TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Diplomarbeit

Verification of a RiPP Precursor on the Peptide Level

Ausgeführt am

Institut für Chemische Technologien und Analytik

an der

Technischen Universität Wien

unter der Anleitung von

Univ. Prof. Priv.-Doz. Dipl.-Ing. Dr.techn. Ruth Birner-Grünberger

Ass.Prof. Dipl.-Ing. Dr.techn. Matthias Schittmayer-Schantl

durch

Julia Rafetzeder

01618543

Wien, Mai 2022

_____

Julia Rafetzeder

Abstract

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a group of secondary metabolites with unique biosynthetic pathways resulting in a wide variety of structures and bioactivities. RiPP biosynthesis is a two-stage process. First, linear precursor peptides consisting of an N-terminal leader sequence, a core sequence and an optional C-terminal recognition sequence are translated by classical ribosomal translation. After this, the linear precursor is heavily modified by an assembly of tailoring enzymes yielding the mature, often cyclized product. A set of potential RiPP precursors has been identified in the genome of the fungus *Trichoderma reesei* in a recent genome mining approach and gene cluster activity was demonstrated on the transcriptome level.

The aim of this thesis was to determine whether the precursor/leader sequence mRNA is translated into the corresponding precursor/leader peptides. To this end, extracts of wild type and RiPP knockout strains of *T. reesei* were subjected to quantitative LC-IMS-MS/MS analysis. Different protease digestion strategies were tested to maximize sequence coverage of the RiPP precursor. By employing a simple extraction procedure followed by nanoLC-IMS-MS/MS analysis and a common proteomics database search, the translation of the precursor mRNA into the corresponding peptide could be verified. Moreover, several extraction and analysis approaches were tested on model RiPPs produced in the fungus *Aspergillus flavus* to develop an efficient method for the analysis of mature RiPPs which will set a basis for the discovery and identification of RiPPs with yet unknown structure in *T. reesei.*

The second part of this thesis compares the efficiency of different sample preparation methods for proteomics with a special focus on sub-microgram sample input. A recently published paramagnetic bead-based approach (single-pot, solid-phase enhanced sample preparation, SP3) for sample preparation was compared to an established in-solution digestion method and a commercially available kit for sample preparation (PreOmics). The SP3 approach was further optimized by adopting different peptide cleanup strategies in order to increase the number of identified proteins. It was shown that for protein amounts above 10 µg, an increase in protein identifications can be achieved by employing an additional desalting step.

Kurzfassung

Ribosomal synthetisierte und post-translational modifizierte Peptide (RiPPs) sind eine Gruppe von Sekundärmetaboliten, die aufgrund einzigartiger Synthesewege eine hohe strukturelle Vielfalt und unterschiedliche bioaktive Eigenschaften besitzen. Die RiPP-Biosynthese ist ein zweistufiger Prozess. Zuerst wird ein lineares Vorläuferpeptid während der klassischen Translation hergestellt, welches aus einer N-terminalen Leader-Sequenz, einer Core-Sequenz und einer optionalen C-terminalen Followersequenz besteht. Anschließend wird der lineare Vorläufer durch unterschiedliche Enzyme modifiziert. Ein Satz potentieller RiPP-Vorläufer wurde bereits im Genom des Pilzes *Trichoderma reesei* durch Genome Mining identifiziert. Die Aktivität des Clusters wurde außerdem auf Transkriptomebene festgestellt.

Ziel dieser Arbeit war es festzustellen, ob die Vorläufer/Leader-Sequenz mRNA in die entsprechenden Vorläufer/Leader-Peptide translatiert wird. Dafür wurden Extrakte von Wildtypen und RiPP-Deletionsstämmen von *T. reesei* mithilfe von LC-IMS-MS/MS quantitativ analysiert. Durch Verdau mit unterschiedlichen Proteasen soll die Sequenzabdeckung der RiPP-Vorläufer maximiert werden. Mithilfe einer einfachen Extraktionsmethode, der Analyse mittels nanoLC-IMS-MS/MS sowie einer gewöhnlichen Datenbanksuche konnte gezeigt werden, dass die Vorläufer-mRNA tatsächlich in das entsprechende Peptid translatiert wird. Außerdem wurden mehrere Extraktions- und Analysenmethoden an bekannten RiPPs des Pilzes *Aspergillus flavus* untersucht, um eine effiziente Methode für die Analyse von reifen RiPPs zu entwickeln, die den Grundstein für die Entdeckung von unbekannten RiPPs in *T. reesei* liefern wird.

