalin fixed specimens impossible without prior treatment. Being a quite recent technique alcohol fixation seems promising, but cryo-specimens are still the most widely used samples for protein and peptide imaging.

After preservation the next step is the sectioning of the tissue samples. The sections typically have a thickness of 3 to 20 μm and are mounted onto a solid surface i.e. a conductive

microscopic slide. In the case of fresh frozen tissue this is carried out in a cryotome. Depending on the analytes of interest and the tissue type, the sample is subjected to a defined washing procedure which removes interfering substances but keeps the analytes in place. For proteins typically a multi-step washing procedure with increasing alcohol concentrations is carried out. This treatment removes salts, water and fats which would suppress ionization and at the same time it fixates the proteins [58].

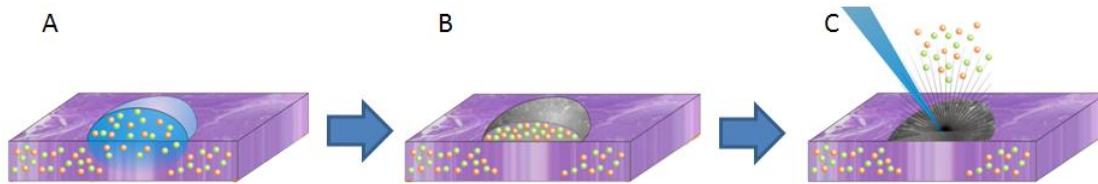


Figure 1-9: Schematic representation of the matrix application process

A| Protein micro extraction by matrix solution. **B|** Mixture of MALDI matrix and proteins after solvent evaporation **C|** Laser desorption of analytes co-crystallized with the matrix molecule.es

As mentioned before, the MALDI process is dependent on a crystalline matrix layer with embedded analyte molecules. In most cases this layer is prepared by the application of a solution of matrix molecules in a mixture of water and an organic solvent (Figure 1-9). Besides dissolving the matrix molecules the liquid droplets also act as micro-chambers, extracting proteins from the underlying tissue. The larger the covered surface area, the wider the analyte can diffuse away from its point of extraction [59]. This as well as the size of the matrix crystals limits the achievable lateral resolution. Therefore the aim is to prepare a homogenous matrix layer with small crystals from small droplets while achieving sufficient extraction from the sample. Besides manual application using simple nozzles, different automated systems are available for a reproducible matrix application. One uses the ink printer principle [60] another one generates the droplets via vibrational vaporization [61] or the pneumatic atomization [62]. Alternatively the matrix layer can be prepared by sublimation which circumvents the floating bias [63]. It is also possible to introduce an additional spraying step before the application of the matrix solution and apply a protease for on tissue digestion [42].

The typical mass range for MALDI-TOF imaging of proteins is between 2 kDa and 20 kDa. The detection of larger species is compromised due to the decreasing sensitivity of the instrument for increasing masses as well as due to the suppression of the ionization of larger species. The

latter can be addressed by more stringent washing procedures [64,65] and even imaging of species above 100 kDa has recently been reported [66]. As the sample leaves the instrument basically undestroyed also successive circles of washing and coating are possible which allows a stepwise coverage of different mass ranges or different analyte types. Furthermore all classical staining techniques and immunohistochemical methods are applicable and digital microscopy can be performed from the very same microscopic slide as used for imaging. Thereby the morphological information of traditional histology can be complemented with the distribution of molecular species making MALDI imaging a molecular microscope [67].

Not only can the image generation within a tissue help to address question like tumor heterogeneity [68] but it is also a very powerful tool if used for the comparison of larger patient cohorts with for example known clinical endpoints. Such studies are quite challenging as not only experimental conditions have to be set up very reproducible but also the data processing and analysis are critical steps for robust results. Data processing aims to remove sources of variation other than biologically relevant information [69]. These processes include normalization of spectra, baseline correction, alignment and calibration as well as peak detection and peak integration. The processing results in lists of m/z values and their respective peak area linked to the x and y coordinates. This list is the basis for visualization and also for further analysis. The aim of the latter is for example to unravel feature patterns which correlate to regions of interest or patient outcome and to test the validity of those findings with sophisticated statistical methods. A detailed review on this topic can be found elsewhere [70].

Biomedical applications of MALDI imaging of proteins

The eye is a highly structured organ but at the same time the proteome especially of the lens is quite simple and mainly consists of few proteins with only little turnover [71]. Hence changes in the proteome of the lens are mainly at the level of post-translational modifications. Posttranslational modifications also play an important role in the indisposition of the eye, making it both a relevant and ideal model organ for MALDI imaging. Studies have investigated the role of the localization of α -crystallin and its modified and truncated forms in the context of cataract development [72–76]. By applying stringent washing conditions or on tissue digestion also the localization of integral membrane proteins in the lens and the retina were elucidated [77–79]. The investigation of glioblastoma was one of the first examples [80] showing the potential of the technology and also connecting two main fields in biomedical research where many MALDI imaging studies are carried out: Cancer research and neuroscience. Important

aspects in the field of cancer research are classification and prognosis prediction. In this respect, MALDI imaging has proven its power for many different cancer types. In breast cancer for example the expression level of human epidermal growth factor receptor 2 (HER2) is a prognostic marker for the response to a treatment with trastuzumab. Walch and coworkers show that the HER2 status can be correlated with peak patterns derived from MALDI imaging [81]. The predictor built from this study was shown to be transferable to gastric tumors [82]. The same group published a study where they demonstrated that patient outcomes in intestinal-type gastric cancer can be correlated with a set of seven proteinous species. Similar study designs were applied in Barrett's adenocarcinoma [83], papillary thyroid carcinomas [84] or for the classification of six different common cancer types [85]. Oppenheimer and coworkers used MALDI imaging in clear cell renal cell carcinoma for the reliable identification of margins separating normal tissue and tumor tissue by aberrant molecular patterns in the tumor microenvironment (Oppenheimer et al., 2010). The brain is not only an important organ but, due to its inherent bilateral symmetry and extensive substructure, also an interesting system for method development in MALDI imaging [69]. One of the first protein MALDI IMS studies was the molecular profiling of proteins and peptides in a Parkinson's disease model as well as the observations of changes based on the administration of L-Dopa [86]. Another study focused on the isoform-specific profiles of the myelin basic protein in different regions throughout the mouse brain under normal and pathological conditions [87]. Latter was induced by the usage of ethidium bromide leading to myelin disorders which play a role in multiple sclerosis. Neuropeptides are secreted molecules acting as endocrine or paracrine transmitters and play a central role in neuronal signaling. First experiments targeting the distribution of neuropeptides were carried out using exocrine tissues and ganglia from sea slugs (*Aplysia californica*) [88].

The application of on-tissue digestion allows the peptide based imaging of FFPE tissue specimen. This is especially interesting because of the rich archives of formalin fixed tissues alongside with clinical data that is available. Tissue microarrays (TMA, basically a collection of different tissues in a block of paraffin) are especially interesting as they allow the uniform processing and measurement of many samples in parallel. Using such a TMA with biopsies derived from adenocarcinoma and squamous cell carcinoma was imaged and the resulting data was used to build a classifier, using a support vector machine algorithm to separate both cancer types [89]. In another TMA-based study, principal component analyses was used to establish molecular profiles which allow the distinction of normal and cancer tissue in pancreas. This study is interesting as they employed an ion mobility analyzer and therefore were able to

separate isobaric ions [90]. Based on FFPE tissue and on-tissue digestion, it was demonstrated that a proteomic profile could successfully classify Spitz nevus (SN) and Spitzoid malignant melanoma (SMM) with high specificity and sensitivity [91] .

Challenges in MALDI imaging of proteins

MALDI imaging of proteins has proven its power but is still an evolving field. Like every analytical technology there is always a striving for improvement. One of the main challenges in the field is the lateral resolution which is not only limited by sample preparation and crystallization but also by the smallest achievable diameter of the laser, which is dependent on the wavelength. A smaller laser irradiated area implicates that fewer ions are created, which leads to the need for higher detection sensitivity. Furthermore the higher the lateral resolution, the more data points relative to the area have to be acquired which limits through-put, drives the need for faster acquisition and increases the amount of generated data. Related to the sensitivity issue there is a need for a higher dynamic range, which is especially limited due to the relative high noise level in spectra from TOF analyzers if coupled to MALDI sources. For a comprehensive analysis of the proteome the accessible mass range is still too small. Improvements in mass accuracy and resolution are needed to discriminate overlapping species and ion mobility might be needed to address the problem of isobaric species. Furthermore quantification in MALDI is not trivial as changes in peak intensity can have many reasons and are not directly linkable to relative changes in peak abundances. This is especially true for MALDI imaging as the biological matrix is very complex and influences the ionization efficiency. Nevertheless MALDI imaging has proven its huge potential in classification and prognosis prediction, which would allow for applications similar to the biotyper [92]. The latter uses spectral libraries of pure bacterial colonies to identify bacteria by comparing protein profiles. One of the main remaining issues of protein imaging is the identification of the proteins constituting the MALDI MS peaks. The next paragraphs will give an overview how protein identification is addressed in the field of proteomics and how these principles can be applied to identify proteins from MALDI imaging.

Mass Spectrometry based Protein Identification

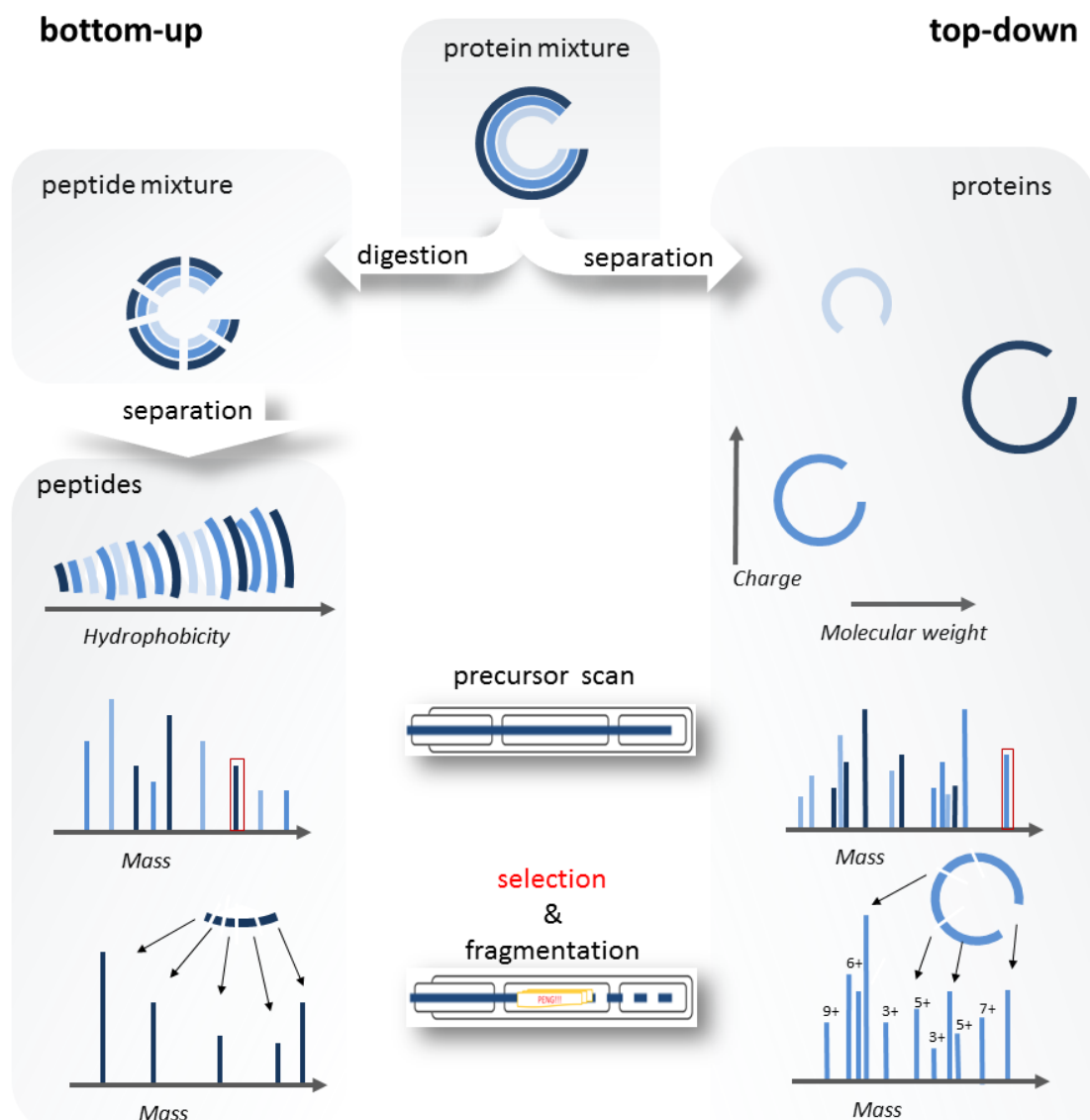


Figure 1-10: Schematic representation of bottom-up and top-down workflows.

A **Bottom-up** strategy always involves an enzymatic digestion. The complex peptide mixture is separated and precursor ion spectra are acquired. Based on that, peptide precursors get selected and fragmented. The resulting fragment spectra allow deriving (partial) sequence information used for peptide based protein identification. **Top-down** approaches aim to identify intact proteoforms. Therefore no enzymatic cleavage is used and separation is carried out on protein level. Mass spectrometric analysis is similar but the derived data is more complex. Some workflows start with top-down protein separation and employ proteolytic cleavage at a later stage.

Developed 30 years ago by Mullis [93] the polymerase chain reaction facilitates the amplification of DNA. Millions of copies can be created starting from only two antiparallel molecules per cell. The nearly unlimited availability of sample material dramatically lowered the sensitivity requirements for analytical techniques and catalyzed the investigation of genes and sequencing of whole genomes [94,95]. In contrast proteins cannot be amplified in a similar fashion. The characterization of the proteome composed from a complex mixture of large and small, physico-chemically different proteins with a dynamic range from some to millions of copies per cell (not even accounting for their modification status) is a very challenging analytical task [96]. Two principal strategies for the analysis of complex protein mixtures are used today. The bottom-up approach adopts the idea of shotgun sequencing of DNA and involves the proteolytic cleavage of proteins into peptides of a similar length. Top-down proteomic approaches aim to completely characterize intact proteins or better proteoforms. Both strategies include sophisticated separation techniques as only purified species can be sequenced by mass spectrometry.

Protein identification by bottom-up proteomics

Decreasing diversity by increasing complexity

Nowadays protein identification in most cases refers to bottom-up proteomic approaches as being the most widely used technique [97] for this purpose. Common to all bottom-up workflows is the usage of a protease and the mass determination of the generated peptides. Trypsin is the most frequently used enzyme as it cuts sequence specific after the two basic amino acids arginine and lysine generating peptides with a basic residue at the C-terminus. This is beneficial for the ionization efficiency and charge stabilization of the ions [98]. Furthermore the frequency of arginine and lysine leads to a quite even peptide length distribution having a maximum at about 11 amino acids [99]. Even though digestion increases sample complexity [100], the resulting peptides have similar properties and are easy to handle. The sample complexity has to be reduced prior to mass spectrometric analysis to meet the specifications of the instrument in terms of detection limit, dynamic range and scan speed [101]. This is often achieved by chromatographic separation, where complexity is resolved on a time scale by 'sorting' and thereby concentrating the analytes in narrow time increments. The standard for this purpose is reversed phase chromatography where peptides are separated due to their

different hydrophobicities. For this purpose analytes have to be solved in an aqueous solvent and are transported with the flow of the so called mobile phase through a column. The column is filled with small particles possessing a hydrophobic surface that interacts with the peptides and retains them. As the hydrophobicity of peptides is dependent on their amino acids sequence [102] peptides show different affinity to the stationary phase and different solubility in the mobile phase. Local equilibria between the bound state (retention) and the solved state (transport) for each molecular species are established until the end of the column is reached. In gradient elution an over-time increasing percentage of a second more hydrophobic solvent is mixed into the mobile phase. The second solvent weakens the interaction of the peptides with the stationary phase and allows tuning the analytes' retention. Reverse phase chromatography not only offers excellent separation and continuous elution of peptides but also the composition of the mobile phase can be chosen to be compatible with electrospray ionization allowing to directly interface it with a mass spectrometer.

Depending on the complexity of the sample as well as the performance of the LC-MS setup, additional prior separation steps might be needed to achieve the desired analytical depth. Historically gel-based electrophoretic separations on protein level were employed followed by in-gel digestion and LC-MS [103]. A drawback of gel based approaches and especially of isoelectric focusing, which is used as the first dimension of separation in two-dimensional gel separation, is the loss of hydrophobic proteins (i.e. membrane bound), extremely basic or acidic proteins and the low peptide recovery yields of in-gel digestion workflows. To address these issues in-solution and filter aided digestion [104] became popular over the last years often combined with multi-dimensional peptide separation strategies [105]. The final step in the separation process, the isolation of single molecular species, is carried out in the mass analyzer.

Peptide based protein identification

An early MS based protein identification strategy is the peptide mass fingerprint (PMF) where the experimental peptide masses are compared to the peptide masses generated from the digestion of known proteins. This technique allows the unambiguous identification of already characterized proteins if the digested sample contains a few, in an ideal case only one, protein. The drawback of this approach is the need for a sufficient amount of purified protein for the analysis. Small proteins are hard to identify as the number of possible peptides is low which makes the matching less specific. If sequence information is available the described workflow can also be implemented *in silico*. Here the theoretical fragment masses can be calculated from

protein sequences [106,107]. This also allows considering modifications of amino acid residues as they directly influence the peptide mass. This approach is mainly combined with two-dimensional gel electrophoresis as this technique offers a very high separation power on protein level [108].

As mentioned, single peptide species can be isolated by means of different analyzers. The selected species (precursor) can then be fragmented. As peptides tend to fragment along their backbone it is often possible to deduce (partial) peptide sequence information. To distinguish the possible peptide fragments Roepstorff and Fohlman [109] suggested the nomenclature as depicted in Figure 1-11.

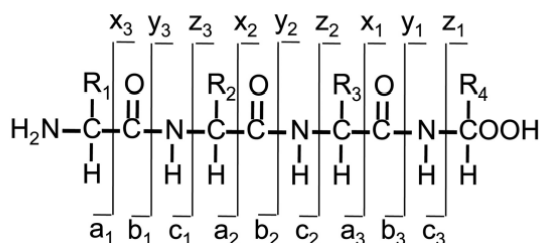


Figure 1-11 Roepstorff and Fohlman Nomenclature.

Possible fragmentation locations and the names of the resulting ions. (adapted from [110]).

A complete ion series would enable to (*de novo*) sequence [110] a peptide by matching peak distances with the residue masses of the single amino acids. But spectra where this is possible are rare. Depending on the resolution and accuracy of the spectrum, some amino acids with similar molecular masses cannot be distinguished (i.e. glutamine and lysine $\Delta m = 0.0364$ Da). Furthermore the tendency of a certain peptide bond to break is dependent on the neighboring amino acids. Therefore not all possible fragment ions have the same intensity or probability of being generated. Also the way in which ions are fragmented has a major influence on the detectable ions. The most commonly used fragmentation techniques are collision-induced dissociation (trap-type CID), higher energy collisional dissociation (HCD or beam-type CID) as well as electron transfer (ETD) [111] and electron capture dissociation (ECD) [112]. On MALDI instruments additionally laser induced dissociation (LID) and in-source decay (ISD) are available. Fragmentation in CID is induced by collisions of the gas phase molecules with an inert gas. These collisions lead to vibrational excitation and finally to breakage of intermolecular bonds. Depending on the kinetic energy of the collision, two types of CID are distinguished. In low-

energy CID, which is mainly carried out in trap analyzers, multiple collisions take place until sufficient vibrational energy is taken up for fragmentation. The energy is distributed within the molecule and eventually results in the dissociation at the weakest bonds. This is in most cases the peptide bond resulting in spectra populated by b- and y- ions. In higher energy dissociation or also beam-type CID, the transferred energies are higher and the dissociation happens nearly instantaneously. This leads to a more random dissociation resulting in spectra with a higher probability of complete b- and y- ion series but also with more internal fragments and neutral losses. This technique is available on quadrupole-TOF instruments or the HCD cell in Orbitrap hybrids instruments. In ECD and ETD electrons are directly or indirectly transferred to cations in the gas phase. The unpaired electron destabilizes the molecule and leads to rapid bond cleavage of the peptide backbone at the N-C alpha bond resulting in c- and z- ions.

The acquisition of fragment spectra is in most cases carried out in a data dependent fashion. First all ions present at this time are scanned in a full scan. Based on this scan and parameters like charge, m/z value and the intensity precursors are selected for fragmentation in the next scans. This sequence of full scan and fragment scans is repeated (duty cycle). To prevent sampling of the same peak already picked m/z values are excluded from fragmentation for a certain time. The outcome of such an experiment is a list of picked precursors with the corresponding fragment spectra. This information even with incomplete ion series can be used by a search engine for protein identification (Figure 1-12).

Based on the precursor mass the search algorithm calculates all possible sequence combinations from a protein database (Figure 1-12 A). As the searching algorithm has only to consider proteolytic peptides the number of possible matching peptide sequences is restricted which limits the searching space and makes the search more specific. For every sequence in this list, theoretical fragment spectra are calculated and compared to the experimental spectrum. Based on the agreement of theoretical and experimental precursor mass as well as the similarity of the theoretical and experimental fragment spectra each calculated hit is scored and ranked. The best scoring entry has the highest probability of being a correct assignment and identifies a protein based on the peptide identification. But even high scoring hits might be wrong. To get an estimate for the frequency of false positive hits of a dataset two approaches are frequently used (Figure 1-12 B). One strategy, the so-called decoy strategy, uses a second search against an additional database which in most cases a randomized or inverted version of the original database. A hit in the decoy database is defined as a false positive identification. The number of hits in the original and the decoy database is an estimate for the global false

discovery rate (FDR) of the search against the real database. The second employed method for FDR estimation uses Bayesian statistics to model a local peptide probability.

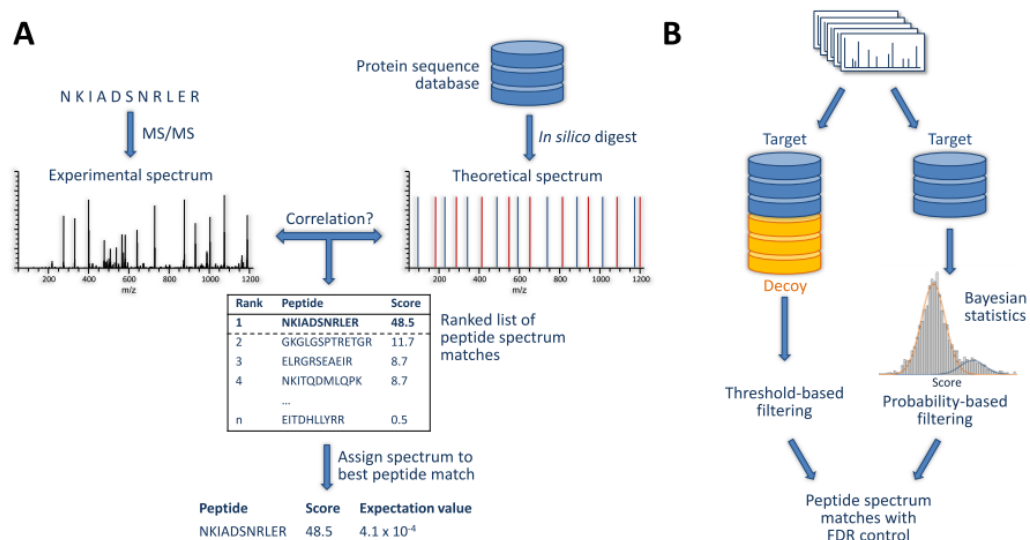


Figure 1-12: MS/MS spectrum based peptide and protein identification.

A| Based on the precursor mass possible sequences are calculated for database entries. For all sequences a theoretical spectrum is calculated and compared with the experimental spectrum. Possible hits are scored and ranked. **B|** Two strategies for false identification rates calculation. The target decoy strategy uses a search against a second 'nonsense' database. Probability filtering uses Bayesian statistics for the FDR calculation of a dataset. (adapted from [13]).

Even though digestion of a complex protein mixture introduces even more complexity bottom-up approaches are very successful in identifying a large number of proteins. A reason for the success of this approach is that in contrast to proteins, peptides have similar properties and are comparably easy to handle. However, their properties are different enough to reduce the complexity of the peptide mixture by means of different chromatographic separation techniques and adapt it to the capabilities (scan speed and sensitivity) of the employed mass spectrometer. State of the art instrumentation allows to profile near to complete proteomes within an hour [113]. This is to a huge extend a result of fast scanning, high resolution and high accuracy mass spectrometers coupled to nanoLC-ESI systems offering high separation power and sensitivity.

Besides the technically driven progress in the successful identification of as many peptides and proteins as possible, proteomic techniques also allow for quantitation. A possibility for that purpose is to compare peak intensities between different LC runs (label free quantification) or count the number of identified peptide-spectrum-matches per protein (spectral counting). The introduction of heavy isotopes is another widely used strategy for quantification (e.g. 'stable isotope labeling by amino acids in cell culture' (SILAC) [114] or 'isobaric tag for relative and absolute quantitation' (iTRAQ) [115]). Those strategies allow combining samples in early stages of the workflow as the differentially labeled peptides can be (only) distinguished by their mass difference.

Bottom-up approaches are also used for the identification of peptides from MALDI imaging. Identifications directly from the tissue are reported [116]. But as the mass-based isolation of peptides on the background of all other peptides is not trivial, only peptides with a high local abundance are identifiable in this way. As outlined in this chapter, a strategy to overcome this bias is the extraction of peptides followed by LC separation and tandem mass spectrometry [117]. The precursor mass is then used to relate back to the mass of the peptide imaging experiment. For the identification of m/z species from protein MALDI imaging, this approach is not applicable as the information of the precursor mass is lost due to the proteolytic digestion.

Top-down proteomics

In principle the identification, characterization and quantification of proteoforms should be one of the main goals of proteomic approaches as those *in vivo* species can be seen as molecular probes and are the direct result of processes within a biological system. Bottom-up approaches often deliver only a limited depth of information as in most cases not the complete protein sequence is covered by identified peptides. The basic idea of top-down approaches is similar to those of bottom-up (Figure 1-10). Top-down strategies also use the information derived from tandem mass spectra of isolated and fragmented precursor ions. But as there are no cleavage enzymes involved, the number of possible sequences matching to the measured mass increases dramatically. Furthermore, the sensitivity of MS instruments for larger molecules is not as good as for smaller molecules. In ESI experiments, this effect is even more dramatic as the protein will be available in different charge states, diluting the signal for precursor as well as for fragment ions. The larger the isolated protein is, the more possible fragmentation channels have to be populated with enough ions to gain a sufficient signal. Partially this can be addressed by optimizing parameters of the MS instrument for different mass ranges which involves a mass

dependent pre-fractionation. As the latter is prone to sample loss large quantities of starting material have to be available.

Only in recent years this approach got more feasible as the instrumentation has been improved and made accessible to a larger group of users. High mass accuracy and resolution provided by the novel generation of orbitrap instruments combined with highly efficient fragmentation and ion transfer abilities allows for the generation of high quality data [118]. Resolution and accuracy are in this respect especially important as the searching space for searches with no enzyme specificity dramatically increases for wide precursor and fragment tolerances. Also the software was adapted and new search engines dedicated for the field of top down proteomics have been developed (ProsightPC [119] and msalign+ [120]). But still the advances in instrumentation are the main driver increasing capabilities for identification and characterization of intact proteins and proteoforms. A recent paper [121] demonstrates the possible power of native high accuracy mass spectrometry by being able to distinguish 59 different proteoforms of 'one protein' (chicken ovalbumin). Discriminating the differentially modified proteoforms would have not been possible by a bottom-up approach. A landmark paper from Kelleher and co-workers showed the general feasibility of high throughput top-down proteomics. By mimicking two-dimensional gel electrophoresis in solution, they were able to identify over 4000 proteoforms derived from over thousand proteins [122].

As it is the intact mass of a proteoforms which is measured in MALDI imaging as well as in top-down proteomics a combination of both approaches would link the two experimental strategies help to fully unlock the potential of MALDI imaging. But as the instrumentation is only emerged in the last years a few studies already employed such an combinatorial approach [123–125].

Objective and outline of the thesis

MALDI imaging mass spectrometry has proven its considerable diagnostic and prognostic value. One of the main concerns of the field is that the molecular identity of potential markers mostly remains unknown. The goal of this thesis is therefore to address this issue by the development of strategies for the identification of proteoforms detected in MALDI imaging experiments and thereby unlock the full potential of this technique.

Following the idea that only proteins co-crystallized with the matrix molecules can be ionized and detected, a straightforward matrix layer extraction procedure was established. This allowed the isolation of the matrix proteome from MALDI imaging specimens derived from human tissues and its subsequent characterization via bottom-up proteomics (Chapter 2).

As the usage of proteases destroys the integrity of the proteoforms detected in an imaging experiment, bottom-up approaches cannot be utilized for the identification of intact proteinous species. Therefore a top-down strategy using high resolution and high accuracy mass spectrometry was chosen to unravel the identity of the proteoforms from human tissues (Chapter 3). To enhance the sensitivity employed approach an iterative exclusion list strategy was developed. This strategy profited from the subset database compiled from the bottom-up identified matrix proteome.

The established top-down approach was further applied for the identification of proteoforms derived from human tumors xenotransplanted in mice (Chapter 4). To increase the identifications of longer proteoforms the mass range was iterated by employing the so called 'mass mode' which allows mass instead of m/z based precursor selection.

Chapter 5 describes the assessment of a new mass calibration standard designated for the calibration of MALDI spectra in the typical mass range of imaging experiments. The standard consists of a 20 kDa protein expressed in *E.coli* with three tryptic cleavage sites. Mixing the protein with its tryptic peptides results in a calibration standard with equally distributed calibration points. Furthermore a calibration standard tailored for bottom-up proteomic approaches was synthesized and evaluated for its utilization as MS and MS/MS calibration standard.

Chapter 2:

Bottom up Characterization of the MALDI Matrix Proteome

Introduction

Matrix-assisted laser desorption/ionization imaging mass spectrometry (MALDI IMS) is an emerging technique that can be described as a multi-color molecular microscope as it allows visualizing the distribution of many molecules as mass to charge (m/z) signals in parallel *in situ* [126]. Originally described some 15 years ago [127] the method has been successfully adapted to different analyte classes including small molecule drugs [128], metabolites [129], lipids [130], proteins [131] and peptides [132] using e. g. formalin fixed paraffin embedded (FFPE) as well as fresh frozen tissue [133]. Because the tissue stays intact in the process, MALDI IMS is compatible with histochemistry [134] as well as immunohistochemistry and thus adds an additional dimension of molecular information to classical microscopy based tissue analysis [135]. Imaging of proteins is appealing as it conceptually allows determining the localization and abundance of proteoforms that naturally occur in the tissue under investigation, including modifications such as phosphorylation, acetylation or ubiquitination, protease mediated cleavage or truncation [52]. Therefore a proteinous m/z species detected by MALDI IMS can be viewed as an *in situ* molecular probe of a particular biological process. In turn, m/z abundance patterns that discriminate different physiological or pathological conditions might be used as diagnostic or even prognostic markers [83,136] In recent years, MALDI IMS of proteins has been successfully applied to different cancer types from the brain [137], breast [81,138], kidney [139], prostate [140] and skin [141]. Furthermore, the technique has been applied in the context of colon inflammation [142], embryonic development [143], Alzheimer's disease [144] and amyotrophic lateral sclerosis [145]. With a few notable exceptions [136,138,139,141,145–

153], the identity of the proteins constituting the observed characteristic m/z patterns has generally remained elusive. This not only precludes the validation of the putative biomarkers by for example immunohistochemistry, but also the elucidation of the biological processes that might underlie the observed phenotype.

Here a straightforward extraction and identification method for proteins embedded in the MALDI matrix layer which represent the molecular species amenable to MALDI IMS is presented. Using a bottom-up proteomic approach including tryptic digestion and liquid chromatography tandem mass spectrometry (LC-MS/MS), an inventory list of proteins derived from this layer was created which is termed MALDI matrix proteome. Although the bottom-up approach breaks the link between the identified proteins and the m/z species detected in MALDI IMS, the list of identifications serves as the pool of proteins from which all potential proteins detectable in MALDI imaging are most likely derived. The proposed extraction method is generic and can be applied to any MALDI IMS specimen to characterize the sample specific matrix proteome.

Methods

MALDI IMS Sample preparation

Cryosections (10 μm) of fresh frozen tissue (esophagus, stomach carcinoma stomach normal tissue, colon mucosa, colon muscle, colon adenoma, colon carcinoma, two different mamma carcinomas and osteosarcoma) were cut on a cryostat (CM1950, Leica Microsystems, Wetzlar, Germany), transferred to cooled ($-20\text{ }^{\circ}\text{C}$) conductive Indium-Tin-Oxide (ITO) coated glass slides (Bruker Daltonik, Bremen, Germany) and treated with 1:1 poly-lysine: 0.1% NP40 (nonyl phenoxyethylpolyethoxyethanol 40). The sections were washed in 70% ethanol and 100% ethanol for one minute each, air-dried and subsequently coated with MALDI matrix (10 g/l sinapinic acid, Sigma Aldrich, in 60% acetonitrile, 0.2% trifluoroacetic acid, TFA) using the ImagePrep spray device (Bruker Daltonik) according to Table 2-1.

Table 2-1: Detailed method for matrix coating of imaging specimens using the Bruker ImagePrep.

	Matrix thickness	Nebulization	Incubation	Dry
Phase 1	10 cycles	25% power 30% modulation 2.2 s spray	5 s	60 s
Phase 2	1 cycle	-	-	60 s
Phase 3	Matrix layer thickness 0.1 V 4- 18 cycles	25% power 30% modulation Sensor 0.1 V	30 s	20% Complete dry every 2 nd cycle Safe dry 20 s
Phase 4	Matrix layer thickness 0.3 V 8- 40 cycles	25% power 30% modulation Sensor 0.2 V	30 s	20% Complete dry every 4th cycle Safe dry 50 s
Phase 5	Matrix layer thickness 0.35 V 4- 64 cycles	25% power 35% modulation Sensor 0.3 V	30 s	20% Complete dry every 4th cycle Safe dry 60 s

Protein extraction from matrix coated cryosections

Proteins were extracted from IMS samples in three steps which are illustrated in Figure 2-1. In the first step, the matrix coated area of the cryosection (approx. surface area 5 cm²) was covered with 150 µl of 7.5% acetonitrile in 0.2% TFA, incubated for one minute and the liquid containing the protein extract was recovered using a pipette. This step was repeated using a further 300 µl of solvent and both extracts were combined. In the second step, the same area was covered with a total of 150 µl of 60 % acetonitrile 0.2% TFA and the liquid was immediately collected again. Note that the first extraction step (low organic content) allowed extracting the matrix without liquid spreading beyond the covered area. The second extraction step (high organic content) dissolved the matrix completely to recover all of the embedded protein in a separate tube. In a third step, the solvent extracted remaining tissue was removed from the slide by means of a scalpel and collected in a tube as a third sample. The two solvent extracts and the excised tissue section were dried using a vacuum concentrator.

SDS gel electrophoresis and in gel digestion

Seventy-five µl of 2x LDS sample buffer (lithium dodecyl sulfate, Invitrogen, Carlsbad, CA) containing 100 µM dithiothreitol (DTT) were added to each of the dried samples. After incubation for one hour at 90°C on a shaker, 17.5 µl of a 550 mM iodoacetamide (IAA) was added, and the sample was incubated for 30 min in the dark to alkylate cysteines. After 10 min of centrifugation at 11,340 x g, samples were heated to 90 °C for 10 minutes and 25 µl of the sample were loaded onto a denaturing gradient gel (4–12% NuPAGE, Invitrogen). The run time for acetonitrile extracts was 5 minutes (just to run the samples into the gel) and 45 minutes for the total tissue extracts (to separate the proteins) at a constant voltage of 200 V. Gel lanes derived from total tissue extracts were cut into 12 equally sized slices. Samples derived from the matrix extracts were cut as a single piece. In-gel tryptic digestion of the proteins was performed according to [154] using sequencing grade trypsin (Promega, Mannheim, Germany)

LC-MS/MS analysis for bottom-up protein identification

Nanoflow LC-MS/MS was performed by coupling a nanoLC-Ultra 1D (Eksigent, Dublin, CA, USA) to an LTQ-Orbitrap XL (Thermo Scientific, Bremen, Germany). Tryptic peptides were dissolved in 20 µl of 0.1% formic acid, and 10 µl were injected for each analysis. Peptides were delivered to a trap column (100 µm inner diameter 2 cm length, packed with 5 µm C18 resin, Reprosil PUR AQ; Dr. Maisch, Ammerbuch, Germany) at a flow rate of 5 µl/min in 100% buffer A (0.1% FA in HPLC grade water). After 10 minutes of loading and washing, peptides were transferred to an

analytical column (75 μm x 40 cm C18 Reprosil PUR AQ, 3 μm ; Dr. Maisch). Peptides from acetonitrile extracts of the tissues were separated using a 210 minute gradient from 7 to 35% of buffer B (0.1% FA in acetonitrile) at 300 nl/min flow rate. Peptides from total lysates of tissue sections were separated by a 90 minute gradient from 2 to 35% of buffer B at 300 nl/min. The LTQ-Orbitrap was operated in data-dependent mode, automatically switching between MS (in Orbitrap) and MS2 (in ion trap). Full scan MS spectra were acquired in the Orbitrap at 60,000 resolution and an AGC (automatic gain control) target value of 1e6. Internal lock mass calibration was performed using the ion signal $[(\text{Si}(\text{CH}_3)_2\text{O})_6\text{H}]^+$ at m/z 445.120025 present in ambient laboratory air. Tandem mass spectra were generated by collision induced dissociation (CID) for up to 15 peptide precursors in the linear ion trap at a target value of 5,000 and a normalized collision energy of 35. Precursors selected for fragmentation were dynamically excluded for 30s.

MS data processing and protein identification

Mascot Distiller version 2.2.1 (Matrix Science, London, UK) was used for peak picking, charge deconvolution and de-isotoping. The resulting peak list files were searched against the IPI human database (version 3.68, 87,061 sequences) using the Mascot search engine version 2.2.04 (Matrix Science). Search parameters included a precursor mass tolerance of 10 ppm and fragment tolerance of 0.6 Da. Enzyme specificity was set to trypsin with a maximal number of two missed cleavages. Carbamidomethylation of cysteine residues was set as a fixed modification and deamidation of asparagine and glutamine, oxidation of methionine and acetylation at the peptide N-terminus were set as variable modifications. Mascot search results were loaded into Scaffold (version 3.00.08, Proteome Software, Portland, OR, USA) and filters in Scaffold [155] were applied to result in false discovery rates (FDR) of 0.8% on protein level and 0.4% on peptide level.

Bioinformatic data analysis

Lists of identified proteins were analyzed for cellular compartment, biological function and canonical pathways using Ingenuity Pathway Analysis, IPA (Ingenuity Systems, www.ingenuity.com). For the survey of proteins from published human MALDI IMS studies, the reported protein accession numbers were mapped to IPI accession numbers using the ID mapping tool of Uniprot (www.uniprot.org). For papers reporting proteins from other organisms, respective sequences were first analyzed by BLAST (www.uniprot.org) and subsequently mapped to the corresponding human ortholog.

Results

Characterization of the matrix proteome

In keeping with the idea that any protein must be contained in the MALDI matrix layer in order to be detectable by the mass spectrometer, the goal was to identify these proteins systematically. Tissue samples were prepared for MALDI IMS according to standard procedures but instead of executing an imaging experiment, the protein containing matrix layer was dissolved and collected from the MALDI matrix layer using 7.5% acetonitrile (ACN) followed by 60% ACN and finally recovered the entire remaining material from the glass slide (Figure 2-1).

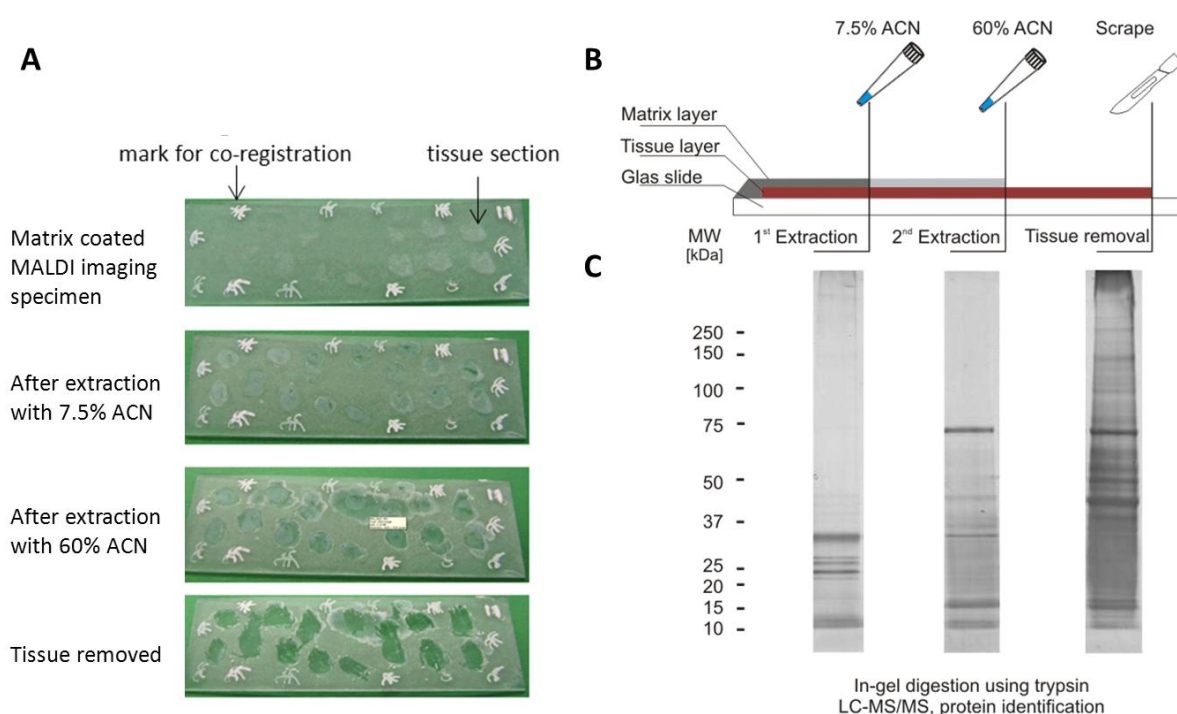


Figure 2-1 : Extraction of proteins from MALDI IMS slides.

A| MALDI imaging specimen during the different stages of the sample preparation process of a ready-to-measure MALDI imaging specimen (stomach normal tissue) during the sequential extraction of the sinapinic acid matrix layer and tissue removal. **B|** schematic representation of the two-step extraction procedure using 7.5% and 60% acetonitrile in 0.2% tri-fluoro acetic acid (TFA). The remaining tissue is recovered by means of a scalpel. **C|** Silver stained SDS PAGE of the proteins recovered from the microscopy slide showing that the acetonitrile extractions primarily recover low molecular weight proteins.

The extraction step with 7.5 % ACN was chosen as, due to the high water content of this buffer, it did not spread across the tissue and the resulting partially de-coated area also prevented spreading of the high ACN extraction buffer. The 60% ACN concentration was chosen to mimic the concentrations during the matrix application process.

Analytical SDS-PAGE separation of the three protein pools (Figure 2-1 C) shows that the 7.5% ACN extract mainly contains small proteins (<35 kDa). Increasing the concentration of the organic solvent to 60% ACN dissolves the MALDI matrix layer completely and extracts proteins up to ~70 kDa in size (the strong band is human albumin). The total tissue protein extract (Figure 2-1 C) shows the typical range of proteins expressed in human tissues covering the entire mass scale. Samples from ten different tissues were processed in this way (Figure 2-2 A) and subjected all to tryptic digestion and protein identification using LC-MS/MS (bottom-up proteomic approach).

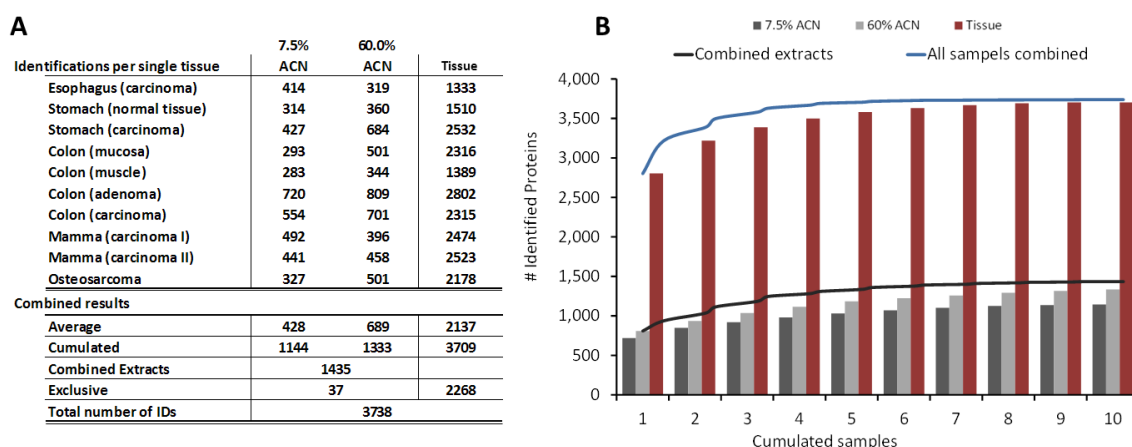


Figure 2-2: Bottom-up proteomic characterization of MALDI IMS samples.

A] Protein identification summary of ten human tissues. **B]** Cumulated number of identified proteins from the ten different tissues and three protein extractions. Samples are ordered such that the highest number of identifications is obtained for the smallest number of samples combined.

Using 20% of each extract of a single microscopic slide with approximately 5 cm² tissue surface, 400, 700 and 2,100 proteins were identified on average from the 7.5% ACN, 60% ACN and total extracts respectively (Figure 2-2 A). The total number of proteins identified from all matrix extracts was 1,435 and 3,709 from all total extracts. With a few exceptions (37 proteins), all

proteins identified from the matrix extracts were also present in the total extracts (the core proteome [156]) showing that the sample preparation used for MALDI IMS studies extracts a subset of the total proteome of the human tissue. Notably, the core tissue proteomes were rather similar between tissues as accumulating all distinct protein identifications across the tissues lead to rapid saturation (Figure 2-2 B). A similar but less pronounced effect was also observed for the acetonitrile extracts indicating that the data mainly covers ubiquitously expressed abundant proteins.

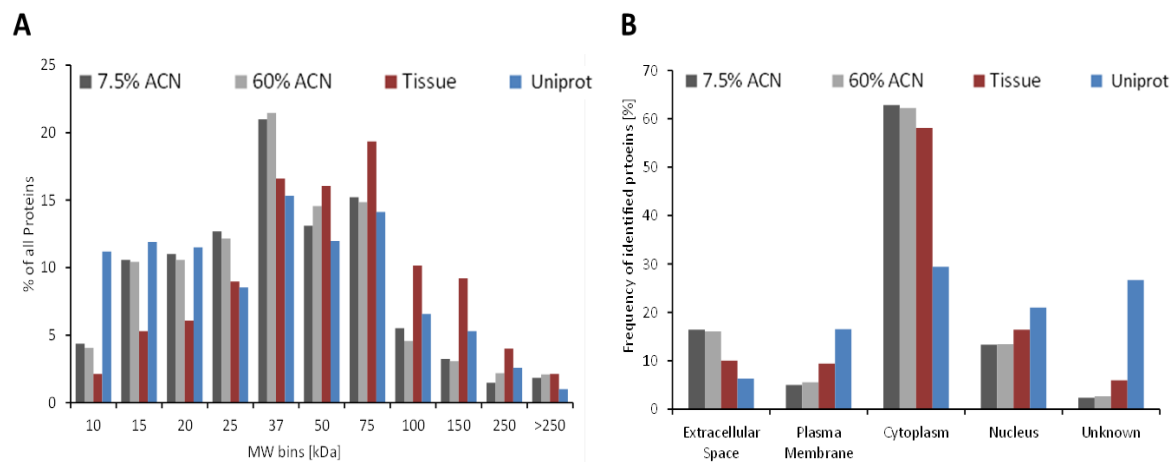


Figure 2-3: Properties of identified proteins.

A] Distribution of identified proteins by molecular weight. Bins were chosen according to the marker used in the SDS PAGE shown in Figure 2-1 B. Blue bars represent the distribution of entries in the Uniprot database (human only). **B]** relative distribution of proteins by cellular compartment.

An analysis of the molecular weight (MW) distribution of the identified proteins (Figure 2-3 A) shows that many more small proteins (<25kDa) are detected in the matrix layer compared to the total tissue, which is consistent with the data obtained by SDS-PAGE analysis and MALDI IMS data from the literature where most protein features are detected between 2-25 kDs [126] and seldom beyond 40 kDa [157]. The bottom-up protein identification data from the matrix layer did, however, also identify many nominally medium sized and large proteins (by MW in Uniprot). This is inconsistent with the SDS-PAGE data shown in Figure 2-2 C and indicates that the larger proteins are either present in quantities below the detection limit of silver staining or that these represent protein fragments arising from protease cleavage (see also below). The majority of all proteins identified from extracts are cytoplasmic proteins (Figure 2-3 B) and the

matrix extracts contain relatively more extracellular proteins but relatively fewer plasma membrane proteins than the tissue extracts. Taken together, the characterization of the matrix proteome shows that it mainly consists of small, soluble, abundant and well annotated human proteins.

A near complete list of published MALDI IMS biomarkers

As noted in the introduction, the molecular identities of m/z patterns detected in MALDI IMS studies often remain unknown. A comprehensive literature survey was conducted and resulted in a list of 254 m/z species that were identified from human and rodent tissue (Table 2-2). Thirteen of the 22 reports analyzed human tissue specimens and identified 60 distinct human proteins. These proteins were then matched to the proteins identified in the ten tissues and, if so, in which extracts and how often they were found. Interestingly, 90% of all IMS markers reported in the literature were found in the matrix proteomes despite the fact that the organs used in the literature were different from the ones used here. It is also noteworthy that the same proteins tend to be identified in multiple studies indicating they are rather abundant and not necessarily specific for a particular tissue or disease.

Many IMS studies use rodents as disease models and the list of identified human matrix proteins also covers nearly 80% of the rodent orthologs (Table 2-2). To assess protein abundance specifically, all 3,700 proteins identified were ranked in the total tissue extracts by the number of tandem mass spectra with which they were identified (divided by the size of the protein) and proteins identified from the literature were highlighted in orange (Figure 2-4 A). The expression range of proteins in the total tissue extract spans approximately four orders of magnitude and most of the IMS biomarkers appear to be abundant cellular proteins.

When plotting the 1,400 matrix proteins on the same scale (Figure 2-2 B), most of the reported IMS markers are also abundant in the matrix. However, there are also some proteins with quite high abundance in the matrix but only medium or low abundance in the tissue (e.g. cytochrome c oxidase). This may suggest that MALDI IMS, at least in principle, can access proteins across several orders of magnitude in cellular protein expression. An alternative explanation is that the spray coating process may strongly favor the solubilization of some particular proteins.

Table 2-2: Proteins previously reported in MALDI IMS studies and their frequency of observation in the present study. The maximum number of observations 'in extracts' is 20 (i.e. 10 human samples, two acetonitrile extractions each) and 10 'in tissue'. '# of references' counts how often a proteins has been reported in a publication either in human or rodents.