Im zweiten Teil der Arbeit wurde die Effizienz verschiedener Methoden zur Probenvorbereitung für Proteomik-Analysen mit Fokus auf Probenmengen im sub-Mikrogrammbereich verglichen. Eine kürzlich veröffentlichte Methode basierend auf paramagnetischen Beads (SP3) wurde mit einer Methode basierend auf Proteinverdau in Lösung sowie einem kommerziell erhältlichen Kit zur Probenvorbereitung (PreOmics) verglichen. Die SP3-Methode wurde weiters optimiert, indem verschiedene Strategien zur Peptidreinigung eingesetzt wurden, um die Anzahl identifizierter Proteine zu erhöhen. Es konnte gezeigt werden, dass die Anzahl an identifizierten Proteinen bei Proteinmengen über 10 µg mithilfe eines einfachen Entsalzungsschrittes erhöht werden kann.

Table of contents

1 Introduction

1.1 Sample preparation for proteomics

Proteomics is the large-scale analysis and characterization of proteins in a biological system. There are different ways for protein analysis, mass spectrometry (MS) based approaches are, however, most commonly used for the analysis of either intact proteins ("top-down") or peptides obtained from enzymatic digestion ("bottom-up") due to high sensitivity and throughput [1, 2].

The first step of a typical proteomic analysis is the extraction of the proteins from a biological sample. Proteins are very heterogeneous in terms of their size, structure, charge and hydrophobicity and often carry different post-translational modifications which makes their isolation extremely challenging. Reagents, such as buffers, salts and detergents are usually added to support cell lysis as well as the release and solubilization of proteins [3]. Cell lysis is often promoted by mechanical disruption which physically breaks apart the cells, e.g. sonication or French pressing [4].

For bottom-up approaches, proteins have to be enzymatically digested to reduce the complexity of the heterogeneous protein sample which circumvents problems associated with the analysis of intact proteins such as ionization and characterization. During digestion, solubilization and unfolding of proteins is vital to ensure proper accession to all cleavage sites. This is usually achieved by the use of sodium dodecyl sulfate (SDS), which denatures macromolecules very effectively [2]. To achieve complete protein unfolding, reduction of disulfide bridges (within and between protein subunits) is necessary. This is usually accomplished with reductants containing free thiols such as dithiothreitol (DTT) or trialkylphosphines such as Tris-(2-carboxyethyl)phosphine (TCEP). Possible renaturation is prevented by alkylation of the reduced cysteine residues, e.g. with iodoacetamide, to irreversibly block free sulfhydryl groups [5].

However, detergents and chaotropes at high concentrations are not compatible with proteases used for enzymatic protein digestion as they reduce proteolytic activity [6]. One widely used method for removing these contaminants is precipitation where proteins are precipitated by the addition of organic solvents, e.g. acetone, a mixture of

methanol and chloroform or trichloroacetic acid. However, precipitation is often accompanied by sample loss. Acetone precipitation, for example, provides 50 to 100 % recovery, depending on the ionic strength of the solution and the initial protein concentration with higher protein concentrations resulting in higher recovery [4].

Purified proteins are then enzymatically digested into peptides within the mass range of around 500 – 3,000 Da. Different proteases can be employed for this step with different specificities, efficiencies and digestion conditions. The most commonly used protease is trypsin which cleaves the peptide bonds C-terminal to lysine and arginine, except when followed by proline [7].

Peptides usually have to be purified and desalted prior to mass spectrometry analysis as buffers, salts and other contaminants interfere with subsequent MS analysis by reducing detection sensitivity for peptides [8]. Several desalting techniques exist which are in many cases based on the hydrophobic binding of peptides to a reversed-phase material. After the removal of salts and other contaminants, the purified peptides can be eluted in an organic solvent. Examples for desalting methods are solid phase extraction, commercially available tips containing reversed-phase material or self-made tips. While commercial desalting tools often suffer from recovery problems, large elution volumes, limited capacity and high costs, self-made tips overcome these problems and can easily be adapted in terms of capacity and functionality [9, 10].

Important aspects which have to be considered for sample preparation are the compatibility with different reagents such as chaotropes, salts or detergents, unbiased protein enrichment and recovery with respect to molecular weight, hydrophobicity and abundance, quantitative recovery over a wide range of protein input, a simple and rapid execution, low cost per sample, high throughput and the possibility for automation [6, 11]. To tackle these issues, various methods for sample preparation have been developed.

1.1.1 In-solution digestion

One common and widely used method for preparation of bottom-up proteomic samples is in-solution digestion. It is a very simple and powerful technique that involves denaturation, reduction, alkylation and digestion of proteins in the liquid phase in contrast to gels or filters. The in-solution method can be performed in a single vessel

to reduce sample loss resulting from solution transfer between different tubes [4]. Although in-solution digestion approaches meet many of the aforementioned criteria, they are usually incompatible with many reagents such as SDS which increase the efficiency of cell disruption and enhance protein yield [6].

Therefore, additional processing steps can be employed for removing contaminants prior to digestion or MS analysis. These approaches include protein precipitation, affinity capture [12, 13] and spin-filter enrichment [14]. Cleanup steps allow the use of a wider collection of reagents during sample preparation. The addition of these steps, however, leads to longer processing times, more complex protocols and higher costs per sample [6].

1.1.2 Single-pot, solid-phase enhanced sample preparation (SP3)

Recently, an alternative protein cleanup method was presented by Hughes *et al.* called SP3 which is short for single-pot, solid-phase enhanced sample preparation. This approach uses paramagnetic beads with a hydrophilic coating functionalized with carboxylate groups. Proteins are captured on the surface of the beads initiated by an organic solvent like ethanol or acetonitrile via a hydrophilic interaction mechanism. Contaminants can then be rinsed away easily as proteins are trapped in a solvation layer on the paramagnetic beads. Enzymatic digestion is performed in the presence of the beads. According to the protocol, the digestion solution can then be directly subjected to analysis by liquid chromatography-tandem mass spectrometry (LC-MS/MS) without the need for further cleanup. However, additional washing steps can be employed by immobilizing the peptides on the beads and rinsing away unwanted contaminants [6]. A combination of rinsing steps with 70 % ethanol and 100 % acetonitrile have been shown to be effective in removing various reagents [15]. The standard SP3 workflow is depicted in Figure 1.

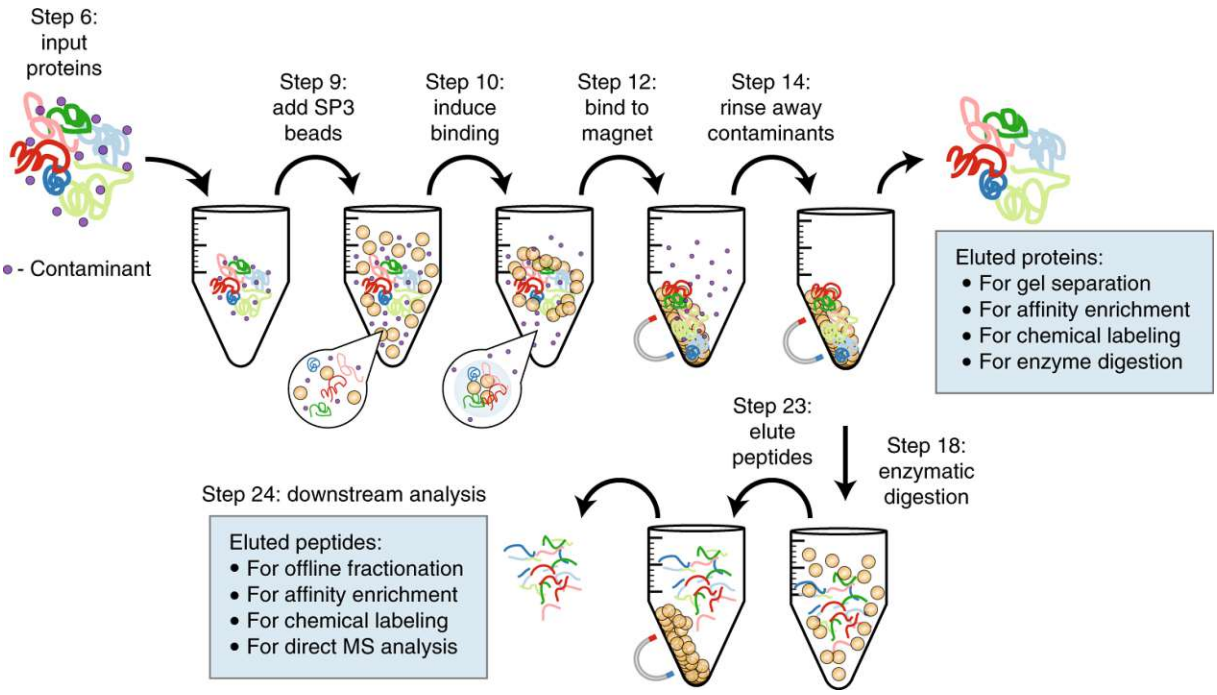

Figure 1: SP3 workflow for processing of protein samples [6].