Gene name	UniProtKB entry	Frequency of obs.		# of references		Gene name	UniProtKB entry	Frequency of obs.		# of references	
		extracts	in tissue	rodent	human			extracts	tissue	rodent	human
ACBG1	Q96GR2	0	0	1	0	MBP	P02686	0	0	1	1
ACBP	P07108	15	9	5	2	MIF	P14174	20	10	3	2
ACTB	P60709	20	10	1	0	MT1G	P13640	8	4	1	0
AGR3	Q8TD06	6	5	0	1	MT2	P02795	8	4	1	0
ALBU	P02768	20	10	2	0	MTPN	P58546	10	10	1	1
APOA2	P02652	5	8	1	0	NDUA4	O00483	9	10	1	0
ATOX1	O00244	0	5	1	0	NEDD8	Q15843	4	9	1	0
ATP5I	P56385	14	10	1	0	NEUG	Q92686	0	0	1	0
ATP5J	P18859	3	4	1	1	PCP4	P48539	0	0	1	0
B2MG	P61769	18	9	2	1	PPIA	P62937	19	10	1	0
BASP1	P80723	8	2	1	0	PROF1	P07737	20	10	1	0
CALM	P62158	16	9	5	1	PSME1	Q06323	20	10	1	0
CH10	P61604	20	10	2	1	PTGDS	P41222	0	0	1	0
COF1	P23528	20	10	1	0	PTMA	P06454	11	10	2	0
COX1	P00395	0	2	1	0	Q32Q12	Q32Q12	17	10	1	0
COX17	Q14061	7	1	1	1	Q496I0	Q496I0	5	9	4	0
COX5A	P20674	13	10	1	0	Q499G5	Q499G5	0	0	1	0
COX5B	P10606	13	7	1	0	QCR6	P07919	10	8	1	0
COX6A2	Q02221	10	8	1	0	QCR8	O14949	5	9	1	0
COX6C	P09669	18	10	1	2	RL22	P35268	15	10	1	0
COX7C	P15954	4	8	1	1	RL38	P63173	18	10	1	0
COX8A	P10176	0	0	1	0	RLA2	P05387	12	10	1	1
CRIP1	P50238	13	6	1	2	RS21	P63220	15	10	1	1
CRIS1	P54107	0	0	1	0	RS27A	P62979	20	10	4	2
CX6B1	P14854	0	0	2	0	RS28	P62857	18	10	2	1
CX7A1	P24310	1	4	1	0	RS29	P62273	14	8	1	0
CX7A2	P14406	5	9	2	0	S100B	P04271	1	1	1	0
CYC	P99999	20	10	3	2	S10A1	P23297	1	3	1	0
CYTB	P04080	18	10	2	1	S10A4	P26447	17	10	1	0
CYTC	P01034	8	5	1	0	S10A6	P06703	19	10	1	5
DCD	P81605	12	10	1	0	S10A8	P05109	16	9	1	2
DEF1	P59665	18	10	4	0	S10A9	P06702	15	9	2	0
DLRB1	Q9NP97	13	9	1	1	S10AA	P60903	11	9	1	3
FKB1A	P62942	18	10	3	0	S10AB	P31949	20	10	3	3
FKBP2	P26885	12	9	1	0	SCG2	P13521	0	0	1	0
FXD3	Q14802	0	0	1	0	TAGL2	P37802	20	10	1	0
GPX5	O75715	0	0	1	0	TBCA	O75347	17	9	1	0
GSTM1	P09488	0	3	1	0	THIO	P10599	20	10	2	3
GSTM3	P21266	0	9	1	0	TIM8A	O60220	5	6	1	0
GSTP1	P09211	15	10	1	0	TOMM7	Q9P0U1	0	0	1	0
H2A1C	Q93077	12	9	1	0	TTHY	P02766	17	10	2	0
H2A2A	Q6F113	20	10	2	0	TYB10	P63313	9	4	2	1
H2B1B	P33778	18	9	1	1	TYB4	P62328	10	5	8	1
H2B1H	Q93079	20	10	1	0	UBA52	P62987	0	0	1	0
H31	P68431	20	10	3	0	USMG5	Q96IX5	7	10	1	0
H4	P62805	20	10	5	2						
HAKAI	Q75N03	0	0	1	0						
HBA	P69905	20	10	3	3						
HBB	P68871	20	10	2	3						
HINT1	P49773	18	10	1	0						
HOPX	Q9BPY8	0	0	1	0						
ISK1	P00995	5	2	1	0						
LEG1	P09382	18	10	1	1						
LSM6	P62312	9	9	1	0						
LYSC	P61626	16	9	2	0						
M3K2	Q9Y2U5	0	0	1	0						

	# proteins identified		
	in extracts	in tissues	
in human	60	54 (90%)	55 (92%)
in rodents	72	56 (78%)	59 (82%)

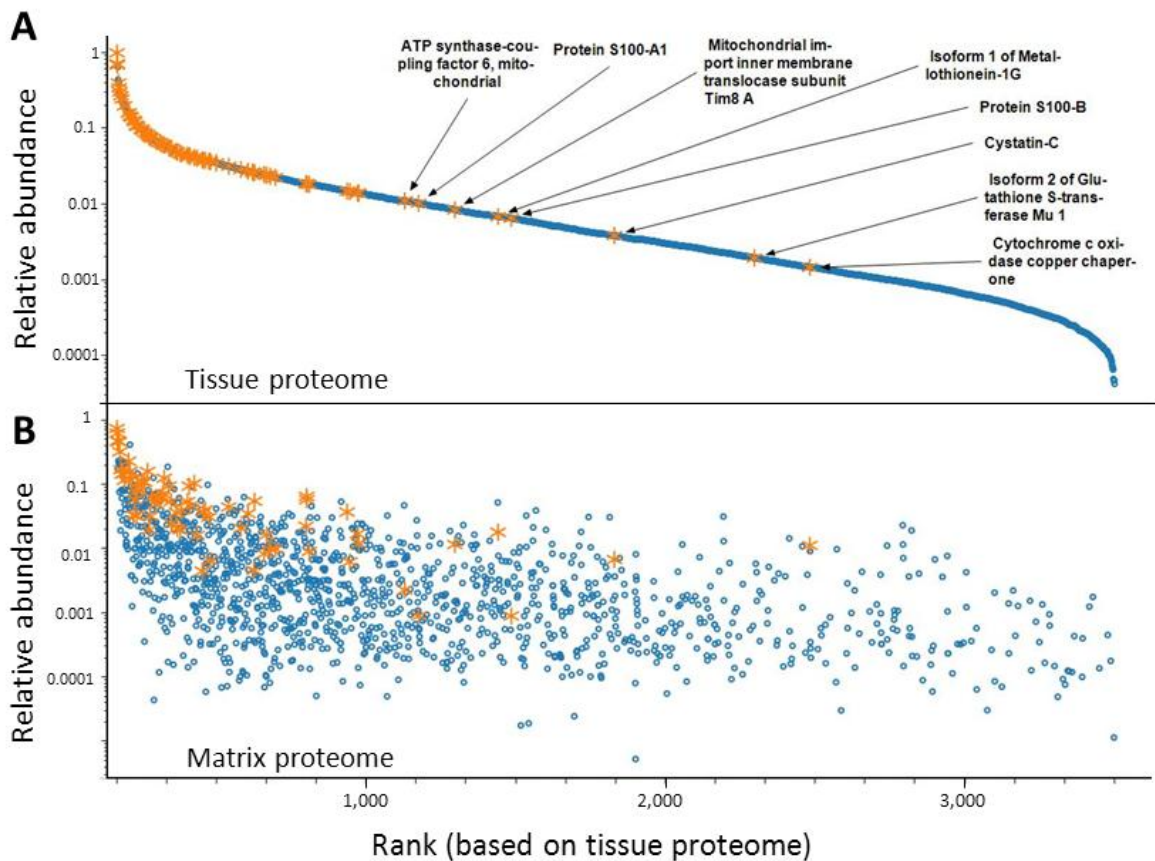


Figure 2-4: Relative abundance of proteins identified in MALDI IMS.

A] Proteins from all 10 human tissues were ranked according to the number of assigned spectra divided by the molecular weight of the proteins' database entry. MALDI IMS markers extracted from the literature (see also Figure 2-2) are indicated in orange showing that these are mainly abundant cellular proteins. **B]** Proteins identified from the MALDI matrix layer plotted against the ranked protein abundance used in panel A.

Discussion

The application of MALDI IMS of proteins has grown steadily over the past 15 years. At the same time, there still is a striking discrepancy between the sophistication with which such images can be generated and the ability to identify the underlying proteins. Out of the many hundreds of publications reporting MALDI IMS data and biomarkers only about 22 (13 human, 9 rodent), at the time of completing this study and mostly very recent papers, have also revealed the name of the protein. There are many reasons for this general inability to identify a protein using MALDI IMS. First, a mass measurement of a protein or fragment thereof as provided by MALDI IMS is not a good indicator for protein identity because any such mass can be generated from nearly any protein sequence within the mass accuracy of the experiment (1-2 Da). Second, the direct mass spectrometric sequencing of the detected m/z feature in a MALDI mass spectrum is generally not possible owing to the generally inadequate sensitivity, resolution and mass accuracy and mass range of these instruments in tandem MS mode. Many researchers have therefore decoupled the imaging experiment from the protein identification experiment and resorted to detailed bottom-up proteomic analysis of a separate tissue sample from which the image was created [150]. In a more sophisticated approach, on-tissue tryptic digestion followed by MS analysis has been employed for protein identification [132]. While these have sometimes been successful, they are also time consuming and, in fact, there has often been a fair amount of 'educated guessing' in trying to connect molecular weight information from the MALDI IMS experiment to a list of proteins identified from a tissue. The above shortcomings have plagued the field to an extent that there is growing concern, at least in parts of the community, about the general value of m/z feature information for biomarker discovery in the absence of protein identification information. This is because the identity of any biomarker must be known in order to validate and develop it into a useful clinical tool.

In this chapter an alternative way to approach this issue was taken by reasoning that any protein detectable by MALDI IMS must not only be present in the underlying tissue, but also reside in the matrix layer of the tissue preparation. Therefore a simple extraction procedure that recovers the proteins from the matrix was developed. Proteins were digested and with the protease trypsin and the 'matrix proteome' was identified by high performance LC-MS/MS. Exemplified on ten human tissues, this enabled the generation of a shortlist of ~1,400 mostly small, abundant and soluble proteins that contain nearly all MALDI IMS protein biomarkers

reported in the literature thus far. Proteins (e.g., E2F3, S100B) that were found only rarely (or never) may have been difficult to distinguish from closely related proteins, possibly misidentified in the original publication, or of too low abundance in the tissues analyzed here (e.g., myelin basic protein, MBP, P02686, which is essentially only expressed in central nervous system tissue, www.proteinatlas.org/ENSG00000197971). Many of the reported biomarkers were identified in many if not most of the tissues analyzed here. For example, the pro-inflammatory cytokine MIF (macrophage migration inhibitory factor, P14174) was found in all ten tissues as well as each of the acetonitrile extracts and this is also true for many other cases. Ingenuity pathway analysis (IPA) of protein identified here and in the MALDI IMS literature (both human and rodents) show significant overrepresentation of, for example, 'mitochondrial dysfunction' and 'EIF2 signaling' both of which are intimately linked to cancer (Table 2-3 to 5). In light of the above and because IMS biomarkers are generally abundant proteins, this observations strongly suggest that IMS biomarkers are mostly molecular surrogate markers of quite broad specificity rather than proteins involved in the molecular mechanism underlying a specific disease or physiology.

Compared to direct MALDI MS/MS or on-tissue digestion, the bottom-up LC-MS/MS approach has the advantage of superior sensitivity and analytical depth (i.e., the number of proteins that can be identified). Still, it is important to connect protein identification back to the m/z feature measured in the IMS experiment.

Table 2-3 Proteins identified in MALDI IMS studies from **human tissues**

Ingenuity Canonical Pathways	-log(p-value)	Ratio	# of molecules
Mitochondrial Dysfunction	1.11E01	8.03E-02	11
EIF2 Signaling	3.12E00	2.75E-02	5
Nur77 Signaling in T Lymphocytes	2.9E00	5.26E-02	3
Role of IL-17A in Psoriasis	2.89E00	1.54E-01	2

Table 2-4 Proteins identified in MALDI IMS studies from **rodent tissues**

Ingenuity Canonical Pathways	-log(p-value)	Ratio	# of molecules
Mitochondrial Dysfunction	9.39E00	7.3E-02	10
EIF2 Signaling	4.94E00	3.85E-02	7
mTOR Signaling	2.88E00	2.63E-02	5
LXR/RXR Activation	2.68E00	3.17E-02	4

Table 2-5 Matrix Extracts of the **ten human tissues**

Ingenuity Canonical Pathways	-log(p-value)	Ratio	# of molecules
EIF2 Signaling	3.43E01	4.18E-01	76
Mitochondrial Dysfunction	2.02E01	3.8E-01	52
Regulation of Actin-based Motility by Rho	1.06E01	3.37E-01	29
Actin Cytoskeleton Signaling	1.03E01	2.17E-01	50

Overall the proposed method is a significant contribution to the field of protein MALDI imaging as it enables the targeted isolation of all proteins co-crystallized in the MALDI matrix layer. Furthermore the strategy could be seamlessly integrated in the imaging workflow, as the matrix has to be removed after the imaging in experiment in order to prepare the tissue sections for digital microscopy. The bottom-up identification drastically reduces the list of candidate proteins possibly detectable in MALDI imaging experiments. It is therefore highly valuable for result interpretation and fundamental for hypothesis generation and validation.

Chapter 3:

Proteoform Identification from MALDI Imaging Specimens using High Resolution Mass Spectrometry

Introduction

MALDI imaging allows the visualization of the spatial distribution of proteoforms directly on the tissue. This, for example, allows to differentiate different tumor types solely based on their m/z patterns [85] or the investigation of the heterogeneity [68,124] of tumors on a molecular level. The latter is a promising application as there is no straightforward approach which would allow identifying the different sub-populations of a tumor. In the context of MALDI imaging studies aiming to reveal biomarkers however, the identification of proteinous m/z species is one of the most important tasks as only known target molecules can be validated by an independent technique like immunohistochemistry. Moreover, the elucidation of the complete sequence of a protein including all its (posttranslational) modifications combined with the spatial information on the background of the tissues' morphology offers a unique starting point for the investigation of the biological mechanisms leading to this proteoform [52]. Bottom-up proteomic approaches are very successful in identifying hundreds of proteins in the course of hours [113]. The success of these approaches is linked to the usage of a protease cleaving proteoforms into peptides. However, owing to the proteolytic digestion the link between the m/z value of the intact species is lost. The goal is to identify the intact species by a top-down approach [158] and thereby exploiting the complete potential of MALDI IMS. The most appealing way to achieve this goal would be to conduct the imaging experiment and the identification of the single species on the same instrument in one experiment. Many MALDI

instruments offer the possibility to isolate and fragment single species, but there are only a few examples in the literature [146] where this approach was successfully used. The reasons for this are several fold. Imaging studies targeting proteins are carried out with the 'cold' matrix sinapinic acid which allows a less fragmentation prone desorption and ionization of large molecules [159]. For the generation of fragment spectra, often 'hot' matrices are chosen as the transfer of vibrational energy is higher, leading -compared to 'cold' matrices- to a more efficient fragmentation. Furthermore, imaging experiments are often performed in the linear TOF mode which offers high sensitivity but lacks mass resolution and only delivers average masses at low resolution. Moreover some MALDI instruments have an upper mass limitation (5 kDa for the UltrafleXtreme) for which they allow performing MS/MS experiments. For higher masses, a strategy called in-source decay (ISD) [160] can be applied. Due to the fact that this technique can only deal with single proteins, it was mainly used in combination with excessive separation on protein level. Only recently it was shown that identifying proteins directly on tissues is possible by ISD if a protein is the dominating species in a certain tissue region [161].

Most protein identification strategies in the MALDI imaging field employ whole tissue lysis and a protein extraction, often followed by a molecular weight cut-off filtration or other weight based separation strategies along with separation and fractionation [86,162,163]. These fractions are screened again by MALDI for the m/z peaks of interest and ideally contain only a single species which allows employing all strategies for sequence elucidation e.g. differential proteolysis (mainly in early publications) [164], followed by mass spectrometric analysis or direct fragmentation. By using such liquid based approaches the spatial information is lost, but the mass still links the imaging and identification experiments. In recent years the available MS instrumentation has significantly improved and now offers fast scanning high resolution, high mass accuracy and highly sensitive instruments [118,165]. This should facilitate the successful implementation of top-down approaches as confident identification is driven by these parameters.

Methods

Sample preparation

MALDI imaging specimens and protein containing tissue extracts thereof were prepared as described for the bottom-up approach in the previous chapter. The 7.5% and 60% acetonitrile extracts were combined and dried in a vacuum centrifuge. The dried samples were reconstituted in 50 % acetonitrile, 0.2% TFA and passed through STAGE tips [166] packed with C18 Empore™ extraction disks (3M, St. Paul, US). The flow-through fraction containing the proteins was dried and reconstituted in buffer A (0.1% formic acid).

LC-MS/MS analysis and proteoform identification from esophagus and colon samples

Nano-LC-MS/MS measurements of intact proteoforms extracted from a MALDI imaging specimen of oesophagus and of colon were carried out using a nanoLC-Ultra 1D+ (Eksigent, Dublin, CA, USA) coupled to a LTQ-Orbitrap Elite mass spectrometer (Thermo Scientific, Bremen, Germany). Samples were loaded onto the same columns set-up, using the same solvents as in Chapter 2. Gradient separation was performed as follows: 2-10% B in 1 min; 10-35% B in 34 min; 35-90% B in 2 min; 90% B for 4 min; 90-2% B in 0.5 min; 2% B for 3.5 min. The Orbitrap Elite was operated in data dependent mode automatically switching between full MS and MS/MS mode. Full MS spectra were acquired in the Orbitrap over a mass range from 430 to 1800 m/z summing up four microscans. The three most intense precursor ions with charge states greater than three were selected for fragmentation using higher-energy collisional dissociation (HCD) and a normalized collision energy of 30. Once sequenced, precursors were dynamically excluded from selection for 10 seconds. Eight, respectively four microscans were acquired for each precursor in separate LC-MS/MS runs to generate tandem mass spectra with sufficient quality for proteoform identification. Both MS and MS/MS were acquired at a resolution of 60,000 in the Orbitrap using a target value of $1e6$. Raw mass spectra were processed using Mascot Distiller version 2.4.2 (Matrix Science). The resulting charge-deconvoluted and de-isotoped spectra were searched against the IPI human database (version 3.68, 87,061 sequences) using the Mascot search engine version 2.3 (Matrix Science). The same data was also searched against a subset database (fasta file generated using the export function of Scaffold) containing 5,465 sequences assembled from the bottom-up analysis described in the previous chapter (3,738 protein sequences and 1,727 isoform sequences). All searches were carried out using 7 ppm precursor tolerance and 0.04 Da fragment ion tolerance. Acetylation of

lysine, protein and peptide N-terminus, deamidation of glutamine and asparagine as well as oxidation of methionine were set as variable modifications. No enzyme was specified for searching. In order to sequence less abundant precursors via the data dependent acquisition the search results of the first measurement were used to compile an exclusion list for further LC-MS/MS iterations. The peak list files of all runs were merged into one file using Mascot daemon. Searches were run without the decoy option i) because searching against human IPI would have taken excessively long (an estimated 8 days on a 12 CPU server) and ii) because the target-decoy approach is not well suited to estimate FDRs in low complexity samples such as the ones analyzed here. Search results were loaded into Scaffold version 3.6.3. For protein identifications to be accepted, the Mascot ion score had to be equal or above the Mascot homology score of 52. In addition, proteins identified by a single peptide were verified by manual inspection of the tandem mass spectra. Identifications based on tryptic peptides originating from carry over from previous runs were excluded.

Data analysis

Protein sequence coverage was calculated from the 'peptides start and stop index' column in the scaffold spectrum export function and normalized to the protein length using MS Excel. The heatmap visualization was carried out using R (performed by Mathias Wilhelm, Chair of Proteomics and Bioanalytics, TUM) and clustered for N-/C- terminal coverage. Fragment spectrum annotation was performed using the web based expert system [167] and the mgf file generated by Mascot distiller.

Results

Proteoform identification using high resolution MS and a trimmed database

In the previous chapter it was shown that the introduced extraction method is well suited for the isolation of the matrix proteome. Moreover, it minimizes the needed sample amount which is especially critical if human tissues are investigated. The modified STAGE tip protocol has proven to be fast and effective to remove particles that were eventually co-extracted from the MALDI imaging specimen and interfered with LC-MS/MS measurements. A forty-five minute gradient originally developed for the separation of a tryptic digest with modest complexity was chosen as a starting point and optimized for longer and therefore inherently more hydrophobic proteoforms. A steep increase of buffer B at the beginning of a linear gradient ending 5 % higher than typically used for tryptic peptides allowed for a time efficient separation. The acquisition of MS and MS/MS spectra at a resolution of 60.000 and at least 4 microscans enabled charge state recognition also of highly charged precursors as well fragment ions thereof (Figure 3-1 A and B). This is important as it allows calculating actual precursor and fragment masses (deconvolution) and therefore facilitates the identification of longer sequences (Figure 3-1 C).

But even though spectrum quality was sufficient, most of the selected precursors were in the low mass range and only a few species above 3 kDa were identified. The data dependent acquisition selects (based on the fixed parameters m/z range and charge state) the most intense ions of a full scan for fragmentation. This decision algorithm therefore discriminates against the sampling of heavier species as the available molecules are likely to be distributed over several charge states and into more isotopic peaks [168]. The dynamic exclusion method was only partially capable to address this problem and low intensity as well as highly charged peaks were only occasionally selected for fragmentation. To overcome this bias a list of precursor masses of proteoforms identified in the first LC-MS iterations was generated and used as a static exclusion list in the instrument method. This resulted in a more frequent sampling of masses above 4 kDa (Figure 3-2 B).

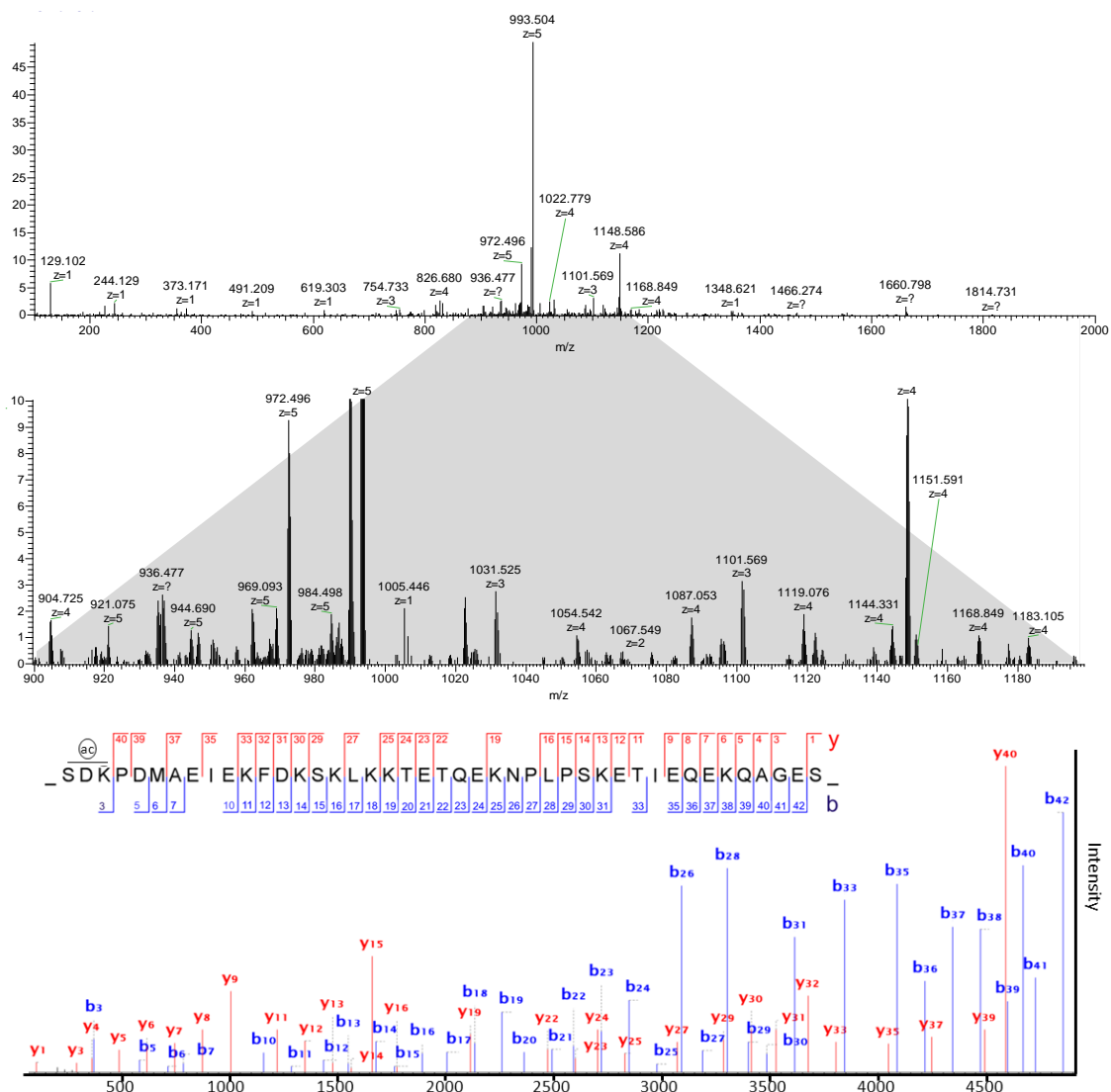


Figure 3-1: High resolution MS/MS spectra enable charge recognition

A| HCD spectrum of precursor 993.51 m/z charge 5+ **B|** zoom in of A from 900 to 1200 showing fragment ions and charges up to +5 **C|** annotated, charge deconvoluted and de-isotoped tandem mass spectrum of intact and acetylated thymosin beta-4 (score 266), internal fragments were omitted.

The exclusion list approach was facilitated by the utilization of a smaller subset database which enabled a reduction of searching time and therefore allowed to iterate the sample in shorter intervals. This subset database was created from sequences of proteins identified by the bottom-up strategy in the previous chapter and only contained 5455 entries. The application of this strategy successfully increased the number of unique identifications in the higher mass

range. In total 126 distinct proteins and 530 proteoforms were identified from 768 unique spectra when searching against the subset database (Figure 3-2 A). To review whether using the smaller database leads to a loss in identifications the search was repeated against the complete IPI human database [169]. Containing 15x more sequence entries dramatically increases the searching space and resulted in an about three times longer searching time. 124 proteins were identified in common and only seven proteins were additionally identified by searching the larger database.

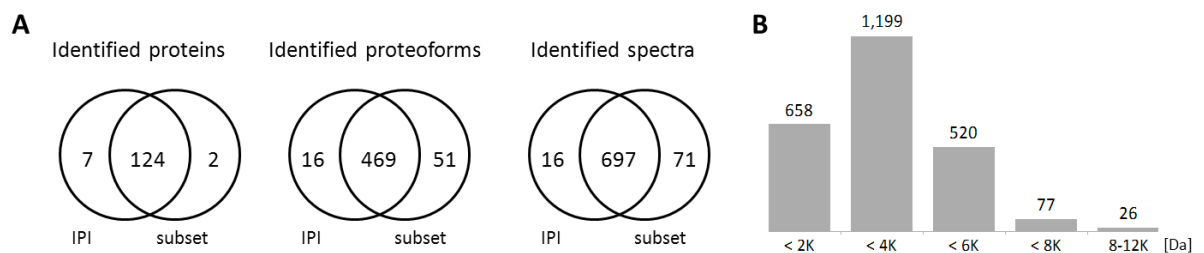


Figure 3-2: Identification of proteoforms using different databases

A Comparison of identified proteins, proteoforms and unique spectra generated by searching the human IPI database (87,061 entries) or the subsetDB (5,455 entries) showing that almost all proteins can be identified in the 15x smaller subsetDB. **B** Mass distributions of precursor ion in bins of 2kDa leading to a successful top-down identification

At peptide and spectrum level only 16 species were exclusively identified by searching the IPI database, whereas the number of peptides and spectra identified by searching the subset databases is more than 3 (peptides) to 4 times (unique spectra) higher. This is mainly a result of the lower Mascot identity score threshold resulting from the smaller search space. Comparing the number of unique spectra (768) to the number of identified spectra (~2,500) shows, that still many species have been repeatedly selected.

The vast majority of the identified peptides are smaller than 4 kDa. However it is noteworthy that peptides as large as 12 kDa could also be identified (Figure 3-2 B) using sample quantities recovered of a single microscopic slide. Interestingly, many of the identified proteins are represented as C-Terminal and N-Terminal fragments (Figure 3-3) suggesting significant protein processing within the tissue that gives rise to these species.

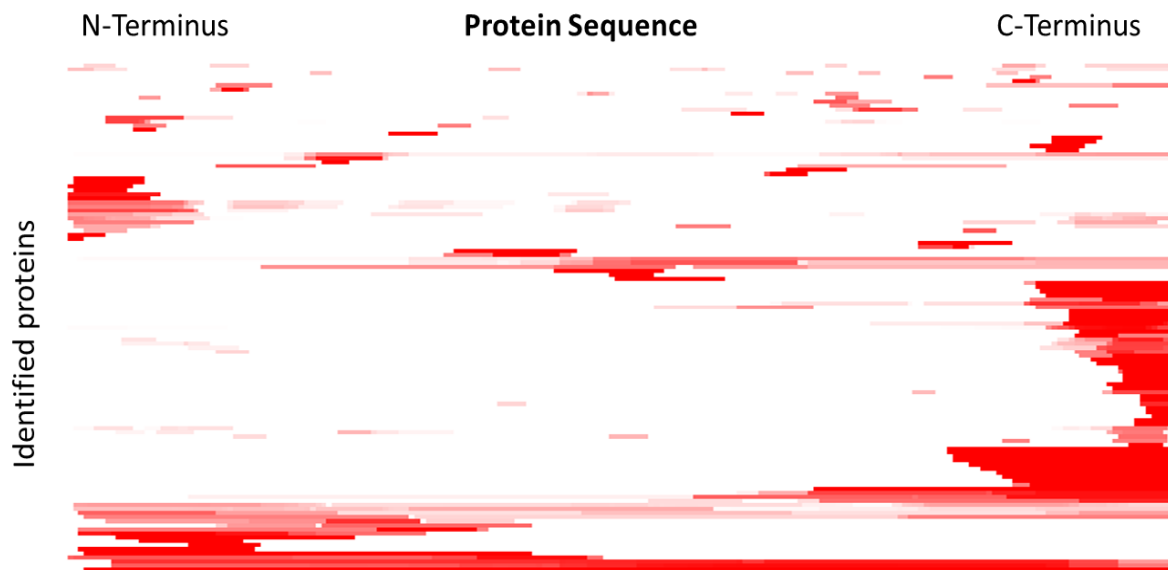


Figure 3-3 Normalized sequence coverage of identified proteins.

Each line represents one protein. The darker the red scale density, the more spectra matched to the respective region of the protein indicating that most proteins identified from MALDI IMS samples are fragments of the original protein sequence.

Discussion

Identification of proteoforms from MALDI IMS experiments provides the link between the images and furthermore allows the investigation of the underlying mechanism leading to these species. But this task is still not addressed very well. Direct sequencing on the tissue is for many reasons a very demanding task and seldom successful. By off-tissue approaches the localization information is lost, but separations based on other techniques allow relating back to the imaging experiment by using the mass as link. Classical tissue disruption and homogenization approaches allow to collect and isolate the complete proteome of the investigated tissue section, but as all proteins are included, excessive separation has to be carried out. The extraction procedure described in Chapter 2 circumvents much of the work intensive and loss prone steps and directly targets the proteins of interest.

High mass accuracy and high resolution instruments facilitate the identification also in higher mass ranges as the spectra allow to deduce charge state and enable deconvolution of MS and MS/MS spectra [170]. But still, identification becomes more challenging the heavier the proteoforms are: The longer the sequence the more possible fragmentation channels have to be populated which means that more ions have to be collected for fragmentation. LTQ-Orbitrap instruments allow varying the number of collected ions for fragmentation only in a static manner via the (predicted [171]) automatic gain control, which is only able to 'count' charges. This means that the higher the charge state of the collected ion is, the fewer molecules are collected for the same AGC value. Therefore, the target value is always a compromise between scan time and spectrum quality [165]. Furthermore, longer proteoforms tend to form ions of different charges leading to an effect called charge state dilution which led to selection and fragmentation of the same 'target' in different charge states during the data dependent acquisition [169]. Additionally the intensity of a single m/z species is low as it is distributed into more and more isotopes. Taken together this leads to a mass dependent decrease of identifications. However, the identification of the most prominent peaks from MALDI imaging experiments is possible using this strategy.

Even with very high quality mass spectrometric information, top-down database searches are time consuming as the precursor mass could have been originated from nearly any protein sequence if no enzyme specificity limits the number of possibilities [107]. This is also reflected in the Mascot identity threshold scores which is a negative logarithmic measure of the probability of a hit being random. For typical bottom-up approaches this threshold is in the range of 25 to 30, whereas in the searches with no enzyme specificity it typically has a value around 50. Clearly the number of possible precursor masses is not only a function of mass accuracy and therefore search tolerance, but also of database size [107]. The effect of a 15x smaller database was shown for the identification of proteins based on searches against the subsetDB compared to the IPI database [169], which led to decreased score thresholds and therefore an increased number of identifications.

An interesting observation from literature survey and the top-down analysis is that the majority of the IMS biomarkers reported in the literature were detected as full length, unmodified proteins while the top-down analysis clearly showed that many of the m/z species in the matrix layer are fragments of much larger proteins (Figure 3-3) or are post-translationally modified (Figure 3-2). This has been generally overlooked in the past because these fragments could

technically not be readily identified. It further suggests that many more IMS biomarkers could now be identified without the common 'educated molecular weight guessing' based on the simple (but often invalid) assumption that a protein would be detected as an intact, unmodified molecule. Recent evidence suggests, that the fragmented proteins are not the result of sample preparation and degradation artifacts but are genuine and discriminating surrogates of biological processes occurring in the respective tissue [149,163,172].

Chapter 4:

Proteoform Identification from Human Tumor Xenografts in Mice

Introduction

Xenotransplantation offers the possibility to grow various solid cancer types hosted in model organisms. This allows for the investigating different clinical conditions such as the response of the tumor to drug treatment [173]. A project of Katharina Huber (Institute of Pathology, Helmholtz Center Munich) focused on the response of a head and neck squamous cell carcinoma to the treatment with different kinase inhibitors. Besides monitoring the drug distribution, changes in proteomic patterns were investigated by MALDI imaging. The established workflow was employed to identify proteoforms from xenografts of different treatments and time points.

Methods

LC-MS/MS analysis

The MALDI imaging specimens were prepared from 6 differentially treated xenografts hosted in mice. The matrix proteome extract for LC-MS/MS analysis was prepared in the same way as in the previous chapter. Gradient separation was performed on a nanoLC-Ultra 1D+ (Eksigent) coupled to a LTQ-Orbitrap Elite mass spectrometer (Thermo Scientific) as follows: 2-15% B in 1 min; 15-45% B in 34 min; 45-90% B in 2 min; 90% B for 4 min; 90-2% B in 0.5 min; 2% B for 3.5 min at a flow rate of 300 nl/min. Buffer A was prepared from water and buffer B from acetonitrile. Both buffers were supplemented with 0.1% FA and 5% DMSO (v/v). Each sample was measured three times with modified acquisition methods on a LTQ-Orbitrap Elite mass

spectrometer operated by the Xcalibur Software (Version 2.2). In the first run, full MS spectra were acquired from 390 to 1800 m/z at a resolution of 60,000 and a maximal scan time of 100 ms collecting $1e6$ charges. Dynamic recalibration was performed using a background ion species at m/z 401.922718 with the elemental composition $[C_6H_{10}O_{14}S_3]^+$ generated from DMSO containing buffers [174]. The three most intense precursor ions with a charge state greater 3 and a m/z value greater than 450 (resulting in a minimum m/z of 1797) were selected for HCD fragmentation collecting $5e5$ charges within a maximum scan time of 500 ms. HCD spectra were acquired at a resolution of 60,000 using 2 microscans. For the second and third iteration a target value of $1e6$ for MS/MS was chosen and data dependent precursor selection was carried out in mass mode picking only species with a mass greater than 5000 Da. Maximum scan time was set to 1 second. Precursors selected once were dynamically excluded for 30 seconds.

Database Searching

Raw mass spectra were processed as stated in Chapter 3. Peak list files of the three runs of one sample were merged into one file using Mascot distiller. The resulting charge-deconvoluted and de-isotoped spectra were searched against a Uniprot human database containing only canonical isoforms (UniProtKB, 20120905 (20,225 sequences; 11,291,209 residues) and a Uniprot mouse database containing only manually reviewed protein entries (UniProtKB 20130617, 16,613 sequences; 9,294,716 residues). All searches used 7 ppm precursor tolerance and 0.07 Da fragment ion tolerance. Methylation, di-methylation and acetylation of lysine, as well as acetylation of protein N-termini were selected as variable modifications. No enzyme was specified for searching. Search results were loaded into Scaffold version 3.6.3. For protein identifications to be accepted, the Mascot ion score had to be equal or above the query dependent Mascot identity score.

Data analysis

Protein identification data was exported from Scaffold using the spectrum export function. Microsoft excel 2010 was used for data handling. Graphics were generated using Tableau 8.0 (Tableau Software). Xcalibur query language (XQL, Thermo Scientific) was used for extracting data from RAW-files. Precursor m/z and charge for searched spectra were extracted from mgf files. Based on the charge state and m/z values, precursor ions selected for fragmentation during LC-ESI-MS experiments were used to calculate the respective average $[M+H]^+$. Average total ion current (TIC) were calculated for bins of the size of 1 Dalton using Tableau.

Reconstructed mass spectra of the selected precursor ions were plotted with MS excel where peak shapes were approximated by the formula:

$$f(x_n) = 0.2x_{n-2} + 0.4x_{n-1} + x_n + 0.4x_{n+1} + 0.2x_{n+2}$$

where x is the average TIC of bin n . In the same way reconstructed spectra of successfully identified precursor ions were generated using the m/z values from Scaffold spectrum export. Those reconstructed spectra were overlaid with MALDI spectra exported from Flex Analysis (Bruker).

Results

Proteoform identification and discrimination of organisms

In total seven samples from MALDI imaging specimens were prepared and three LC-MS experiments were performed with varied parameters. In the first run the minimal required mass to be selected for fragmentation was set to 1796 Da. For the second and third run the 'mass mode' in the instruments software was enabled. In this mode the data dependent precursor selection is based on the deconvoluted mass instead of m/z values and was adjusted to fragment only precursors with a minimum mass of 5 kDa. The effect of the modified acquisition method is reflected in the increased number of selected precursors (Figure 4-1 A, blue bars) in and above the 5 kDa bin. The number of scans however does not translate into a higher number of identifications in this mass range (green bars). A simple explanation would be that the number of available precursors was just too low for sufficient MS/MS spectra even, if the maximum fill time used (Figure 4-1 B, red area). The decrease of the identification rate for higher masses could then be addressed to charge state dilution and isotope dilution effects as discussed in chapter 3. However, the drop of the MS/MS TIC intensities for masses above 5 kDa not reaching the maximum injection time (Figure 4-1 B, green area) can not only be explained in that way.

Linked the low intensities and the long fill times, another effect called space charging [175] might be involved in the identification rate decrease for highly charged ions of low abundance. As trap instruments first collect all available ions species and then selectively eject the ones above and below the targeted m/z range the long accumulation times may have led to an

overfilling of the trap. The resulting high charge density building up inside the trap during the ion accumulation interacts with the imposed quadrupolar field and eventually comprises the performance and therefore also the selection efficiency of the trap.

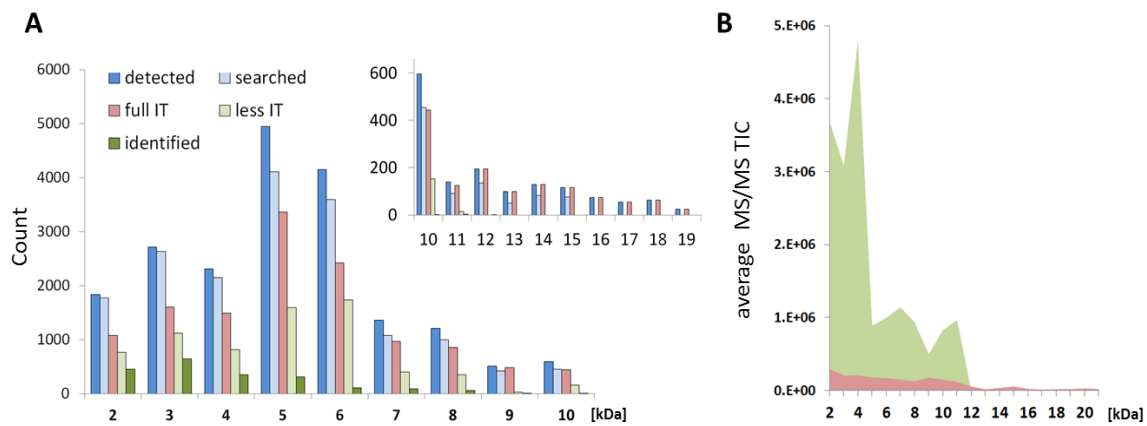


Figure 4-1: Mass dependence of scan events and identifications.

A] Count of spectra that were acquired, searched, reached maximum injection time (full IT) of 500ms or 1000ms respectively, not reached max. injection time (less IT) and those that were successfully identified **B]** Average TIC per bin of spectra reaching (red) or not reaching (green) maximum injection time.

Still the complete measurement resulted in 2014 spectra with a Mascot ion score equal or above the Mascot identity score and 550 proteoforms derived from 306 proteins were identified (Figure 4-2 A).

Interestingly, for over 75% of the identified precursors it was, due to the long sequences, possible to discriminate whether the proteoform was derived from the human tumor tissue or from the tumor hosting organism. The analysis on the level of spectral counts of the subcellular localization of the 3 groups (human, mouse, not distinguishable) reveals a relative overrepresentation of proteins from mouse in the category 'extracellular space' (Figure 4-2 B). Most of these spectra are derived from fragments of serum albumin. Furthermore many of the spectra in the category 'cytoplasm' are derived from mouse hemoglobin alpha and beta. Both have a high abundance in red blood cells [176]. A relative underrepresentation of proteins from mouse is shown for the nucleolus. These results are in good accordance with the biological situation as the human tumor tissue is integrated into tissue and supplied by the blood cycle of the host organism. Furthermore the distribution is very similar to the results from the bottom-up approach in chapter 2 (Figure 2-3). An example for a proteoform where no clear

discrimination was possible and which was identified from 86 of 377 spectra identified in this category is the ubiquitin. It is a highly conserved prokaryotic protein [177] and has the same sequence in mouse and human.

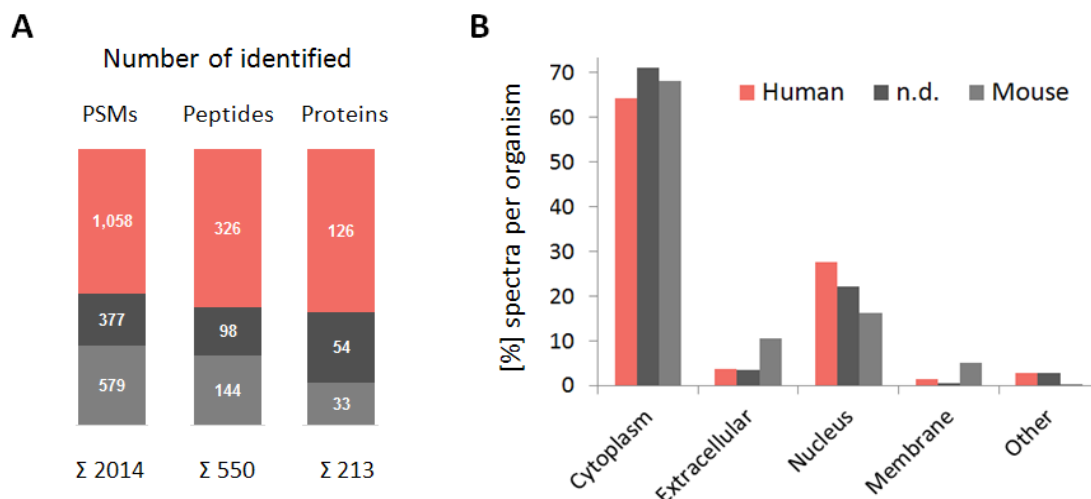


Figure 4-2: Organism specific proteoform identification.

A] Number of identified peptide spectrum matches, identified peptide sequences and proteins.

B] Relative distribution of identified spectral matches sorted according to the subcellular localization of the protein from which it was derived.

Linking identifications and MALDI spectra

The final goal of this approach is to link the identified proteoforms to the peaks from MALDI imaging. As MALDI mainly generates singly protonated species the direct way would be to calculate the average m/z value of the $[M+H]^+$ of the identified proteoforms and match it to the respective peak of the MALDI spectrum by a simple lookup. Another method which allows a more global visual comparison between the MALDI peaks and the species selected for fragmentation is depicted in Figure 4-3. Here the MALDI spectrum (Figure 4-3 A) is aligned on a common mass axis with a spectrum reconstructed from the deconvoluted $[M+H]^+$ values of the precursors ions selected for fragmentation versus their average MS/MS TIC (Figure 4-3 B, red) during the LC-MS/MS analysis. The resemblance of the MALDI and the reconstructed spectrum is good, especially for high intensity peaks in the lower mass range. In the higher mass range intensities decrease in the reconstructed mass spectrum, but also above 10 kDa counterparts in both spectra can be recognized. Successful identifications are plotted accordingly (in blue) enabling a simple annotation of identified species in the MALDI spectrum.

Figure 4-3: Resemblance of MALDI-TOF MS and LC-ESI-MS measurements (next page)

A| MALDI spectrum of a matrix extract spotted in sinapinic acid. **B|** Reconstructed spectrum from selected precursors (as $[M+H]^+$ in bins of 1 Da) plotted against the average TIC intensity of the MS/MS spectra per bin from 3 LC-MS runs in red and the TIC intensities (x10) of the spectra that led to an identification in blue. **C|** Overlay of a MALDI spectrum and the reconstructed spectrum. Peak in MALDI is assembled from different species of which two were identified as proteoforms of ubiquitin (RL40). The peak at 8560 has a Mascot score of above 124. The second peak at 8569 was identified with a score of 78 and a Mascot delta score of 0. The original spectrum does not allow annotating the site of the suggested modifications (twice acetylated at K, twice di-methylated K). Acronyms (blue) indicate the protein name from which the identified proteoform is derived (see below).

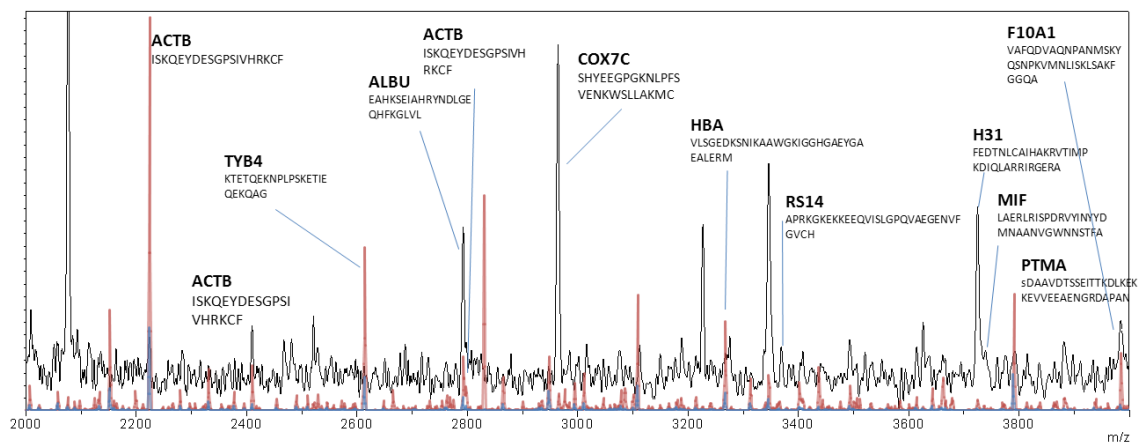
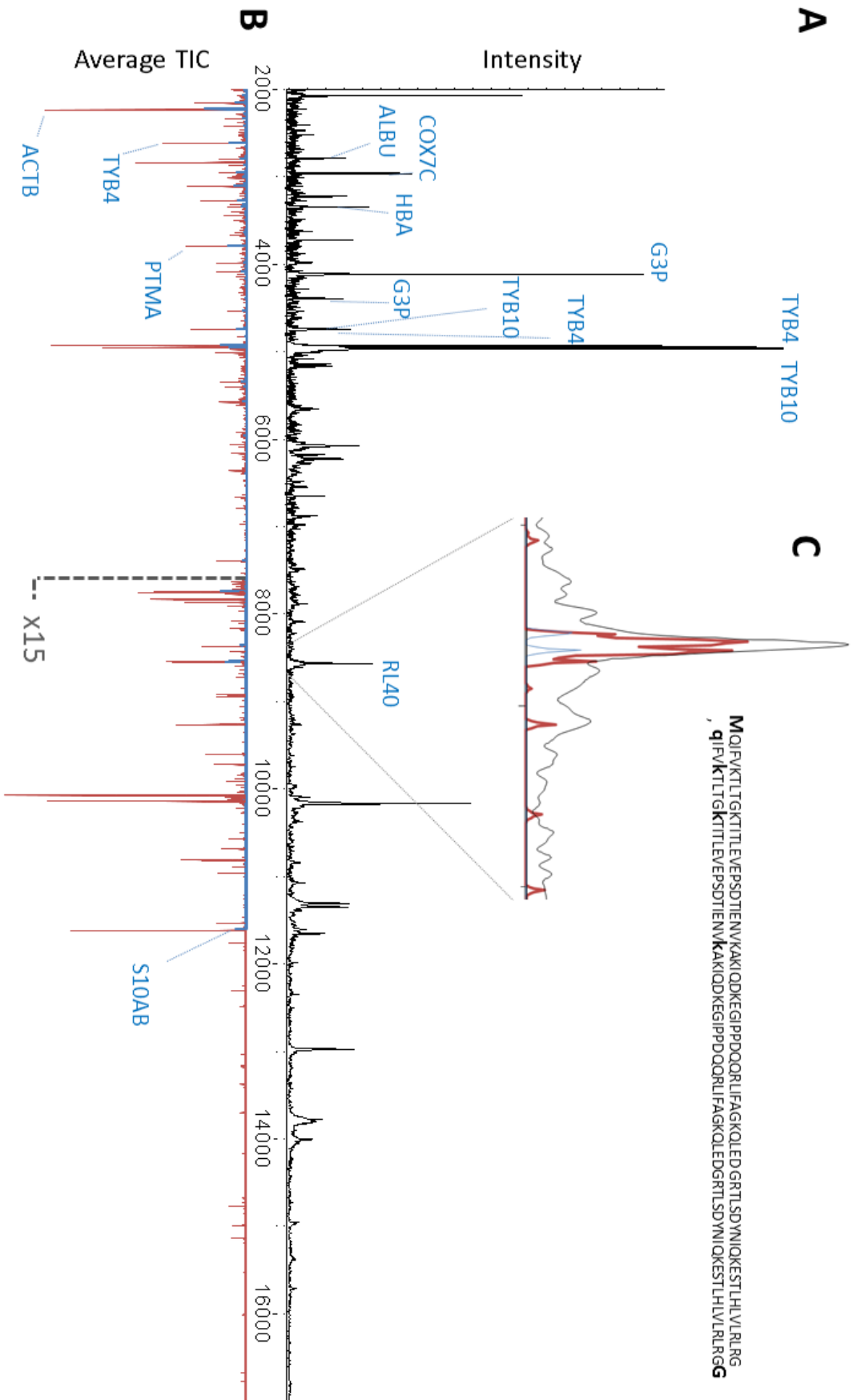


Figure 4-4: Zoom into MALDI spectrum in the mass range from 2000 to 4000 Da.. In red reconstructed: average TIC of MS/MS. In blue: reconstructed average TIC of identified spectra. Peptides sequences and proteins were matched from top down identification data.