The SP3 approach has several advantages compared to conventional sample preparation methods for proteomics. It is compatible with a variety of chemicals (i.e. detergents, chaotropes and salts), adaptable to a wide range of sample quantities down to 100 ng of input material and provides high recovery. The protocol is simple and requires minimal handling as all steps can be performed in a single tube. Furthermore, the SP3 method is very cost efficient and can be automated easily which can further decrease costs and increase throughput [6, 15].

1.2 Proteome analysis via LC-MS/MS

After enzymatic digestion and possible cleaning steps, peptides are first separated via liquid chromatography and then introduced into a tandem mass spectrometer through electrospray ionization (ESI) for analysis and identification [16].

1.2.1 Chromatographic separation

One of the most commonly used techniques for peptide separation is high-performance liquid chromatography (HPLC). A typical HPLC system used for proteomics consists of a column packed with reversed-phase material as stationary phase. As the mobile phase containing the analytes is pumped through the column, the peptides interact with alkyl chains (e.g. C_4, C_8 or C_{18}) on the surface of the column

material via hydrophobic interaction. Hydrophilic peptides will therefore elute faster than hydrophobic peptides. As a result, peptides with similar or even identical masses but different hydrophobicity will have different retention times and can easily be resolved and subsequently detected by the mass spectrometer [17, 18].

Peptides are usually eluted under acidic conditions using a gradient of water and an organic solvent such as acetonitrile [18]. The main advantages of reversed-phase liquid chromatography are the high peak capacity and the compatibility of mobile phase solvents with ESI [19].

1.2.2 Electrospray ionization

In the next step, peptides eluting from the LC column have to be ionized for mass spectrometric analysis. The most commonly used method for ion generation of liquid samples is ESI, a so-called “soft ionization” technique that produces intact ions from complex macromolecules with high ionization efficiency [20, 21]. The sample solution is introduced into the ESI chamber through a stainless-steel needle which is held at a few kilovolts relative to the counter electrode [22]. Due to the electric field applied to the surface of the solution emerging from the needle, a Taylor cone starts to form that emits charged liquid droplets [23]. The droplets migrate towards the capillary inlet and undergo solvent evaporation which leads to an increasing charge density on the droplet surface until the Rayleigh limit is reached. At this point, the Coulomb repulsion between the charges becomes more powerful than the surface tension at the surface of the droplet causing a so-called coulomb explosion into many smaller droplets [22].

To explain the final formation of gas-phase ions, two theories have been developed. The ion evaporation model assumes that solvent evaporation and droplet shrinkage occur until the field strength on the droplet surface is large enough for the ejection of small, solvated ions. According to the charge residue model, droplets generated by cycles of evaporation and Coulomb explosions contain only one ion [21]. In the case of peptides, analyte ions are commonly multiply charged. The resulting m/z values are therefore lower and typically fall in the mass ranges of common mass analyzers [24].

1.2.3 Nano liquid chromatography

Cellular protein abundance can range from only few copies per cell to a million copies per cell [25]. This enormous dynamic range makes the detection of low concentrated proteins very challenging. Thus, the suitability of an analytical method in the field of proteomics strongly depends on its ability to sensitively detect low abundant compounds despite the presence of compounds which are orders of magnitudes more concentrated [18]. During the chromatographic separation, a sample compound is subjected to dilution which increases proportionally with the square of the column radius (r^2), with the square root of the column length and with the plate height. Therefore, the final compound concentration at the peak maximum increases inversely to r^2 under identical injection and chromatographic conditions [26]. It can be calculated that a reduction of the inner diameter of a column from 4.6 mm to 75 µm results in an approximately 3700-fold increase in peak height and mass sensitivity. For this reason, nanoLC combined with ESI-MS is the most commonly used platform for proteomics. NanoLC columns have an inner diameter of 0.1 mm or lower and flow rates in the nL/min region [27]. Apart from higher sensitivity provided by nanoLC, the lower flow rates are also advantageous for the ESI mechanism as the initial size of the charged droplets produced in the spraying process is reduced and the desolvation process of analyte molecules is more efficient. As a result, a larger part of analyte molecules present in the droplets is ionized and transferred to the mass spectrometer [28].