Acronyms:

ACTB: actin beta, **ALBU:** albumin, **COX7C:** cytochrome c oxidase subunit VIIc, **F10A1:** Hsc70-interacting protein **G3P:** glyceraldehyde 3-phosphate, **H31:** histone H3, **HBA:** hemoglobin alpha, **MIF:** Macrophage migration inhibitory factor, **PTMA** Prothymosin alpha **RL40:** ubiquitin-60S ribosomal protein L40, **RS14:** 30S ribosomal protein S14, **S10A8:** protein S100-A8, **TYB10:** thymosin beta-10, **TYB4:** thymosin beta-4



Overlying the MALDI spectrum and the reconstructed spectrum and focusing on a smaller mass range permits a more detailed comparison (Figure 4-4). Such an overlay also enables to judge the calibration of the MALDI measurement as the mass values from the LC-MS/MS experiment are accurate and precise. The resemblance of the mass values is still good which often allows annotating the peaks identity but often the intensities do not correspond very well. Furthermore this view enables to recognize if a peak in the MALDI spectrum is generated from two or more ion species but could not be resolved in the linear mode (Figure 4-3 C). Still for some peaks no corresponding peak can be found in either direction. This may have several reasons. One might be that the MALDI species are sodium or potassium adducts whereas most ESI species are protonated. A not corresponding peak might also be the result of in-source fragmentation during the ESI process, which would lead to shorter and therefore lighter species as measured in MALDI MS. An indicator for such a process would be a high frequency of prolines at the terminal position of the identified sequences as peptides including proline are prone to induce fragmentation [178]. However, an investigation of the terminal amino acids (as well as the previous and following amino acid in the sequence of the respective protein) of the identified proteoforms did not show any over-representation of proline compared to its frequency in the human proteome. In-source fragmentation can therefore be excluded.

Discussion

High resolution MS instruments coupled to nanoLC systems permit to acquire data with sufficient quality for the identification of long and highly charged proteoforms. As shown in Chapter 3, employing a reduced database not only aids the identification of proteoforms but also allows generating exclusion lists. This strategy prevents repeated sequencing of already identified high intensity peaks in consecutive runs and therefore a gain in analytical depth. For the samples analyzed in this chapter no such bottom-up generated database was available and minimal versions of the Uniprot database for human and mouse had to be employed for database searching. This resulted in a searching time of about one day per sample which made the exclusion list approach not appropriate. Therefore a different approach was used to iterate the sample in the higher mass range. But even though the choice of the mass mode option led to a high number of sequencing events for precursors with a deconvoluted mass above 5 kDa,

the success was limited. Besides the acquisition related issues this result might be due to several possible reasons: i) the species selected possess modifications which were not expected and therefore not included in the search, or ii) the modification is an internal crosslink e.g. disulfide bridges like in the case of defensins, iii) the database was chosen too small and did not contain the respective protein.

Nevertheless a large number of proteoforms was identified and some of those could be unambiguously linked to the MALDI spectrum of the respective sample. Missing counterparts in both directions might also be addressed by complementary ionization methods [179,180]. Still this approach is promising and with further development into the fragmentation of larger precursors, this mass range should also be more easily addressable.

Furthermore it was shown that long identified sequences reveal the species from which it was derived. This might open the possibility to distinguish intra and extracellular effects of the tumor tissue.

Chapter 5:

PAS-cal: Calibration Standards for MALDI Mass Spectrometry

Introduction

Matrix assisted laser desorption/ionization (MALDI) [181] is widely used in life science research. Beside the molecular weight determination of intact proteins and application to protein and small molecule imaging [182], MALDI instruments are frequently employed for protein identification purposes using either peptide mass fingerprinting (PMF) [107] or liquid chromatography coupled tandem mass spectrometry (LC-MALDI MS/MS) [16,183]. In all these applications, accurate mass determinations are highly important, requiring careful and frequent mass calibration of the instrument. For time-of-flight (TOF) mass analyzers, hardware components such as delayed extraction elements and ion mirrors render mass calibration more complex than the simple quadratic correlation between an ion's flight time and its mass and, as a result, a number of higher order calibration functions have been introduced [184,185]. Especially in MALDI-TOF MS, further factors influence mass accuracy such as temperature, spot morphology [186] and target plate geometry as well as the sample position on the target [187,188], which leads to the need for frequent calibration during measurements.

A generic mass calibration standard suited for bottom-up proteomic applications is therefore highly desirable and should meet a number of requirements; i) cover the typical mass range of (tryptic) peptides, ii) provide many calibration points with small mass increments to enable fitting higher order calibration functions, iii) yield robust intensities in MS and MS/MS, iv) have a molecular composition that is not easily confused with the analytes under investigation and v) have a long shelf life. These criteria are not fully met by widely used peptide calibration

standards that typically encompass few naturally occurring peptides or non-peptidic polymers such as poly(propyl-glycol), PPG [185].

Here the design, synthesis and application of a calibration standard termed PAS-cal that addresses the above needs is described. The design was inspired by the protein PASylation technology recently introduced by Schlapschy et al. [189] in which recombinant therapeutic proteins are furnished with multiple sequential copies of a [ASPAAPSAPPAA]_n peptide cassette consisting solely of the structurally disordered amino acids proline, alanine and serine (hence the name PAS), which leads to an expanded hydrodynamic volume and strongly increases the half-life of said proteins in the blood stream by retarding kidney filtration. Among several useful features of these sequences, they do not occur in natural proteins, they are highly soluble under aqueous conditions and their lack of stable structure allows peptide synthesis in a straightforward manner. Based on these considerations, 30 PAS-cal peptides were synthesized ranging from 8 to 37 amino acids in length utilizing the same sequence motif that can be used in combination or alone to calibrate MALDI MS and MS/MS spectra. This enabled the efficient multi-point calibration of such spectra, resulting in an increased number of peptide and protein identifications from complex mixtures analyzed by LC-MALDI MS/MS.

Based on the same idea a PAS-cal protein calibration standard covering a mass range up to 20 kDa was designed and expressed as a SUMO fusion protein in *E.coli*. The sequence includes three arginine residues resulting in a set of four PAS-polypeptides of different length after digestion with trypsin. A mixture of the intact protein and its tryptic peptides measured by MALDI MS leads to a spectrum with mass distances of about 2000 Da making it a perfect standard for MALDI MS of smaller proteins and MALDI imaging applications.

Methods

Peptide synthesis

(Performed by Ksenia Bashkueva during her Bachelor thesis at the chair of proteomics and bioanalytics)

PAS-cal peptides were produced by solid-phase synthesis following the standard Fmoc strategy on a parallel peptide synthesizer (MultiPep, Intavis, Cologne, Germany) at 2- μ mol scale. For each peptide, proline, alanine or serine (Intavis) were initially coupled to a solid support containing an immobilized Lys or Arg residue (TentaGel S Trt-Arg(Pbf)Fmoc, TentaGel S Trt-Lys(Boc)Fmoc Rapp Polymere GmbH, Tübingen, Germany) followed by the sequential coupling of further amino acids (compare Table 5-1 for sequences). To prevent the generation of erroneous sequences, free amino groups were capped by acetylation after each synthesis cycle and prior to the next Fmoc de-protection step. After completion of the synthesis scheme, peptides were released from the resin using 92.5% trifluoroacetic acid (TFA), 5% tri-isopropylsilane and 2.5% water and subsequently lyophilized. Synthesis control was performed by MALDI MS analysis and the crude PAS-cal peptides were subsequently used in isolation or mixed empirically to obtain a balanced MS intensity distribution.

Preparation of *E. coli* proteolytic digest

(Performed by Christoph Rösli, DKFZ, Heidelberg)

E. coli cells (TG1 Electroporation-Competent Cells, Agilent, Waldbronn, Germany) were grown in LB broth at 37°C. Supernatant was removed after centrifugation at 10,000 x g for 20 min at 4 °C, and 10 ml lysis buffer (10 mM Tris-HCl (pH 7.5), 5 mM EDTA, 1 x complete Protease Inhibitor Cocktail, Roche Applied Sciences, Mannheim, Germany) per gram wet bacteria was added. Cells were lysed using a sonifier (Branson Sonifier W250D, Danbury, US) at 40% intensity for 6 intervals (10 sec on, 15 sec off). After centrifugation (10'000 x g, 4 °C, 25 min), the supernatant was frozen at -20 °C until further use. Disulfide bridges were reduced using dithioerythritol (5 mM, 60°C, 30 min) and free sulfhydryl groups were alkylated with iodoacetamide (15 mM, room temperature, in the dark, 30 min). Proteins were denatured in 6 M urea and 50 mM Tris-HCl (pH 8.0), followed by protein digestion using a mixture of LysC and Trypsin (Promega, Mannheim, Germany, protein:protease ratio of 25:1) for 4 h at 37 °C. Subsequently, the buffer system was adjusted to a urea concentration of 0.86 M by adding 50 mM Tris-HCl (pH 8.0), 1

mM CaCl₂ in water. The digestion mixture was held at 37 °C und agitated overnight. Protease activity was quenched by the addition of TFA to a final concentration of 0.1% v/v. The resulting peptide mixture was desalted using OMIX C18 Tips (Agilent) according to the manufacturer's protocol and the eluted sample was dried in a vacuum concentrator and stored at -20 °C.

MALDI MS measurements of PAS-cal peptides and peptide mass fingerprinting

PAS-cal peptides and mixtures thereof were dissolved in water to 2 µmol/ml (estimated from synthesis scale) and mixed 1:1 with α-cyano-4-hydroxycinnamic acid (HCCA, 10 mg/ml in 50% acetonitrile, ACN, 0.1% TFA, Bruker Daltonik, Bremen, Germany). Samples were spotted onto a stainless steel target (Bruker Daltonik) using 0.5 µl per spot. Samples were then recrystallized using 0.1 µl ethanol, resulting in a homogeneous matrix layer. MALDI TOF spectra were acquired in positive ion reflectron mode on an UltrafleXtreme MALDI-TOF/TOF instrument (Bruker Daltonik, operated by FlexControl 3.3, sample rate of 4 GS/s) by summing up to 1000 laser shots over a mass range of 500-3500 *m/z*. Tryptic peptides from bovine cytochrome C and bovine serum albumin as well as peptide calibration standard II (all from Bruker Daltonik) were prepared accordingly. For peptide mass fingerprinting, UltrafleXtreme mass spectra were converted into peak lists using flexAnalysis (version 3.3, Bruker Daltonik). Resulting xml files were submitted to Mascot (2.4.1 Matrix Science, London, UK) and searched against the SwissProt database (version 57, taxonomy filter set to 'other mammals' considering carbamidomethylation of cysteines and oxidation of methionines as variable modifications. Peptide mass tolerance was set to 75 ppm, full trypsin specificity was required, allowing up to 2 missed cleavages.

LC-MALDI MS/MS measurements and data processing of E. coli protein digests

(performed by Christoph Rösli, DKFZ, Heidelberg)

E. coli peptides were dissolved in 5% ACN, 0.1% TFA in water to a concentration of 1 µg/µl and 10 µl were loaded directly onto a reversed phase column (1.7 µm BEH130 C18, 75 µm x 250 mm, Waters, Milford, USA). Peptides were separated on a nanoACQUITY UPLC (Waters) using a binary solvent system (solvent A: 0.1 % TFA in water; solvent B: 0.1 % TFA in ACN) and using the following gradient: 5-11% B in one min; 11-14% B in 3 min; 14-30% B in 63 min, 30-40% B in 13 min. Eluting peptides were mixed online with HCCA (3 mg/ml HCCA, ProteoChem, Cheyenne, USA in 70% ACN, 0.1% TFA in water containing the standard peptides Gonadoliberin, Angiotensin I, Neurotensin and Adrenocorticotrophic Hormone Fragment (ACTH) 18-39, each at a

concentration of 150 pmol/ml. The mixture was spotted onto an Opti-TOF LC/MALDI insert (AB SCIEX) using an automatic MALDI spotter (SunCollect, SunChrom, Friedrichsdorf, Germany). Spotting was started with a delay of 10 min after sample injection. 1200 spots were prepared per LC run, corresponding to 4 s wide fractions per spot. MALDI MS and MS/MS measurements were carried out on a TOF/TOF 5800 System (ABSCIEX, Framingham, USA) controlled by TOF/TOF Series Explorer Software V4.1.0 (build 12). Per spot, 8 x 250 shots for MS spectra and up to 12 x 250 shots per MS/MS spectrum were summed up. MS spectra were internally calibrated using the 4 peptides spiked into the matrix solution. MS/MS spectra were calibrated using 6 calibration points from ACTH 18-39 or 50 calibration points of the 37 amino acid PAS-cal peptide (PAS-cal37). Tandem mass spectra from LC-MALDI MS/MS experiments were processed by Protein Pilot (Version 4.5, ABSCIEX) employing the Paragon Algorithm 4.5.00, 1654.

Protein identification from LC-MALDI data

The resulting peak lists (in Mascot generic format, MGF) were searched against the SwissProt database (v57, taxonomy filter *Escherichia coli*) using Mascot (version 2.4.1) with a mass tolerance for precursors of 25 ppm and 0.15 Da for fragment ions. Enzyme specificity was set to trypsin, allowing a maximum of two missed cleavages while carbamidomethylation of cysteines was set as fixed modification. Search results were exported from Mascot and further analyzed using Microsoft Excel 2010 and Tableau 8.1 (tableausoftware, Seattle, US). Peptide spectrum matches (PSMs) with a score lower than the Mascot identity score of 30 were categorically removed prior to subsequent analysis.

PAS-cal protein expression in *E.coli* and digest tryptic digestion

(performed by Joscha Breibeck at the Lehrstuhl für Biologische Chemie, WZW, TUM)

The design and expression of the Pas-cal proteins were performed by Joscha Breibeck and will only be briefly described. The SUMO-PAS-cal fusion protein was expressed in *E. coli* BL21 (DE3). Cells were grown at 37 °C in LB medium. The cell pellet was homogenized in a French press, clarified by centrifugation and sterile filtration. The protein solution was loaded onto a Ni/NTA resin and eluted using an imidazole gradient. The resulting fractions were dialyzed and subjected to ion exchange chromatography. SUMO-proteins were concentrated by ultracentrifugation and cleaved by recombinant truncated SUMO protease. Preparative size exclusion chromatography was used to clean the PAS-cal protein. Aliquots were incubated with sequence grade trypsin.

MALDI MS of PAS-protein and peptides

The solutions of the intact PAS-protein and the trypsin-digested PAS-cal (0.28 mg/ml in 50 mM NH_4HCO_3) were mixed in the ratio 3:1 in order to trigger the formation of dimeric and trimeric peptide adduct ions in the MALDI spectrum. To the resulting peptide mixture an equal volume of 2,5-dihydroxybenzoic acid solution (DHB, 10 $\mu\text{g}/\text{mL}$ in 30% acetonitrile in water, 0.2% trifluoroacetic acid) was added and the mixture was spotted according to the dried droplet method [3]. MALDI-MS analysis was performed with the instrument UltrafleXtreme from Bruker (Bremen, Germany) in the linear positive mode (source voltage: 25 kV). Using 1000 shots per spectrum, an m/z range from 1000 to 23000 was analyzed with ion suppression below an m/z ratio of 1500. m/z values for the various expected peptide species were calculated with Compass IsotopePattern software (Version 1.3) based on the amino acid sequences.

Results and Discussion

PAS-cal peptides: mass calibration standard for bottom-up proteomics

To establish a peptidic mass calibration standard for MALDI MS in proteomics that meets the requirements outlined in the introduction, two sets of 30 peptides each (termed PAS-cal) were synthesized based on the basic PAS repeat "ASPAAPSAPPAA" with a C-terminal lysine or arginine residue added to aid ionization in the MALDI process. This design principle led to tryptic-like peptides covering a mass range from about 750 Dalton (8 amino acids) to 3200 (37 amino acids) Dalton (Table 5-1) which is also the typical mass range of peptides generated by trypsin digestion of proteins or protein extracts.

While lysine-terminated PAS-cal peptides tended to form sodium and potassium adducts (data not shown), arginine containing peptides were dominantly observed as $[M+H]^+$ and yielded higher intensities as described before [190]. Due to the fact that PAS-cal peptides are intrinsically unstructured [189], peptide synthesis even for the longer peptides was straightforward and allowed to directly use the crude synthesis products for subsequent experiments. The high solubility of these peptides also rendered them stable under aqueous conditions, which permitted use of the same PAS-cal preparations over more than two years without noticeable losses.

Two applications for the PAS-cal peptides were investigated: i) the mass calibration of MALDI TOF spectra for protein identification by peptide mass fingerprinting and ii) the calibration of MALDI MS/MS spectra for protein identification via fragment ion based database searching. For the first application, 30 arginine terminated PAS-cal peptides were mixed to yield a balanced distribution of intensities in a MALDI spectrum (Figure 5-1 A) covering the range from 756.39 to 3140.62 m/z . The peak spacing corresponds to the residue masses of one of the three amino acids that make up the PAS-cal peptides, thus allowing multi-point calibration of MALDI MS spectra. To evaluate which of the calibration methods provided by the instrument software (flex control) performs best, PAS-cal and the commercially available "peptide calibration standard II" (Bruker) were mixed and analyzed together.

Table 5-1: Arginine-terminated PAS-cal peptides.

The PAS repeat unit used in this study is indicated by the dotted lines.

Monoisotopic [M+H] ⁺	Sequence length	Peptide sequence
756.399878	8	ASPAAPSR
827.436992	9	ASPAAPSAR
924.489756	10	ASPAAPSAPR
1021.54252	11	ASPAAPSAPPR
1092.57963	12	ASPAAPSAPPAR
1163.61675	13	ASPAAPSAPPAA:R
1234.65386	14	ASPAAPSAPPAA:AR
1321.68589	15	ASPAAPSAPPAA:ASR
1418.73865	16	ASPAAPSAPPAA:ASPR
1489.77577	17	ASPAAPSAPPAA:ASPAR
1560.81288	18	ASPAAPSAPPAA:ASPAAR
1657.86565	19	ASPAAPSAPPAA:ASPAAPR
1744.89767	20	ASPAAPSAPPAA:ASPAAPSR
1815.93479	21	ASPAAPSAPPAA:ASPAAPSAR
1912.98755	22	ASPAAPSAPPAA:ASPAAPSAPR
2010.04032	23	ASPAAPSAPPAA:ASPAAPSAPPR
2081.07743	24	ASPAAPSAPPAA:ASPAAPSAPPAR
2152.11454	25	ASPAAPSAPPAA:ASPAAPSAPPAA:R
2223.15166	26	ASPAAPSAPPAA:ASPAAPSAPPAA:AR
2310.18369	27	ASPAAPSAPPAA:ASPAAPSAPPAA:ASR
2407.23645	28	ASPAAPSAPPAA:ASPAAPSAPPAA:ASPR
2478.27356	29	ASPAAPSAPPAA:ASPAAPSAPPAA:ASPAR
2549.31068	30	ASPAAPSAPPAA:ASPAAPSAPPAA:ASPAAR
2646.36344	31	ASPAAPSAPPAA:ASPAAPSAPPAA:ASPAAPR
2733.39547	32	ASPAAPSAPPAA:ASPAAPSAPPAA:ASPAAPSR
2804.43258	33	ASPAAPSAPPAA:ASPAAPSAPPAA:ASPAAPSAR
2901.48535	34	ASPAAPSAPPAA:ASPAAPSAPPAA:ASPAAPSAPR
2998.53811	35	ASPAAPSAPPAA:ASPAAPSAPPAA:ASPAAPSAPPR
3069.57523	36	ASPAAPSAPPAA:ASPAAPSAPPAA:ASPAAPSAPPAR
3140.61234	37	ASPAAPSAPPAA:ASPAAPSAPPAA:ASPAAPSAPPAAR

PAS- repeat

Acquired mass spectra were calibrated internally against either the mass list of the Bruker standard or the mass list of the PAS peptides. For both standards the ‘cubic- enhanced’ calibration method outperformed the other three options (linear, linear correction and quadratic) and resulted in relative standard deviations of 2 ppm each, indicating that this is the best achievable mass accuracy on this particular instrument (data not shown).

As the mass accuracy of MALDI TOF instruments depends on several parameters including e.g. matrix morphology [186] or spot localization on the target plate, each sample has to be

calibrated. This may be achieved by internal spectrum calibration using e.g. endogenous or spiked trypsin autolysis products [191]. However, this approach is often not robust as the endogenous levels of trypsin autolysis peptides are largely unpredictable and ‘guessing’ the appropriate amount of spiked standard can be difficult.

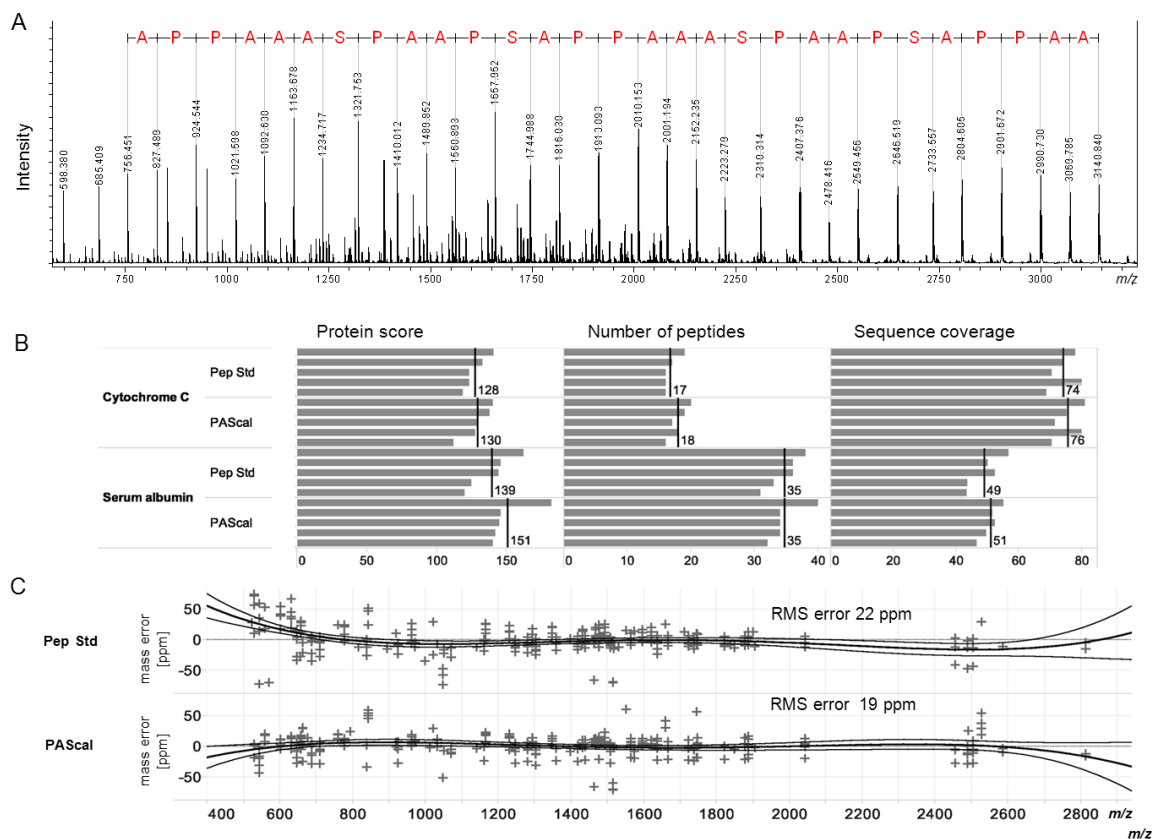


Figure 5-1: Evaluation of PAS-cal peptides for protein identification by peptide mass fingerprinting. A] MALDI TOF MS spectrum of the PAS-cal peptide mixture (30 peptides). **B]** PMF search results for bovine serum albumin and cytochrome c tryptic digests. Vertical lines indicate averages of five measurements. **C]** Relative mass errors for all measurements including a 4th order polynomial trend line (black line) along with the 95% confidence interval (grey lines).

In addition, correcting the higher order effects of time-to-mass calibration usually requires multi-point calibration. As a result, MALDI MS spectra are most often externally calibrated. To evaluate the PAS-cal standard (comprising the 30 mixed peptides) as an external calibrant for peptide mass fingerprinting, digests of bovine serum albumin (BSA) and bovine cytochrome C were analyzed by MALDI TOF MS in reflectron mode and the measured spectra were either calibrated on the PAS-cal standard or on the commercial Bruker standard (both spotted

nearby). Mascot search results showed that the performance of both standards was comparable in terms of number of detected peptides and sequence coverage but PAS-cal calibration exhibited slight improvements in terms of mass error and protein score (Figure 5-1 B, C; note that the Mascot score is a log₁₀ score, thus a difference of 1 score points translates into 10x higher confidence). It was also observed that the intensities of the high *m/z* PAS-cal peptides decreased as the MALDI ion source of the instrument became contaminated over time owing to the use of high frequency lasers. This effect may serve as a simple QC measure to define an instrument maintenance interval.

Following the trend in proteomic research to analyze more complex mixtures, today's analytical strategies typically combine peptide separation by liquid chromatography with tandem mass spectrometry (LC-MS/MS) both on electrospray ionization (ESI) and on MALDI instruments. The success of peptide and protein identification in such experiments critically depends on mass accuracy [192] and largely the same requirements apply with respect to the properties of calibration standards for MS and MS/MS calibration. One notable difference is that a single peptide generating fragment ions covering the entire mass range may suffice to calibrate a tandem mass spectrum. PAS-cal peptides are ideally suited for this purpose as the C-terminal arginine residue promotes the formation of a strong γ -ion series. The MALDI tandem mass spectrum of the longest PAS-cal peptide (PAS-cal37) is shown in Figure 5-2. Its appearance resembles that of the PAS-cal mixture because the singly charged γ -ions of PAS-cal37 have the same chemical composition as the protonated peptides from the synthetic mixture.

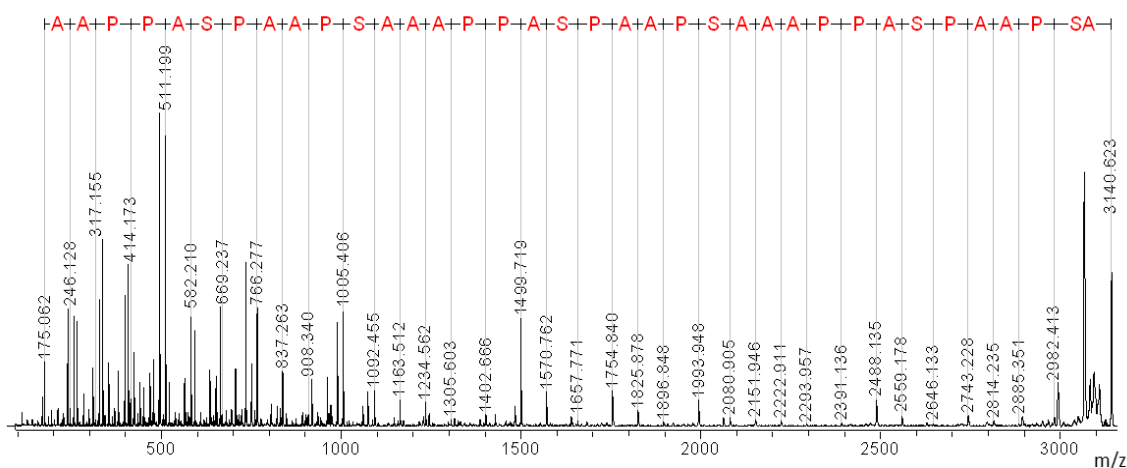


Figure 5-2: Annotated fragment spectrum pattern of PAS-cal37 peptide.
Exhibiting fragment ions covering the entire sequence.

To evaluate the utility of PAS-cal37 (3140.61 Da, 50 calibration points) for the multi-point calibration of tandem mass spectra, two LC-MALDI MS/MS experiments using tryptic digests of *E. coli* were performed and the results were compared with analogous duplicate experiments calibrated on the fragment ion spectra of ACTH 18-39 (2464.19 Da, 6 calibration points), which is widely used as MS2 calibration standard. As shown in Table 5-3, PAS-cal37 calibration led to improvements in all metrics applied, including an increase of peptide spectrum matches (PSM) by 18% with a concomitant 15% increase in peptide identifications and 13% increase in protein identifications. In addition, the average and median Mascot scores were also slightly improved in the PAS-cal37 calibrated experiments (again note log 10 space of the Mascot ion score so that even small score differences of three points correspond to a considerable effect in absolute space).

Table 5-2: Result summary of LC-MALDI MS/MS measurements:

Note that only identifications were accepted that exceeded the Mascot peptide identity threshold of 30.

Sample	# distinct protein identifications	# distinct peptide identifications	# peptide spectrum matches	# submitted queries	Average mascot score	Median mascot score
PAS 1	994	4,069	4,277	16,446	65	56
PAS 2	1,002	4,130	4,338	16,418	64	56
Total	1,178	4,738	8,615	32,864	65	56
ACTH 1	838	3,310	3,437	15,268	62	53
ACTH 2	910	3,722	3,868	16,124	63	53
Total	1,043	4,123	7,305	31,392	62	53

Further analysis of PSMs with ion scores of equal or greater than the Mascot identity threshold (of 30) revealed two factors that are responsible for the observed improvements when using PAS-cal37 for the calibration of tandem mass spectra. First, more fragment ions were matched within the search tolerance (Figure 5-3 A, B upper panel) and, second, the median fragment ion mass error was smaller compared to ACTH calibration (Figure 5-3 B middle and lower panel). Obviously, the number of matching fragments drives the identification score and the lower mass error of the fragments aids in discriminating against alternative PSMs for the same tandem mass spectrum.

Furthermore the suitability of the PAS-cal standard for retention time and mass calibration in LC-ESI-MS/MS experiments (data not shown) was investigated. While the standard can, in principle, be used for these purposes, there are limitations. First, all PAS-cal peptides are

hydrophilic which renders their elution time profile in reversed-phase HPLC separations of peptides rather narrow. This may, however, be addressed in the future by adapting the underlying design principle to include more hydrophobic amino acids. Second, PAS-cal peptides (like any other peptides) pick up more charges as the sequence gets longer which somewhat narrows the m/z range across which ESI spectra can be calibrated. This is not a strong limitation though, as tryptic peptides from complex digests show a very similar behavior. Alternatively, the QCAL standard reported by Simon Gaskell's group [193] or the PAS-cal protein standard can be used for extending the mass range for the calibration of ESI spectra.

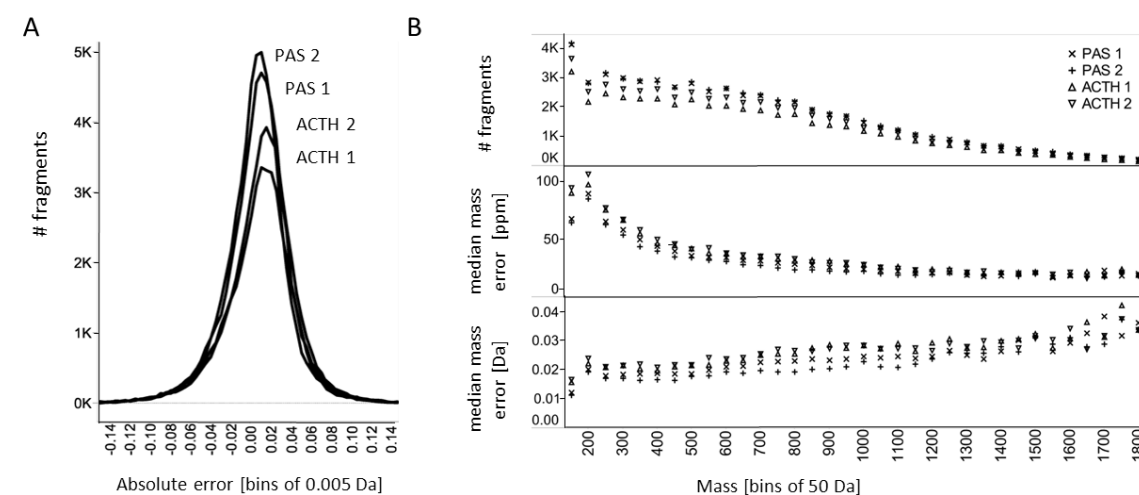


Figure 5-3: Multi-point calibration of tandem mass spectra using PAS-cal37 and ACTH.

A] Distribution of absolute mass errors of matching fragment ions **B]** Upper panel: number of matched fragments (within a tolerance of 0.15 Da) as a function of fragment ion mass (in bins of 50 Da). Middle panel: median mass error of matched fragments in ppm as a function of fragment ion mass. Lower panel: same as in B but in Dalton.

PAS-protein and its tryptic peptides for mass calibration in the range 2-20kDa

For the calibration of higher mass ranges a PAS-cal protein was synthesized. Its sequence includes three arginine residues (Figure 5-4) which are cleavage sites for trypsin.

```
AASPAAPSAPPAAASPAAPSAPPAAASPAAPSAPPAAASPAAPSAPPAR
AASPAAPSAPPAAASPAAPSAPPAAASPAAPSAPPAAASPAAPSAPPAA
ASPAAPSAPPAAASPAAPSAPPARASPAAPSAPPAAASPAAPSAPPAA
ASPAAPSAPPAAASPAAPSAPPAAASPAAPSAPPAAASPAAPSAPPAA
SPAAPSAPPAAASPAAPSAPPARASPAAPSAPPAAASPAAPSAPPARA
```

Figure 5-4: Sequence of the expressed PAS-cal protein containing 245 amino acids including four arginine residues (highlighted in yellow)

Measured with MALDI TOF in the linear mode intact PAS-cal was observed in three (seldom four) charge states whereas the doubly protonated ion species showed the highest intensity (not shown). A mixture of the intact polypeptide and its tryptic peptides, showed the four standard peptides (A, B, C, D) as well as the singly (CS 1) and doubly (CS 2) charged states of the intact concatamer at their expected m/z ratios (Figure 5-5 and Table 5-3).

By adjusting the relative amount of the tryptic peptide fragments to a threefold excess over the intact PAS-cal polypeptide, the formation of di- and trimeric adduct species between the peptides could be observed. It is well-known from MALDI experiments that several peptides can “share” one single charge by coordinating the same proton, leading to m/z ratios resulting from the sum of the single species involved in the complex [10]. Most likely in this case, the bridging protons are shared by the basic arginine residues present in each of the PAS-cal peptides. Exploiting this phenomenon of adduct formation, the m/z gap between the singly and doubly charged states of the intact polypeptide could be complemented. Due to the intrinsic sequence repetition of the concatamer, the m/z ratio 12218.5 can be assigned to two different possible dimeric peptide complexes.

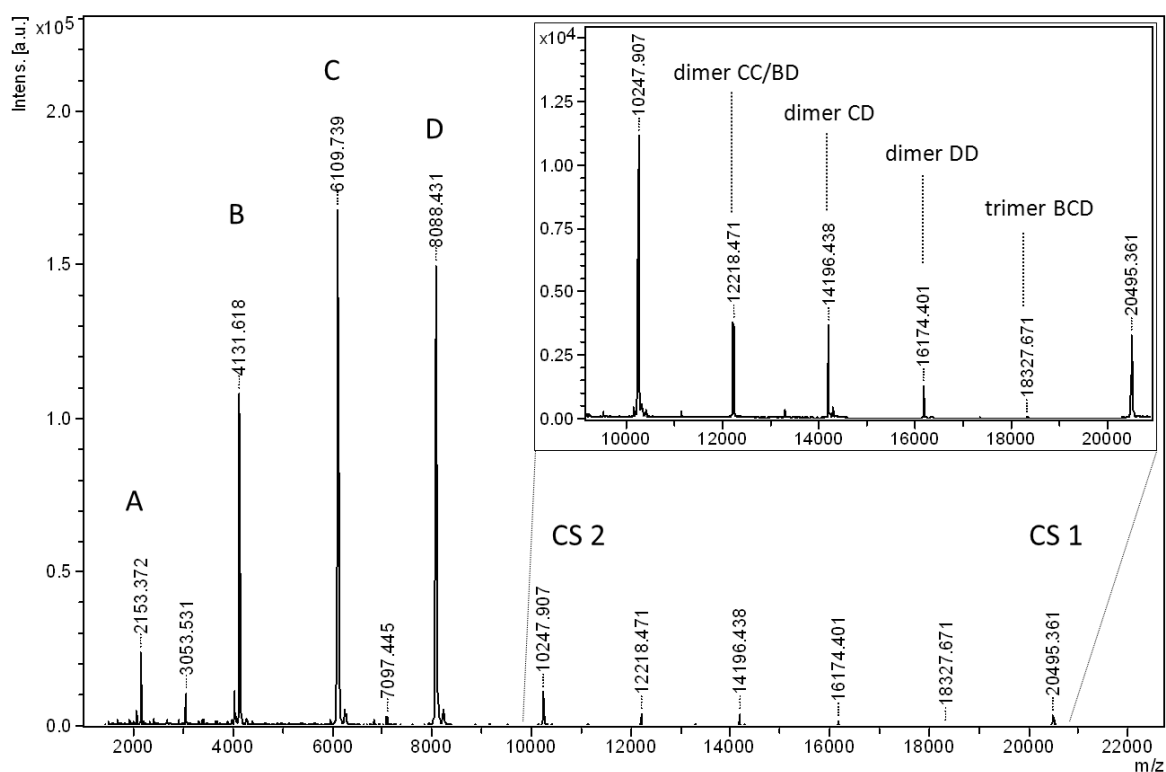


Figure 5-5: MALDI-MS spectrum of PAS-cal and its tryptic peptides acquired in linear mode. The inset shows an enlarged view of the charge states +1 and +2 of the intact polypeptide as well as the di- and trimeric species formed by the tryptic peptide fragments.

Table 5-3: MALDI-MS signals measured for PAS-cal peptides.

Species		calculated m/z-signals	observed m/z-signals
Peptide 1 (A)	[M+H] ⁺	2153.379	2153.372
Peptide 2 (B)	[M+H] ⁺	4131.548	4131.618
Peptide 3 (C)	[M+H] ⁺	6109.718	6109.739
Peptide 4 (D)	[M+H] ⁺	8087.888	8088.431
PAS-cal (intact)	[M] ²⁺ (CS 2)	10247.775	10247.907
	[M] ⁺ (CS 1)	20495.536	20495.361
Dimer CC/BD	[M+M+H] ⁺	12218.439	12218.471
Dimer CD	[M+M+H] ⁺	14196.599	14196.438
Dimer DD	[M+M+H] ⁺	16174.769	16174.401
Trimer BCD	[M+M+M+H] ⁺	18327.147	18327.671

These additional effects of the PAS-cal polypeptide standard in MALDI experiments make it suited for the calibration of a broad m/z range. Even though the di- and tri-mere formation could only be observed for DHB matrix, all other peaks were also observed with sinapic acid guaranteeing a calibration point at 20495 m/z . This makes the PAS-protein and peptide mixture an ideal calibrant for MALDI imaging applications. Analogue to the synthesised PAS-cal37 peptide the tryptic peptides (Table 5-3) could be used as calibration standards for MALDI-MS/MS spectra up to a mass range up to typically used in middle-down (similar to bottom-up but proteases/agents with less frequent cleavage sites are used [194,195]) or top-down experiments targeting small to medium proteoforms.

Chapter 6: Final Conclusions

Protein MALDI imaging is an appealing methodology as it enables to study proteomic patterns in a spatially resolved fashion on the background of tissue morphology. One of the main challenges of the field is the identification of the molecules detected in the imaging experiment. This thesis was set out to address these issues.

By mimicking the matrix layer preparation conditions, the here established sample preparation procedure offers a straightforward way for the isolation of the matrix proteome with minimal sample amount needed and which can seamlessly be integrated into the MALDI imaging workflow. The bottom-up analysis showed that mainly small, soluble and high abundant proteins co-crystallize (or at least conglomerate) with the MALDI matrix molecules and are therefore detectable in the imaging experiment. A comparison with identified species from literature underlines these findings and defines the list of possible markers that can be derived from MALDI imaging studies.

The sample preparation method also enables the direct analysis via LC-ESI-MS and allows for the direct sequencing of proteoforms from the matrix layer. It was shown that the employment of subset-databases, an interactive exclusion list strategy and the utilization of the 'mass mode' facilitated the identification of proteoforms up to a mass of 12 kDa. Still the employed strategies were not perfectly able to address the specific issues of data dependent acquisition methods of the used instrument when measuring intact proteoforms. A recently published data acquisition/identification pipeline called autopilot [168] unifies those strategies and enables the instrument software to perform advanced data dependent acquisition. The selection criteria are based on the 'targets' (mass species in different charge states) instead of m/z values. Those are compared in real-time to a database of already identified proteoforms. Sequencing events are only triggered for unknown target masses or to improve sequence coverage of not yet sufficiently covered proteoforms by varying fragmentation conditions. Full sequence coverage including all post-translational modifications is desirable as it results from all processes its precursor underwent.

The comparably long sequences of the identified proteoforms often enable to deduce the organism from which it was derived. This might be valuable information to investigate the interplay between the tumor tissue and the host organism in xenograft models. The analysis of the identified proteoforms showed that there are many C- and N- terminal fragments of proteins. This suggests that there are proteolytic processes in those tissues. However, the question if this is an effect of sample handling or of processes before deriving the sample was not investigated.

The introduced graphical method for matching the identified proteoforms to the peaks of a MALDI spectrum showed that the resemblance of both methods is good but not perfect. This might be a result of the different ionization efficiency of the same species in ESI and MALDI. The established method employing ESI-MS can be seen as a shortcut which often leads to identification of MALDI imaging species. For unmatched peaks of interest the classical way of fractionation and screening of the fractions under imaging MS conditions might not be circumventable.

For the correlation of the identified proteoforms with the peaks of the MALDI spectrum the mass accuracy of the latter is an important criterion. This is achieved by a proper calibration of the MS spectrum. For this purpose a protein was designed only containing proline, alanine and serine residues as well as 3 basic arginine residues. Mixing the intact protein with its tryptic peptides results in an excellent standard covering the typical imaging mass range with evenly distributed calibrations points. Based on the same PAS sequence repeat a multi-peptide calibration standard was developed and has proven its applicability in peptide mass fingerprinting. Single peptides can also be employed as MS/MS calibration standard and it has been shown that MS/MS accuracy and therefore overall performance improves by using many calibration points resulting from the fragment spectra of the standard peptide.

Appendix

Abbreviations and Acronyms

ACN	acetonitrile
AGC	automatic gain control
BSA	bovine serum albumin
CID	collision induced dissociation
Da	Dalton
DESI	desorption electro spray ionisation
DHB	2,5-dihydroxybenzoic acid
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
e.g.	exempli gratia
ECD	Electron-capture dissociation
ESI	electro spray ionisation
ETD	Electron-transfer dissociation
FA	formic acid
FDR	false discovery rate
FFPE	formalin fixed paraffin embedded
FWHM	full width at half maximum
g	gram
HCCA	alpha-cyano-4-hydroxycinnamic acid
HCD	higher-energy collisional dissociation
HCl	hydrogen chloride
HPLC	high performance liquid chromatography
i.e.	id est
IAA	iodoacetamide

IMS	imaging mass spectrometry
ISD	in-source decay
ITO	indium-tin-oxide
LC	liquid chromatography
LDS	lithium dodecyl sulphate
LID	laser desorption/ionization
<i>m/z</i>	mass-to-charge ratio
MALDI	matrix-assisted laser desorption/ionization
mgf	mascot generic format
min	minute
mM	millimolar
MSI	mass spectrometry imaging
NP40	nonyl phenoxy polyethoxylethanol 40
PAGE	Polyacrylamide gel electrophoresis
PSM	peptide spectrum match
s	second
SA	sinapinic acid
SDS	Sodium dodecyl sulfate
SIMS	secondary ion mass spectrometry
STAGE	stop and go extraction
TFA	trifluor acetic acid
TIC	total ion current
TOF	time of flight
Tris	tris(hydroxymethyl)aminomethane)
TUM	Technische Universität München
u	atomic mass unit
V	volt

References

- [1] J.J. Thomson, Rays of positive electricity, Lond. Edinb. Dublin Philos. Mag. J. Sci. 20 (1910) 752–767.
- [2] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, Science. 246 (1989) 64–71. doi:10.1126/science.2675315.
- [3] M. Karas, F. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, Anal. Chem. 60 (1988) 2299–2301.
- [4] R.F.K. Herzog, F.P. Viehböck, Ion Source for Mass Spectrography, Phys. Rev. 76 (1949) 855–856. doi:10.1103/PhysRev.76.855.
- [5] A. Benninghoven, Beobachtung von oberflächenreaktionen mit der statischen methode der sekundärionen-massenspektroskopie. I die methode, Surf. Sci. 28 (1971) 541–562. doi:10.1016/0039-6028(71)90061-6.
- [6] H.R. Morris, M. Panico, M. Barber, R.S. Bordoli, R.D. Sedgwick, A. Tyler, Fast atom bombardment: A new mass spectrometric method for peptide sequence analysis, Biochem. Biophys. Res. Commun. 101 (1981) 623–631. doi:10.1016/0006-291X(81)91304-8.
- [7] S.E. Harton, F.A. Stevie, H. Ade, Carbon-13 Labeling for Improved Tracer Depth Profiling of Organic Materials Using Secondary Ion Mass Spectrometry, J. Am. Soc. Mass Spectrom. 17 (2006) 1142–1145. doi:10.1016/j.jasms.2006.03.018.
- [8] T. Hoshi, M. Kudo, High resolution static SIMS imaging by time of flight SIMS, Appl. Surf. Sci. 203–204 (2003) 818–824. doi:10.1016/S0169-4332(02)00834-6.
- [9] E.R. Amstalden van Hove, D.F. Smith, R.M.A. Heeren, A concise review of mass spectrometry imaging, J. Chromatogr. A. 1217 (2010) 3946–3954. doi:10.1016/j.chroma.2010.01.033.
- [10] F. Hillenkamp, M. Karas, R.C. Beavis, B.T. Chait, Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers, Anal. Chem. 63 (1991) 1193A–1203A. doi:10.1021/ac00024a002.
- [11] R. Zenobi, R. Knochenmuss, Ion formation in MALDI mass spectrometry, Mass Spectrom. Rev. 17 (1998) 337–366. doi:10.1002/(SICI)1098-2787(1998)17:5<337::AID-MAS2>3.0.CO;2-S.

- [12] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, et al., Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* 2 (1988) 151–153. doi:10.1002/rcm.1290020802.
- [13] H. Hahne, *Studies towards the proteome-wide detection, identification and quantification of protein glycosylation*, TUM, 2012.
- [14] T.W. Jaskolla, M. Karas, Compelling Evidence for Lucky Survivor and Gas Phase Protonation: The Unified MALDI Analyte Protonation Mechanism, *J. Am. Soc. Mass Spectrom.* 22 (2011) 976–988. doi:10.1007/s13361-011-0093-0.
- [15] M. Wilm, Principles of electrospray ionization, *Mol. Cell. Proteomics.* (2011) mcp.R111.009407. doi:10.1074/mcp.R111.009407.
- [16] M. Mann, M. Wilm, Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags, *Anal. Chem.* 66 (1994) 4390–4399. doi:10.1021/ac00096a002.
- [17] J.V. Iribarne, B.A. Thomson, On the evaporation of small ions from charged droplets, *J. Chem. Phys.* 64 (2008) 2287–2294. doi:10.1063/1.432536.
- [18] A.L. Dill, D.R. Ifa, N.E. Manicke, Z. Ouyang, R.G. Cooks, Mass spectrometric imaging of lipids using desorption electrospray ionization, *J. Chromatogr. B.* 877 (2009) 2883–2889. doi:10.1016/j.jchromb.2008.12.058.
- [19] K.H. Kingdon, A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures, *Phys. Rev.* 21 (1923) 408–418. doi:10.1103/PhysRev.21.408.
- [20] A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat, et al., Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer, *Anal. Chem.* 78 (2006) 2113–2120. doi:10.1021/ac0518811.
- [21] P. Mallick, B. Kuster, Proteomics: a pragmatic perspective, *Nat. Biotechnol.* 28 (2010) 695–709. doi:10.1038/nbt.1658.
- [22] Lottspeich, *Bioanalytik*, 2006.
- [23] Q. Hu, R.J. Noll, H. Li, A. Makarov, M. Hardman, R. Graham Cooks, The Orbitrap: a new mass spectrometer, *J. Mass Spectrom.* 40 (2005) 430–443. doi:10.1002/jms.856.
- [24] X. Han, A. Aslanian, J.R. Yates, Mass Spectrometry for Proteomics, *Curr. Opin. Chem. Biol.* 12 (2008) 483–490. doi:10.1016/j.cbpa.2008.07.024.
- [25] J.R. Yates, C.I. Ruse, A. Nakorchevsky, Proteomics by Mass Spectrometry: Approaches, Advances, and Applications, *Annu. Rev. Biomed. Eng.* 11 (2009) 49–79. doi:10.1146/annurev-bioeng-061008-124934.

- [26] E.G. Solon, A. Schweitzer, M. Stoeckli, B. Prideaux, Autoradiography, MALDI-MS, and SIMS-MS imaging in pharmaceutical discovery and development, *AAPS J.* 12 (2010) 11–26. doi:10.1208/s12248-009-9158-4.
- [27] D.S. Cornett, M.L. Reyzer, P. Chaurand, R.M. Caprioli, MALDI imaging mass spectrometry: molecular snapshots of biochemical systems, *Nat. Methods.* 4 (2007) 828–833. doi:10.1038/nmeth1094.
- [28] H. Liebl, Ion Microprobe Mass Analyzer, *J. Appl. Phys.* 38 (1967) 5277–5283. doi:10.1063/1.1709314.
- [29] R. Castaing, G.J. Slodzian, Optique corpusculaire—premiers essais de microanalyse par emission ionique secondaire, *Microscopie.* 1 (1962) 395–399.
- [30] D.S. McPhail, Applications of Secondary Ion Mass Spectrometry (SIMS) in Materials Science, *J. Mater. Sci.* 41 (2006) 873–903. doi:10.1007/s10853-006-6568-x.
- [31] C. Petibois, Imaging methods for elemental, chemical, molecular, and morphological analyses of single cells, *Anal. Bioanal. Chem.* 397 (2010) 2051–2065. doi:10.1007/s00216-010-3618-7.
- [32] R. Strick, P.L. Strissel, K. Gavrilov, R. Levi-Setti, Cation–chromatin binding as shown by ion microscopy is essential for the structural integrity of chromosomes, *J. Cell Biol.* 155 (2001) 899–910. doi:10.1083/jcb.200105026.
- [33] C. Bich, R. Havelund, R. Moellers, D. Touboul, F. Kollmer, E. Niehuis, et al., Argon Cluster Ion Source Evaluation on Lipid Standards and Rat Brain Tissue Samples, *Anal. Chem.* 85 (2013) 7745–7752. doi:10.1021/ac4009513.
- [34] Y. Magnusson, P. Friberg, P. Sjövall, F. Dangardt, P. Malmberg, Y. Chen, Lipid imaging of human skeletal muscle using TOF-SIMS with bismuth cluster ion as a primary ion source, *Clin. Physiol. Funct. Imaging.* 28 (2008) 202–209. doi:10.1111/j.1475-097X.2008.00796.x.
- [35] R.M. Caprioli, T.B. Farmer, J. Gile, Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS, *Anal. Chem.* 69 (1997) 4751–4760. doi:10.1021/ac970888i.
- [36] T.W. Powers, E.E. Jones, L.R. Betesh, P.R. Romano, P. Gao, J.A. Copland, et al., Matrix Assisted Laser Desorption Ionization Imaging Mass Spectrometry Workflow for Spatial Profiling Analysis of N-Linked Glycan Expression in Tissues, *Anal. Chem.* 85 (2013) 9799–9806. doi:10.1021/ac402108x.

- [37] A. Fülöp, M.B. Porada, C. Marsching, H. Blott, B. Meyer, S. Tambe, et al., 4-Phenyl- α -cyanocinnamic Acid Amide: Screening for a Negative Ion Matrix for MALDI-MS Imaging of Multiple Lipid Classes, *Anal. Chem.* 85 (2013) 9156–9163. doi:10.1021/ac4018154.
- [38] C. Marsching, M. Eckhardt, H.-J. Gröne, R. Sandhoff, C. Hopf, Imaging of complex sulfatides SM3 and SB1a in mouse kidney using MALDI-TOF/TOF mass spectrometry, *Anal. Bioanal. Chem.* 401 (2011) 53–64. doi:10.1007/s00216-011-4802-0.
- [39] S.J. Atkinson, P.M. Loadman, C. Sutton, L.H. Patterson, M.R. Clench, Examination of the distribution of the bioreductive drug AQ4N and its active metabolite AQ4 in solid tumours by imaging matrix-assisted laser desorption/ionisation mass spectrometry, *Rapid Commun. Mass Spectrom.* 21 (2007) 1271–1276. doi:10.1002/rcm.2952.
- [40] S. Khatib-Shahidi, M. Andersson, J.L. Herman, T.A. Gillespie, R.M. Caprioli, Direct Molecular Analysis of Whole-Body Animal Tissue Sections by Imaging MALDI Mass Spectrometry, *Anal. Chem.* 78 (2006) 6448–6456. doi:10.1021/ac060788p.
- [41] M. Lagarrigue, M. Becker, R. Lavigne, S.-O. Deininger, A. Walch, F. Aubry, et al., Revisiting Rat Spermatogenesis with MALDI Imaging at 20- μ m Resolution, *Mol. Cell. Proteomics.* 10 (2011) M110.005991. doi:10.1074/mcp.M110.005991.
- [42] Y. Schober, S. Guenther, B. Spengler, A. Römpf, Single Cell Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging, *Anal. Chem.* 84 (2012) 6293–6297. doi:10.1021/ac301337h.
- [43] M.K. Passarelli, N. Winograd, Lipid imaging with time-of-flight secondary ion mass spectrometry (ToF-SIMS), *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids.* 1811 (2011) 976–990. doi:10.1016/j.bbalip.2011.05.007.
- [44] B. Balluff, C. Schöne, H. Höfler, A. Walch, MALDI imaging mass spectrometry for direct tissue analysis: technological advancements and recent applications, *Histochem. Cell Biol.* 136 (2011) 227–244. doi:10.1007/s00418-011-0843-x.
- [45] S.R. Ellis, S.H. Brown, M. in het Panhuis, S.J. Blanksby, T.W. Mitchell, Surface analysis of lipids by mass spectrometry: More than just imaging, *Prog. Lipid Res.* 52 (2013) 329–353. doi:10.1016/j.plipres.2013.04.005.
- [46] Z. Takáts, J.M. Wiseman, B. Gologan, R.G. Cooks, Mass Spectrometry Sampling Under Ambient Conditions with Desorption Electrospray Ionization, *Science.* 306 (2004) 471–473. doi:10.1126/science.1104404.

- [47] S.R. Ellis, C. Wu, J.M. Deeley, X. Zhu, R.J.W. Truscott, M. in het Panhuis, et al., Imaging of human lens lipids by desorption electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom.* 21 (2010) 2095–2104. doi:10.1016/j.jasms.2010.09.003.
- [48] M. Girod, Y. Shi, J.-X. Cheng, R.G. Cooks, Desorption electrospray ionization imaging mass spectrometry of lipids in rat spinal cord, *J. Am. Soc. Mass Spectrom.* 21 (2010) 1177–1189. doi:10.1016/j.jasms.2010.03.028.
- [49] J. Watrous, N. Hendricks, M. Meehan, P.C. Dorrestein, Capturing Bacterial Metabolic Exchange Using Thin Film Desorption Electrospray Ionization-Imaging Mass Spectrometry, *Anal. Chem.* 82 (2010) 1598–1600. doi:10.1021/ac9027388.
- [50] L.S. Eberlin, R. Haddad, R.C.S. Neto, R.G. Cosso, D.R.J. Maia, A.O. Maldaner, et al., Instantaneous chemical profiles of banknotes by ambient mass spectrometry, *Analyst.* 135 (2010) 2533–2539. doi:10.1039/C0AN00243G.
- [51] The UniProt Consortium, Update on activities at the Universal Protein Resource (UniProt) in 2013, *Nucleic Acids Res.* 41 (2012) D43–D47. doi:10.1093/nar/gks1068.
- [52] L.M. Smith, N.L. Kelleher, T.C. for T.D. Proteomics, Proteoform: a single term describing protein complexity, *Nat. Methods.* 10 (2013) 186–187. doi:10.1038/nmeth.2369.
- [53] M.R. Wilkins, C. Pasquali, R.D. Appel, K. Ou, O. Golaz, J.C. Sanchez, et al., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis, *Biotechnol. Nat. Publ. Co.* 14 (1996) 61–65.
- [54] N.L. Anderson, N.G. Anderson, Proteome and proteomics: new technologies, new concepts, and new words, *Electrophoresis.* 19 (1998) 1853–1861. doi:10.1002/elps.1150191103.
- [55] B. Metz, G.F.A. Kersten, P. Hoogerhout, H.F. Brugghe, H.A.M. Timmermans, A. de Jong, et al., Identification of Formaldehyde-induced Modifications in Proteins REACTIONS WITH MODEL PEPTIDES, *J. Biol. Chem.* 279 (2004) 6235–6243. doi:10.1074/jbc.M310752200.
- [56] M. Shabihkhani, G.M. Lucey, B. Wei, S. Mareninov, J.J. Lou, H.V. Vinters, et al., The procurement, storage, and quality assurance of frozen blood and tissue biospecimens in pathology, biorepository, and biobank settings, *Clin. Biochem.* (n.d.). doi:10.1016/j.clinbiochem.2014.01.002.
- [57] B. Ergin, S. Meding, R. Langer, M. Kap, C. Viertler, C. Schott, et al., Proteomic Analysis of PAXgene-Fixed Tissues, *J. Proteome Res.* 9 (2010) 5188–5196. doi:10.1021/pr100664e.

- [58] S.A. Schwartz, M.L. Reyzer, R.M. Caprioli, Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation, *J. Mass Spectrom.* 38 (2003) 699–708. doi:10.1002/jms.505.
- [59] E.A. Jones, A. van Remoortere, R.J.M. van Zeijl, P.C.W. Hogendoorn, J.V.M.G. Bovée, A.M. Deelder, et al., Multiple Statistical Analysis Techniques Corroborate Intratumor Heterogeneity in Imaging Mass Spectrometry Datasets of Myxofibrosarcoma, *PLoS ONE*. 6 (2011) e24913. doi:10.1371/journal.pone.0024913.
- [60] A.M. Delvolve, A.S. Woods, Optimization of automated matrix deposition for biomolecular mapping using a spotter, *J. Mass Spectrom. JMS*. 46 (2011) 1046–1050. doi:10.1002/jms.1986.
- [61] M. Schuerenberg, C. Luebbert, S.-O. Deininger, R. Ketterlinus, D. Suckau, MALDI tissue imaging: mass spectrometric localization of biomarkers in tissue slices, *Nat. Methods*. 4 (2007). doi:10.1038/nmeth1039.
- [62] M.-C. Djidja, S. Francese, P.M. Loadman, C.W. Sutton, P. Scriven, E. Claude, et al., Detergent addition to tryptic digests and ion mobility separation prior to MS/MS improves peptide yield and protein identification for in situ proteomic investigation of frozen and formalin-fixed paraffin-embedded adenocarcinoma tissue sections, *PROTEOMICS*. 9 (2009) 2750–2763. doi:10.1002/pmic.200800624.
- [63] J.A. Hankin, R.M. Barkley, R.C. Murphy, Sublimation as a method of matrix application for mass spectrometric imaging, *J. Am. Soc. Mass Spectrom.* 18 (2007) 1646–1652. doi:10.1016/j.jasms.2007.06.010.
- [64] J.J. Nicklay, G.A. Harris, K.L. Schey, R.M. Caprioli, MALDI Imaging and in Situ Identification of Integral Membrane Proteins from Rat Brain Tissue Sections, *Anal. Chem.* 85 (2013) 7191–7196. doi:10.1021/ac400902h.
- [65] J. Franck, R. Longuespée, M. Wisztorski, A. Van Remoortere, R. Van Zeijl, A. Deelder, et al., MALDI mass spectrometry imaging of proteins exceeding 30,000 daltons, *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* 16 (2010) BR293–299.
- [66] S.R. Ellis, J.H. Jungmann, D.F. Smith, J. Soltwisch, R.M.A. Heeren, Enhanced Detection of High-Mass Proteins by Using an Active Pixel Detector, *Angew. Chem. Int. Ed.* 52 (2013) 11261–11264. doi:10.1002/anie.201305501.
- [67] K. Schwamborn, R.M. Caprioli, Molecular imaging by mass spectrometry — looking beyond classical histology, *Nat. Rev. Cancer*. 10 (2010) 639–646. doi:10.1038/nrc2917.

- [68] A. Turtoi, A. Blomme, D. Debois, J. Somja, D. Delvaux, G. Patsos, et al., Organized proteomic heterogeneity in colorectal cancer liver metastases and implications for therapies, *Hepatology*. (2013) n/a–n/a. doi:10.1002/hep.26608.
- [69] J.L. Norris, R.M. Caprioli, Analysis of Tissue Specimens by Matrix-Assisted Laser Desorption/Ionization Imaging Mass Spectrometry in Biological and Clinical Research, *Chem. Rev.* 113 (2013) 2309–2342. doi:10.1021/cr3004295.
- [70] T. Alexandrov, MALDI imaging mass spectrometry: statistical data analysis and current computational challenges, *BMC Bioinformatics*. 13 (2012) S11. doi:10.1186/1471-2105-13-S16-S11.
- [71] D.N. Stewart, J. Lango, K.P. Nambiar, M.J.S. Falso, P.G. FitzGerald, D.M. Rocke, et al., Carbon turnover in the water-soluble protein of the adult human lens, *Mol. Vis.* 19 (2013) 463–475.
- [72] A.C. Grey, K.L. Schey, Distribution of bovine and rabbit lens β -crystallin products by MALDI imaging mass spectrometry, *Mol. Vis.* 14 (2008) 171–179.
- [73] A.C. Grey, K.L. Schey, Age-Related Changes in the Spatial Distribution of Human Lens α -Crystallin Products by MALDI Imaging Mass Spectrometry, *Invest. Ophthalmol. Vis. Sci.* 50 (2009) 4319–4329. doi:10.1167/iovs.09-3522.
- [74] J. Han, K.L. Schey, MALDI Tissue Imaging of Ocular Lens α -Crystallin, *Invest. Ophthalmol. Vis. Sci.* 47 (2006) 2990–2996. doi:10.1167/iovs.05-1529.
- [75] D.R. Stella, K.A. Floyd, A.C. Grey, M.B. Renfrow, K.L. Schey, S. Barnes, Tissue Localization and Solubilities of α A-crystallin and its Numerous C-terminal Truncation Products in Pre- and Postcataractous ICR/f Rat Lenses, *Invest. Ophthalmol. Vis. Sci.* 51 (2010) 5153–5161. doi:10.1167/iovs.10-5302.
- [76] S.-P. Su, J.D. McArthur, J. Andrew Aquilina, Localization of low molecular weight crystallin peptides in the aging human lens using a MALDI mass spectrometry imaging approach, *Exp. Eye Res.* 91 (2010) 97–103. doi:10.1016/j.exer.2010.04.010.
- [77] A.C. Grey, P. Chaurand, R.M. Caprioli, K.L. Schey, MALDI Imaging Mass Spectrometry of Integral Membrane Proteins from Ocular Lens and Retinal Tissue†, *J. Proteome Res.* 8 (2009) 3278–3283. doi:10.1021/pr800956y.
- [78] M. Ronci, S. Sharma, T. Chataway, K.P. Burdon, S. Martin, J.E. Craig, et al., MALDI-MS-Imaging of Whole Human Lens Capsule, *J. Proteome Res.* 10 (2011) 3522–3529. doi:10.1021/pr200148k.

- [79] D.B. Thibault, C.J. Gillam, A.C. Grey, J. Han, K.L. Schey, MALDI tissue profiling of integral membrane proteins from ocular tissues, *J. Am. Soc. Mass Spectrom.* 19 (2008) 814–822. doi:10.1016/j.jasms.2008.03.002.
- [80] M. Stoeckli, P. Chaurand, D.E. Hallahan, R.M. Caprioli, Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues, *Nat. Med.* 7 (2001) 493–496. doi:10.1038/86573.
- [81] S. Rauser, C. Marquardt, B. Balluff, S.-O. Deininger, C. Albers, E. Belau, et al., Classification of HER2 Receptor Status in Breast Cancer Tissues by MALDI Imaging Mass Spectrometry, *J. Proteome Res.* 9 (2010) 1854–1863. doi:10.1021/pr901008d.
- [82] B. Balluff, M. Elsner, A. Kowarsch, S. Rauser, S. Meding, C. Schuhmacher, et al., Classification of HER2/neu Status in Gastric Cancer Using a Breast-Cancer Derived Proteome Classifier, *J. Proteome Res.* 9 (2010) 6317–6322. doi:10.1021/pr100573s.
- [83] M. Elsner, S. Rauser, S. Maier, C. Schöne, B. Balluff, S. Meding, et al., MALDI imaging mass spectrometry reveals COX7A2, TAGLN2 and S100-A10 as novel prognostic markers in Barrett’s adenocarcinoma, *J. Proteomics.* 75 (2012) 4693–4704. doi:10.1016/j.jprot.2012.02.012.
- [84] M. Nipp, M. Elsner, B. Balluff, S. Meding, H. Sarioglu, M. Ueffing, et al., S100-A10, thioredoxin, and S100-A6 as biomarkers of papillary thyroid carcinoma with lymph node metastasis identified by MALDI Imaging, *J. Mol. Med.* 90 (2012) 163–174. doi:10.1007/s00109-011-0815-6.
- [85] S. Meding, U. Nitsche, B. Balluff, M. Elsner, S. Rauser, C. Schöne, et al., Tumor Classification of Six Common Cancer Types Based on Proteomic Profiling by MALDI Imaging, *J. Proteome Res.* 11 (2012) 1996–2003. doi:10.1021/pr200784p.
- [86] J. Pierson, J.L. Norris, H.-R. Aerni, P. Svenningsson, R.M. Caprioli, P.E. Andrén, Molecular Profiling of Experimental Parkinson’s Disease: Direct Analysis of Peptides and Proteins on Brain Tissue Sections by MALDI Mass Spectrometry, *J. Proteome Res.* 3 (2004) 289–295. doi:10.1021/pr0499747.
- [87] R. Ceuppens, D. Dumont, L. Van Brussel, B. Van de Plas, R. Daniels, J.-P. Noben, et al., Direct profiling of myelinated and demyelinated regions in mouse brain by imaging mass spectrometry, *Int. J. Mass Spectrom.* 260 (2007) 185–194. doi:10.1016/j.ijms.2006.09.007.
- [88] R. Kruse, J.V. Sweedler, Spatial profiling invertebrate ganglia using MALDI MS, *J. Am. Soc. Mass Spectrom.* 14 (2003) 752–759. doi:10.1016/S1044-0305(03)00288-5.

- [89] M.R. Groseclose, P.P. Massion, P. Chaurand, R.M. Caprioli, High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry, *PROTEOMICS*. 8 (2008) 3715–3724. doi:10.1002/pmic.200800495.
- [90] M.-C. Djidja, E. Claude, M.F. Snel, S. Francese, P. Scriven, V. Carolan, et al., Novel molecular tumour classification using MALDI–mass spectrometry imaging of tissue micro-array, *Anal. Bioanal. Chem.* 397 (2010) 587–601. doi:10.1007/s00216-010-3554-6.
- [91] R. Lazova, E.H. Seeley, M. Keenan, R. Gueorguieva, R.M. Caprioli, Imaging Mass Spectrometry - a new and promising method to differentiate Spitz nevi from Spitzoid malignant melanomas, *Am. J. Dermatopathol.* 34 (2012) 82–90. doi:10.1097/DAD.0b013e31823df1e2.
- [92] K. Sogawa, M. Watanabe, K. Sato, S. Segawa, C. Ishii, A. Miyabe, et al., Use of the MALDI BioTyper system with MALDI–TOF mass spectrometry for rapid identification of microorganisms, *Anal. Bioanal. Chem.* 400 (2011) 1905–1911. doi:10.1007/s00216-011-4877-7.
- [93] K. Mullis, F. Faloon, S. Scharf, R. Saiki, G. Horn, H. Erlich, Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction, *Cold Spring Harb. Symp. Quant. Biol.* 51 Pt 1 (1986) 263–273.
- [94] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, et al., Initial sequencing and analysis of the human genome, *Nature*. 409 (2001) 860–921. doi:10.1038/35057062.
- [95] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, et al., The sequence of the human genome, *Science*. 291 (2001) 1304–1351. doi:10.1126/science.1058040.
- [96] R. Aebersold, Constellations in a cellular universe, *Nature*. 422 (2003) 115–116. doi:10.1038/422115a.
- [97] P.R. Jungblut, H.G. Holzhütter, R. Apweiler, H. Schlüter, The speciation of the proteome, *Chem. Cent. J.* 2 (2008) 16. doi:10.1186/1752-153X-2-16.
- [98] K. Vékey, Multiply charged ions, *Mass Spectrom. Rev.* 14 (1995) 195–225. doi:10.1002/mas.1280140304.
- [99] B.Q. Tran, C. Hernandez, P. Waridel, A. Potts, J. Barblan, F. Lisacek, et al., Addressing Trypsin Bias in Large Scale (Phospho)proteome Analysis by Size Exclusion Chromatography and Secondary Digestion of Large Post-Trypsin Peptides, *J. Proteome Res.* 10 (2011) 800–811. doi:10.1021/pr100951t.

- [100] A. Michalski, J. Cox, M. Mann, More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC-MS/MS, *J. Proteome Res.* 10 (2011) 1785–1793. doi:10.1021/pr101060v.
- [101] N. Delmotte, M. Lasaosa, A. Tholey, E. Heinzle, C.G. Huber, Two-Dimensional Reversed-Phase × Ion-Pair Reversed-Phase HPLC: An Alternative Approach to High-Resolution Peptide Separation for Shotgun Proteome Analysis, *J. Proteome Res.* 6 (2007) 4363–4373. doi:10.1021/pr070424t.
- [102] V. Spicer, A. Yamchuk, J. Cortens, S. Sousa, W. Ens, K.G. Standing, et al., Sequence-Specific Retention Calculator. A Family of Peptide Retention Time Prediction Algorithms in Reversed-Phase HPLC: Applicability to Various Chromatographic Conditions and Columns, *Anal. Chem.* 79 (2007) 8762–8768. doi:10.1021/ac071474k.
- [103] M. Schirle, M.-A. Heurtier, B. Kuster, Profiling Core Proteomes of Human Cell Lines by One-dimensional PAGE and Liquid Chromatography-Tandem Mass Spectrometry, *Mol. Cell. Proteomics.* 2 (2003) 1297–1305. doi:10.1074/mcp.M300087-MCP200.
- [104] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nat. Methods.* 6 (2009) 359–362. doi:10.1038/nmeth.1322.
- [105] U. Kruse, C.P. Pallasch, M. Bantscheff, D. Eberhard, L. Frenzel, S. Ghidelli, et al., Chemoproteomics-based kinome profiling and target deconvolution of clinical multi-kinase inhibitors in primary chronic lymphocytic leukemia cells, *Leukemia.* 25 (2011) 89–100. doi:10.1038/leu.2010.233.
- [106] M. Mann, P. Højrup, P. Roepstorff, Use of mass spectrometric molecular weight information to identify proteins in sequence databases, *Biol. Mass Spectrom.* 22 (1993) 338–345. doi:10.1002/bms.1200220605.
- [107] D.J.C. Pappin, P. Hojrup, A.J. Bleasby, Rapid identification of proteins by peptide-mass fingerprinting, *Curr. Biol.* 3 (1993) 327–332. doi:10.1016/0960-9822(93)90195-T.
- [108] T. Rabilloud, M. Chevallet, S. Luche, C. Lelong, Two-dimensional gel electrophoresis in proteomics: Past, present and future, *J. Proteomics.* 73 (2010) 2064–2077. doi:10.1016/j.jprot.2010.05.016.
- [109] P. Roepstorff, J. Fohlman, Letter to the editors, *Biol. Mass Spectrom.* 11 (1984) 601–601. doi:10.1002/bms.1200111109.
- [110] B. Ma, R. Johnson, De Novo Sequencing and Homology Searching, *Mol. Cell. Proteomics.* 11 (2012) O111.014902. doi:10.1074/mcp.O111.014902.

- [111] J.E.P. Syka, J.J. Coon, M.J. Schroeder, J. Shabanowitz, D.F. Hunt, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 9528–9533. doi:10.1073/pnas.0402700101.
- [112] R.A. Zubarev, N.L. Kelleher, F.W. McLafferty, Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process, *J. Am. Chem. Soc.* 120 (1998) 3265–3266. doi:10.1021/ja973478k.
- [113] A.S. Hebert, A.L. Richards, D.J. Bailey, A. Ulbrich, E.E. Coughlin, M.S. Westphall, et al., The One Hour Yeast Proteome, *Mol. Cell. Proteomics.* (2013) mcp.M113.034769. doi:10.1074/mcp.M113.034769.
- [114] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Accurate quantitation of protein expression and site-specific phosphorylation, *Proc. Natl. Acad. Sci.* 96 (1999) 6591–6596. doi:10.1073/pnas.96.12.6591.
- [115] L.R. Zieske, A perspective on the use of iTRAQ™ reagent technology for protein complex and profiling studies, *J. Exp. Bot.* 57 (2006) 1501–1508. doi:10.1093/jxb/erj168.
- [116] A. Römpf, B. Spengler, Mass spectrometry imaging with high resolution in mass and space, *Histochem. Cell Biol.* 139 (2013) 759–783. doi:10.1007/s00418-013-1097-6.
- [117] J.O.R. Gustafsson, J.S. Eddes, S. Meding, T. Koudelka, M.K. Oehler, S.R. McColl, et al., Internal calibrants allow high accuracy peptide matching between MALDI imaging MS and LC-MS/MS, *J. Proteomics.* 75 (2012) 5093–5105. doi:10.1016/j.jprot.2012.04.054.
- [118] A. Michalski, E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Müller, et al., Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes, *Mol. Cell. Proteomics.* 11 (2012) O111.013698. doi:10.1074/mcp.O111.013698.
- [119] G.K. Taylor, Y.-B. Kim, A.J. Forbes, F. Meng, R. McCarthy, N.L. Kelleher, Web and Database Software for Identification of Intact Proteins Using “Top Down” Mass Spectrometry, *Anal. Chem.* 75 (2003) 4081–4086. doi:10.1021/ac0341721.
- [120] X. Liu, Y. Sirotkin, Y. Shen, G. Anderson, Y.S. Tsai, Y.S. Ting, et al., Protein Identification Using Top-Down Spectra, *Mol. Cell. Proteomics.* 11 (2012) M111.008524. doi:10.1074/mcp.M111.008524.
- [121] Y. Yang, A. Barendregt, J.P. Kamerling, A.J.R. Heck, Analyzing Protein Micro-Heterogeneity in Chicken Ovalbumin by High-Resolution Native Mass Spectrometry Exposes Qualitatively and Semi-Quantitatively 59 Proteoforms, *Anal. Chem.* 85 (2013) 12037–12045. doi:10.1021/ac403057y.

- [122] J.C. Tran, L. Zamdborg, D.R. Ahlf, J.E. Lee, A.D. Catherman, K.R. Durbin, et al., Mapping intact protein isoforms in discovery mode using top-down proteomics, *Nature*. 480 (2011) 254–258. doi:10.1038/nature10575.
- [123] R. Ait-Belkacem, L. Sellami, C. Villard, E. DePauw, D. Calligaris, D. Lafitte, Mass spectrometry imaging is moving toward drug protein co-localization, *Trends Biotechnol.* 30 (2012) 466–474. doi:10.1016/j.tibtech.2012.05.006.
- [124] E.A. Jones, N. Schmitz, C.J.F. Waaijer, C.K. Frese, A. van Remoortere, R.J.M. van Zeijl, et al., Imaging Mass Spectrometry-based Molecular Histology Differentiates Microscopically Identical and Heterogeneous Tumors, *J. Proteome Res.* 12 (2013) 1847–1855. doi:10.1021/pr301190g.
- [125] K.L. Schey, D.M. Anderson, K.L. Rose, Spatially-Directed Protein Identification from Tissue Sections by Top-Down LC-MS/MS with Electron Transfer Dissociation, *Anal. Chem.* 85 (2013) 6767–6774. doi:10.1021/ac400832w.
- [126] B. Balluff, C. Schöne, H. Höfler, A. Walch, MALDI imaging mass spectrometry for direct tissue analysis: technological advancements and recent applications, *Histochem. Cell Biol.* 136 (2011) 227–244. doi:10.1007/s00418-011-0843-x.
- [127] R.M. Caprioli, T.B. Farmer, J. Gile, Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS, *Anal. Chem.* 69 (1997) 4751–4760.
- [128] A. Römpf, S. Guenther, Z. Takats, B. Spengler, Mass spectrometry imaging with high resolution in mass and space (HR(2) MSI) for reliable investigation of drug compound distributions on the cellular level, *Anal. Bioanal. Chem.* 401 (2011) 65–73. doi:10.1007/s00216-011-4990-7.
- [129] L. Signor, E. Varesio, R.F. Staack, V. Starke, W.F. Richter, G. Hopfgartner, Analysis of erlotinib and its metabolites in rat tissue sections by MALDI quadrupole time-of-flight mass spectrometry, *J. Mass Spectrom. JMS.* 42 (2007) 900–909. doi:10.1002/jms.1225.
- [130] Y. Liu, Y. Chen, A. Momin, R. Shaner, E. Wang, N.J. Bowen, et al., Elevation of sulfatides in ovarian cancer: an integrated transcriptomic and lipidomic analysis including tissue-imaging mass spectrometry, *Mol. Cancer.* 9 (2010) 186. doi:10.1186/1476-4598-9-186.
- [131] P. Chaurand, J.C. Latham, K.B. Lane, J.A. Mobley, V.V. Polosukhin, P.S. Wirth, et al., Imaging mass spectrometry of intact proteins from alcohol-preserved tissue specimens: bypassing formalin fixation, *J. Proteome Res.* 7 (2008) 3543–3555. doi:10.1021/pr800286z.

- [132] Y. Schober, S. Guenther, B. Spengler, A. Römpf, High-resolution matrix-assisted laser desorption/ionization imaging of tryptic peptides from tissue, *Rapid Commun. Mass Spectrom. RCM*. 26 (2012) 1141–1146. doi:10.1002/rcm.6192.
- [133] M. Wisztorski, J. Franck, M. Salzet, I. Fournier, MALDI direct analysis and imaging of frozen versus FFPE tissues: what strategy for which sample?, *Methods Mol. Biol. Clifton NJ*. 656 (2010) 303–322. doi:10.1007/978-1-60761-746-4_18.
- [134] F. Deutskens, J. Yang, R.M. Caprioli, High spatial resolution imaging mass spectrometry and classical histology on a single tissue section, *J. Mass Spectrom. JMS*. 46 (2011) 568–571. doi:10.1002/jms.1926.
- [135] E.H. Seeley, K. Schwamborn, R.M. Caprioli, Imaging of intact tissue sections: moving beyond the microscope, *J. Biol. Chem.* 286 (2011) 25459–25466. doi:10.1074/jbc.R111.225854.
- [136] B. Balluff, S. Rauser, S. Meding, M. Elsner, C. Schöne, A. Feuchtinger, et al., MALDI imaging identifies prognostic seven-protein signature of novel tissue markers in intestinal-type gastric cancer, *Am. J. Pathol.* 179 (2011) 2720–2729. doi:10.1016/j.ajpath.2011.08.032.
- [137] S.A. Schwartz, R.J. Weil, M.D. Johnson, S.A. Toms, R.M. Caprioli, Protein profiling in brain tumors using mass spectrometry: feasibility of a new technique for the analysis of protein expression, *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 10 (2004) 981–987.
- [138] J.A. Bauer, A.B. Chakravarthy, J.M. Rosenbluth, D. Mi, E.H. Seeley, N. De Matos Granja-Ingram, et al., Identification of markers of taxane sensitivity using proteomic and genomic analyses of breast tumors from patients receiving neoadjuvant paclitaxel and radiation, *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 16 (2010) 681–690. doi:10.1158/1078-0432.CCR-09-1091.
- [139] S.R. Oppenheimer, D. Mi, M.E. Sanders, R.M. Caprioli, Molecular analysis of tumor margins by MALDI mass spectrometry in renal carcinoma, *J. Proteome Res.* 9 (2010) 2182–2190. doi:10.1021/pr900936z.
- [140] K. Schwamborn, R.C. Krieg, M. Reska, G. Jakse, R. Knuechel, A. Wellmann, Identifying prostate carcinoma by MALDI-Imaging, *Int. J. Mol. Med.* 20 (2007) 155–159.
- [141] W.M. Hardesty, M.C. Kelley, D. Mi, R.L. Low, R.M. Caprioli, Protein signatures for survival and recurrence in metastatic melanoma, *J. Proteomics.* 74 (2011) 1002–1014. doi:10.1016/j.jprot.2011.04.013.

- [142] A.E. M'Koma, E.H. Seeley, M.K. Washington, D.A. Schwartz, R.L. Muldoon, A.J. Herline, et al., Proteomic profiling of mucosal and submucosal colonic tissues yields protein signatures that differentiate the inflammatory colitides, *Inflamm. Bowel Dis.* 17 (2011) 875–883. doi:10.1002/ibd.21442.
- [143] K.E. Burnum, S. Tranguch, D. Mi, T. Daikoku, S.K. Dey, R.M. Caprioli, Imaging mass spectrometry reveals unique protein profiles during embryo implantation, *Endocrinology.* 149 (2008) 3274–3278. doi:10.1210/en.2008-0309.
- [144] M. Stoeckli, D. Staab, M. Staufenbiel, K.-H. Wiederhold, L. Signor, Molecular imaging of amyloid beta peptides in mouse brain sections using mass spectrometry, *Anal. Biochem.* 311 (2002) 33–39.
- [145] J. Hanrieder, T. Ekegren, M. Andersson, J. Bergquist, MALDI Imaging of Post Mortem Human Spinal Cord in Amyotrophic Lateral Sclerosis, *J. Neurochem.* (2012). doi:10.1111/jnc.12019.
- [146] L.H. Cazares, D. Troyer, S. Mendrinios, R.A. Lance, J.O. Nyalwidhe, H.A. Beydoun, et al., Imaging mass spectrometry of a specific fragment of mitogen-activated protein kinase/extracellular signal-regulated kinase kinase 2 discriminates cancer from uninvolved prostate tissue, *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 15 (2009) 5541–5551. doi:10.1158/1078-0432.CCR-08-2892.
- [147] M. Elsner, S. Rauser, S. Maier, C. Schöne, B. Balluff, S. Meding, et al., MALDI imaging mass spectrometry reveals COX7A2, TAGLN2 and S100-A10 as novel prognostic markers in Barrett's adenocarcinoma, *J. Proteomics.* (2012). doi:10.1016/j.jprot.2012.02.012.
- [148] H.K. Kim, M.L. Reyzer, I.J. Choi, C.G. Kim, H.S. Kim, A. Oshima, et al., Gastric cancer-specific protein profile identified using endoscopic biopsy samples via MALDI mass spectrometry, *J. Proteome Res.* 9 (2010) 4123–4130. doi:10.1021/pr100302b.
- [149] R. Lemaire, S.A. Menguellat, J. Stauber, V. Marchaudon, J.-P. Lucot, P. Collinet, et al., Specific MALDI imaging and profiling for biomarker hunting and validation: fragment of the 11S proteasome activator complex, Reg alpha fragment, is a new potential ovary cancer biomarker, *J. Proteome Res.* 6 (2007) 4127–4134. doi:10.1021/pr0702722.
- [150] S. Meding, B. Balluff, M. Elsner, C. Schöne, S. Rauser, U. Nitsche, et al., Tissue Based Proteomics Reveals FXYD3, S100A11 and GSTM3 as Novel Markers for Regional Lymph Node Metastasis in Colon Cancer, *J. Pathol.* (2012). doi:10.1002/path.4021.
- [151] M. Nipp, M. Elsner, B. Balluff, S. Meding, H. Sarioglu, M. Ueffing, et al., S100-A10, thioredoxin, and S100-A6 as biomarkers of papillary thyroid carcinoma with lymph node

- metastasis identified by MALDI imaging, *J. Mol. Med. Berl. Ger.* 90 (2012) 163–174. doi:10.1007/s00109-011-0815-6.
- [152] S.M.J. Rahman, Y. Shyr, P.B. Yildiz, A.L. Gonzalez, H. Li, X. Zhang, et al., Proteomic patterns of preinvasive bronchial lesions, *Am. J. Respir. Crit. Care Med.* 172 (2005) 1556–1562. doi:10.1164/rccm.200502-274OC.
- [153] S. Rauser, C. Marquardt, B. Balluff, S.-O. Deininger, C. Albers, E. Belau, et al., Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry, *J. Proteome Res.* 9 (2010) 1854–1863. doi:10.1021/pr901008d.
- [154] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, *Anal. Chem.* 68 (1996) 850–858.
- [155] B.C. Searle, Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies, *Proteomics.* 10 (2010) 1265–1269. doi:10.1002/pmic.200900437.
- [156] M. Schirle, M.-A. Heurtier, B. Kuster, Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry, *Mol. Cell. Proteomics MCP.* 2 (2003) 1297–1305. doi:10.1074/mcp.M300087-MCP200.
- [157] A.C. Grey, P. Chaurand, R.M. Caprioli, K.L. Schey, MALDI imaging mass spectrometry of integral membrane proteins from ocular lens and retinal tissue, *J. Proteome Res.* 8 (2009) 3278–3283. doi:10.1021/pr800956y.
- [158] J.R. Yates, N.L. Kelleher, Top Down Proteomics, *Anal. Chem.* 85 (2013) 6151–6151. doi:10.1021/ac401484r.
- [159] E. Stimson, O. Truong, W.J. Richter, M.D. Waterfield, A.L. Burlingame, Enhancement of charge remote fragmentation in protonated peptides by high-energy CID MALDI-TOF-MS using “cold” matrices, *Int. J. Mass Spectrom. Ion Process.* 169–170 (1997) 231–240. doi:10.1016/S0168-1176(97)00227-9.
- [160] P. Chaurand, F. Luetzenkirchen, B. Spengler, Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and MALDI-post-source decay time-of-flight mass spectrometry, *J. Am. Soc. Mass Spectrom.* 10 (1999) 91–103. doi:10.1016/S1044-0305(98)00145-7.
- [161] R. Ait-Belkacem, C. Berenguer, C. Villard, L. Ouafik, D. Figarella-Branger, O. Chinot, et al., MALDI imaging and in source decay for top down characterization of glioblastoma, *PROTEOMICS.* (2013) n/a–n/a. doi:10.1002/pmic.201300329.

- [162] K.E. Burnum, S. Tranguch, D. Mi, T. Daikoku, S.K. Dey, R.M. Caprioli, Imaging mass spectrometry reveals unique protein profiles during embryo implantation, *Endocrinology*. 149 (2008) 3274–3278. doi:10.1210/en.2008-0309.
- [163] M. Lagarrigue, T. Alexandrov, G. Dieuset, A. Perrin, R. Lavigne, S. Baulac, et al., A new analysis workflow for MALDI imaging mass spectrometry: application to the discovery and identification of potential markers of Childhood Absence Epilepsy, *J. Proteome Res.* (2012). doi:10.1021/pr3006974.
- [164] M.L. Reyzer, R.L. Caldwell, T.C. Dugger, J.T. Forbes, C.A. Ritter, M. Guix, et al., Early Changes in Protein Expression Detected by Mass Spectrometry Predict Tumor Response to Molecular Therapeutics, *Cancer Res.* 64 (2004) 9093–9100. doi:10.1158/0008-5472.CAN-04-2231.
- [165] F. Pachl, B. Ruprecht, S. Lemeer, B. Kuster, Characterization of a high field Orbitrap mass spectrometer for proteome analysis, *PROTEOMICS*. 13 (2013) 2552–2562. doi:10.1002/pmic.201300076.
- [166] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, *Nat. Protoc.* 2 (2007) 1896–1906. doi:10.1038/nprot.2007.261.
- [167] N. Neuhauser, A. Michalski, J. Cox, M. Mann, Expert System for Computer-assisted Annotation of MS/MS Spectra, *Mol. Cell. Proteomics*. 11 (2012) 1500–1509. doi:10.1074/mcp.M112.020271.
- [168] K.R. Durbin, R.T. Fellers, I. Ntai, N.L. Kelleher, P.D. Compton, Autopilot: An Online Data Acquisition Control System for the Enhanced High-Throughput Characterization of Intact Proteins, *Anal. Chem.* (2014). doi:10.1021/ac402904h.
- [169] P.J. Kersey, J. Duarte, A. Williams, Y. Karavidopoulou, E. Birney, R. Apweiler, The International Protein Index: an integrated database for proteomics experiments, *Proteomics*. 4 (2004) 1985–1988. doi:10.1002/pmic.200300721.
- [170] X. Liu, Y. Inbar, P.C. Dorrestein, C. Wynne, N. Edwards, P. Souda, et al., Deconvolution and Database Search of Complex Tandem Mass Spectra of Intact Proteins A COMBINATORIAL APPROACH, *Mol. Cell. Proteomics*. 9 (2010) 2772–2782. doi:10.1074/mcp.M110.002766.
- [171] J.V. Olsen, J.C. Schwartz, J. Griep-Raming, M.L. Nielsen, E. Damoc, E. Denisov, et al., A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed, *Mol. Cell. Proteomics*. 8 (2009) 2759–2769. doi:10.1074/mcp.M900375-MCP200.

- [172] L.H. Cazares, D. Troyer, S. Mendrinós, R.A. Lance, J.O. Nyalwidhe, H.A. Beydoun, et al., Imaging Mass Spectrometry of a Specific Fragment of Mitogen-Activated Protein Kinase/Extracellular Signal-Regulated Kinase Kinase Kinase 2 Discriminates Cancer from Uninvolved Prostate Tissue, *Clin. Cancer Res.* 15 (2009) 5541–5551. doi:10.1158/1078-0432.CCR-08-2892.
- [173] W.M. van Weerden, J.C. Romijn, Use of nude mouse xenograft models in prostate cancer research, *The Prostate.* 43 (2000) 263–271. doi:10.1002/1097-0045(20000601)43:4<263::AID-PROS5>3.0.CO;2-I.
- [174] H. Hahne, F. Pachi, B. Ruprecht, S.K. Maier, S. Klaeger, D. Helm, et al., DMSO enhances electrospray response, boosting sensitivity of proteomic experiments, *Nat. Methods.* 10 (2013) 989–991. doi:10.1038/nmeth.2610.
- [175] J.W. Hager, A new linear ion trap mass spectrometer, *Rapid Commun. Mass Spectrom.* 16 (2002) 512–526. doi:10.1002/rcm.607.
- [176] B. Barasa, M. Slijper, Challenges for red blood cell biomarker discovery through proteomics, *Biochim. Biophys. Acta BBA - Proteins Proteomics.* (n.d.). doi:10.1016/j.bbapap.2013.10.002.
- [177] K. Hofmann, L. Falquet, A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems, *Trends Biochem. Sci.* 26 (2001) 347–350. doi:10.1016/S0968-0004(01)01835-7.
- [178] A. Michalski, N. Neuhauser, J. Cox, M. Mann, A Systematic Investigation into the Nature of Tryptic HCD Spectra, *J. Proteome Res.* 11 (2012) 5479–5491. doi:10.1021/pr3007045.
- [179] W.M. Bodnar, R.K. Blackburn, J.M. Krise, M.A. Moseley, Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage, *J. Am. Soc. Mass Spectrom.* 14 (2003) 971–979. doi:10.1016/S1044-0305(03)00209-5.
- [180] Y. Yang, S. Zhang, K. Howe, D.B. Wilson, F. Moser, D. Irwin, et al., A Comparison of nLC-ESI-MS/MS and nLC-MALDI-MS/MS for GeLC-Based Protein Identification and iTRAQ-Based Shotgun Quantitative Proteomics, *J. Biomol. Tech. JBT.* 18 (2007) 226–237.
- [181] M. Karas, D. Bachmann, U. Bahr, F. Hillenkamp, Matrix-assisted ultraviolet laser desorption of non-volatile compounds, *Int. J. Mass Spectrom. Ion Process.* 78 (1987) 53–68. doi:10.1016/0168-1176(87)87041-6.
- [182] P.M. Angel, R.M. Caprioli, Matrix-Assisted Laser Desorption Ionization Imaging Mass Spectrometry: In Situ Molecular Mapping, *Biochemistry (Mosc.).* 52 (2013) 3818–3828. doi:10.1021/bi301519p.

- [183] H. Kamada, T. Fugmann, D. Neri, C. Roesli, Improved protein sequence coverage by on resin deglycosylation and cysteine modification for biomarker discovery, *PROTEOMICS*. 9 (2009) 783–787. doi:10.1002/pmic.200800466.
- [184] M. Bantscheff, B. Duempelfeld, B. Kuster, An improved two-step calibration method for matrix-assisted laser desorption/ionization time-of-flight mass spectra for proteomics, *Rapid Commun. Mass Spectrom. RCM*. 16 (2002) 1892–1895. doi:10.1002/rcm.798.
- [185] J. Gobom, M. Mueller, V. Egelhofer, D. Theiss, H. Lehrach, E. Nordhoff, A Calibration Method That Simplifies and Improves Accurate Determination of Peptide Molecular Masses by MALDI-TOF MS, *Anal. Chem.* 74 (2002) 3915–3923. doi:10.1021/ac011203o.
- [186] R.W. Garden, J.V. Sweedler, Heterogeneity within MALDI Samples As Revealed by Mass Spectrometric Imaging, *Anal. Chem.* 72 (2000) 30–36. doi:10.1021/ac9908997.
- [187] M.W. Duncan, H. Roder, S.W. Hunsucker, Quantitative matrix-assisted laser desorption/ionization mass spectrometry, *Brief. Funct. Genomic. Proteomic*. 7 (2008) 355–370. doi:10.1093/bfgp/eln041.
- [188] V. Egelhofer, J. Gobom, H. Seitz, P. Giavalisco, H. Lehrach, E. Nordhoff, Protein Identification by MALDI-TOF-MS Peptide Mapping: A New Strategy, *Anal. Chem.* 74 (2002) 1760–1771. doi:10.1021/ac011204g.
- [189] M. Schlapschy, U. Binder, C. Börger, I. Theobald, K. Wachinger, S. Kisling, et al., PASylation: a biological alternative to PEGylation for extending the plasma half-life of pharmaceutically active proteins, *Protein Eng. Des. Sel. PEDS*. 26 (2013) 489–501. doi:10.1093/protein/gzt023.
- [190] E. Krause, H. Wenschuh, P.R. Jungblut, The Dominance of Arginine-Containing Peptides in MALDI-Derived Tryptic Mass Fingerprints of Proteins, *Anal. Chem.* 71 (1999) 4160–4165. doi:10.1021/ac990298f.
- [191] A.-C. Gavin, M. Bösch, R. Krause, P. Grandi, M. Marzioch, A. Bauer, et al., Functional organization of the yeast proteome by systematic analysis of protein complexes, *Nature*. 415 (2002) 141–147. doi:10.1038/415141a.
- [192] R. Zubarev, M. Mann, On the proper use of mass accuracy in proteomics, *Mol. Cell. Proteomics MCP*. 6 (2007) 377–381. doi:10.1074/mcp.M600380-MCP200.
- [193] C.E. Eyers, D.M. Simpson, S.C.C. Wong, R.J. Beynon, S.J. Gaskell, QCAL—a Novel Standard for Assessing Instrument Conditions for Proteome Analysis, *J. Am. Soc. Mass Spectrom.* 19 (2008) 1275–1280. doi:10.1016/j.jasms.2008.05.019.

- [194] P. Kraft, J. Mills, E. Dratz, Mass Spectrometric Analysis of Cyanogen Bromide Fragments of Integral Membrane Proteins at the Picomole Level: Application to Rhodopsin, *Anal. Biochem.* 292 (2001) 76–86. doi:10.1006/abio.2001.5072.
- [195] B. Meyer, D.G. Papatiriu, M. Karas, 100% protein sequence coverage: a modern form of surrealism in proteomics, *Amino Acids.* 41 (2011) 291–310. doi:10.1007/s00726-010-0680-6.

Publications

- [1] H. Hahne, F. Pachi, B. Ruprecht, **S. K. Maier**, S. Klaeger, D. Helm, G. Médard, M. Wilm, S. Lemeer, and B. Kuster, "DMSO enhances electrospray response, boosting sensitivity of proteomic experiments," *Nat. Methods*, vol. 10, no. 10, pp. 989–991, Oct. 2013.
- [2] **S. K. Maier**, H. Hahne, A. M. Gholami, B. Balluff, S. Meding, C. Schoene, A. K. Walch, and B. Kuster, "Comprehensive identification of proteins from MALDI imaging," *Mol. Cell. Proteomics*, Jun. 2013.
- [3] M. Aichler, M. Elsner, N. Ludyga, A. Feuchtinger, V. Zangen, **S. K. Maier**, B. Balluff, C. Schöne, L. Hierber, H. Braselmann, S. Meding, S. Rauser, H. Zischka, M. Aubele, M. Schmitt, M. Feith, S. M. Hauck, M. Ueffing, R. Langer, B. Kuster, H. Zitzelsberger, H. Höfler, and A. K. Walch, "Clinical response to chemotherapy in oesophageal adenocarcinoma patients is linked to defects in mitochondria," *J. Pathol.*, vol. 230, no. 4, pp. 410–419, Aug. 2013.
- [4] B. M. Grüner, H. Hahne, P. K. Mazur, M. Trajkovic-Arsic, **S. Maier**, I. Esposito, E. Kalideris, C. W. Michalski, J. Kleeff, S. Rauser, R. M. Schmid, B. Küster, A. Walch, and J. T. Siveke, "MALDI imaging mass spectrometry for in situ proteomic analysis of preneoplastic lesions in pancreatic cancer," *PloS One*, vol. 7, no. 6, p. e39424, 2012.
- [5] M. Elsner, S. Rauser, **S. Maier**, C. Schöne, B. Balluff, S. Meding, G. Jung, M. Nipp, H. Sarioglu, G. Maccarrone, M. Aichler, A. Feuchtinger, R. Langer, U. Jütting, M. Feith, B. Küster, M. Ueffing, H. Zitzelsberger, H. Höfler, and A. Walch, "MALDI imaging mass spectrometry reveals COX7A2, TAGLN2 and S100-A10 as novel prognostic markers in Barrett's adenocarcinoma," *J. Proteomics*, vol. 75, no. 15, pp. 4693–4704, Aug. 2012.
- [6] U. Kruse, C. P. Pallasch, M. Bantscheff, D. Eberhard, L. Frenzel, S. Ghidelli, **S. K. Maier**, T. Werner, C. M. Wendtner, and G. Drewes, "Chemoproteomics-based kinome profiling and target deconvolution of clinical multi-kinase inhibitors in primary chronic lymphocytic leukemia cells," *Leukemia*, vol. 25, no. 1, pp. 89–100, Jan. 2011.

Acknowledgments

Ja, das ist meine Dissertation.

Ohne die mannigfaltige Unterstützung vieler hätte es diese Arbeit mit an Sicherheit grenzender Wahrscheinlichkeit jedoch nie gegeben. Was nun folgt, kann nur ein unvollständiger Versuch bleiben, mich bei all jenen zu bedanken, die auf unterschiedlichste Art und Weise dazu beigetragen haben, dass sie doch entstanden ist.

Zuallererst möchte ich mich bei meinen beiden Betreuern Axel und Bernhard bedanken. Nicht nur weil beide diese Arbeit erst ermöglicht haben, sondern vor allem weil sie mich stets ideenreich und tatkräftig bei der erfolgreichen Umsetzung unterstützt haben.

Mein Dank gilt Prof. Dr. Martin Kingenspor, der sich bereit erklärt hat, den Prüfungsvorsitz zu übernehmen.

Prof. Dr. Carsten Hopf hat vor langer Zeit durch seine hervorragende Vorlesung mein Interesse an Proteomics geweckt. Ihm gilt mein Dank sowie allen Cellzomern insbesondere Marcus und Bea, die bis heute wertvolle Ansprechpartner in Sachen MS und darüber hinaus geblieben sind.

Meine Kollegen am Helmholtz Zentrum haben mich in die bildgebende Massenspektromie eingeführt, mit Proben versorgt und immer unterstützt und willkommen geheißen. Danke Annette, Andreas, Benni, Cédrik, Claudia-Mareike, Katharina, Mareike, Michaela, Patrick, Sandra, Sonja, Stephan und Ulrike!

Zu Dank bin ich den Küsters verpflichtet. Neben all den anregenden Diskussionen, wertvollen Hinweisen und der großen Hilfsbereitschaft, hat jeder einzelne dazu beigetragen, dass die letzten 3 Jahre, 8 Monate und 12 Tage produktiv und lehrreich waren. Es war vor allem auch eine sehr schöne Zeit mit Euch. Danke Zhixiang, Xin, Susan, Stephan, Steffi, Simone, Silvia, Sandra, Michaela, Matthias, Mathias, Martin, Kati, Karl, Jonas, Huichao, Heiner, Harald, Hannes, Guillaume, Gabi, Fiona, Estela, Elke, Elena, Dominic, Christophe, Claudia, Chen, Ben, Barbara, Andrea, Andi und Amin.

Auch, und besonders möchte ich mich bei denen bedanken, für die das, was ich da in den letzten Jahren getrieben habe, vermutlich ähnlich relevant ist wie der sprichwörtliche Sack Reis. Ich habe euch oft versetzt, und trotzdem denkt ihr an mich, geht ans Telefon und öffnet mir die

Tür! Die Sache mit der Tür gilt natürlich vor allem für meine Mitbewohner die wohl mit am schnellsten mitbekommen haben, ob ein Experiment geklappt hat oder schief gegangen ist. Vielen Dank liebe Freunde! Vielen Dank David, Elli, Holger, Caro, Alex, Boris, Thomas, Claudia, Flo, Julia, Anja und Michael.

Und zu guter Letzt und nicht nur weil man das eben so macht möchte ich mich bei meinen Eltern bedanken! Je älter ich werde, desto mehr weiß ich zu schätzen, was Ihr mir alles ermöglicht habt! Vielen Dank!

Curriculum Vitae

Stefan Karl Maier

Date of birth 13.08.1983
 Place of birth Leutkirch im Allgäu

Professional Experience

Since 2010/07 PhD student at the Institute of Analytical Pathology (with Prof. Axel Walch, Helmholtz Zentrum München) in cooperation with the Chair of Proteomics and Bioanalytics (with Prof. Bernhard Küster, Technische Universität München)

2009/11-2010/06 Research Associate in the Department of Analytical Sciences at Cellzome AG (Heidelberg)

2007/03-2007/05 Intern in the Department of Physical Chemistry and Information Technology (Joint Innovation Lab) at BASF SE (Ludwigshafen)

Education

Since 2010/07 PhD thesis: "Towards Comprehensive Identification of Proteins from MALDI imaging" Technische Universität München

2008/12-2009/05 Diploma thesis: "Two-dimensional reversed phase chromatography in mass spectrometry based Proteomics" at Cellzome AG, Heidelberg

2006/09-2009/06 Study of biological chemistry
 at Mannheim University of Applied Sciences
 Final degree: Dipl.-Ing. Biologische Chemie (FH)

2003/09-2006/08 Study of chemistry at the Ruprecht–Karls-Universität Heidelberg
 Pre-degree: 2006/03

1994-2003 Katholisch Freies Gymnasium St. Konrad (Ravensburg)
 Abitur: 2003/07