1.2.4 Mass spectrometric analysis

The peptide ions then enter the mass spectrometer where their mass-to-charge ratios are measured. In cases where the sample consists of only one protein, protein identification can be performed using a method called peptide mass fingerprinting (PMF). This technique is based on the exact measurement of the peptide masses and a subsequent comparison of these masses with theoretical peptide mass lists obtained from an *in silico* digestion of database proteins [18]. In many cases, the peptide mass fingerprint does not contain all the theoretical peptide masses and often, many of the experimental peaks do not correspond to any of the peptide masses in the protein found. Unambiguous protein identification is therefore often challenging [29]. Peptides obtained from more complex protein mixtures elute from the LC column in an arbitrary order. As a consequence, the simple comparison of peptide masses to theoretically

generated mass lists is not possible as no unambiguous assignments can be made [18].

To overcome this limitation, a technique called tandem mass spectrometry or MS/MS is typically employed to gain additional information on the peptide ions. In the first MS step, the mass-to-charge ratio of the intact peptide ion, the so-called precursor ion, is measured. For the second MS measurement, ions with specific m/z-ratios coming from the MS1 are selected, isolated from all other peptide ions by e.g. an analytical quadrupole and then split into smaller fragments. Fragmentation of the precursor ions usually occurs in a process called collision-induced dissociation (CID) where peptide ions collide with inert gas atoms or molecules such as He, Ar, N_2, etc. In a quadrupole time-of-flight (TOF) instrument, peptide ions are accelerated into the collision chamber containing the inert gas by applying an electric potential to increase the kinetic energy. By collision with the inert gas, a peptide ion is fragmented along its backbone at the peptide bonds leading to a set of C-terminally and N-terminally shortened peptide ions whose m/z-ratios are determined in the second MS step yielding sequence specific MS/MS spectra [16, 18].

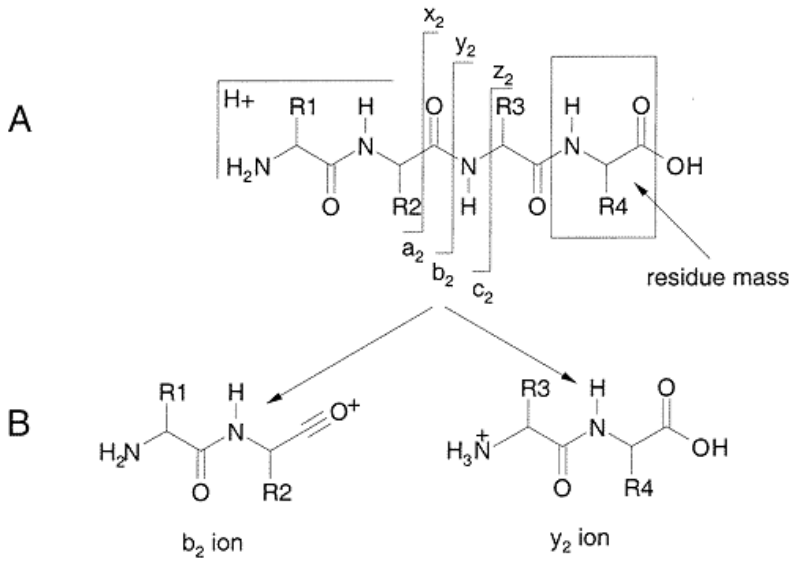

Figure 2: (A) Nomenclature for peptide fragment ions that form upon collision-induced dissociation. (B) Examples for b- and y-ions [30].

The cleavage of the amide bonds leads to formation of different fragment ions. If the charge from the parent peptide ion remains on the N-terminal side of the broken amide bond, the corresponding fragment ion is called an a-, b- or c-ion. However, if the charge remains on the C-terminal side of the cleavage site, the fragment ion is referred to as

x-, y- or z-ion, as shown in Figure 2. At the collision energy levels set in a typical CID process, the by far dominant fragmentation events occur at the amide bond of the peptide backbone, yielding mainly b- and z-ions. The position of the cleaved amide bond resulting in b- or z-ions within the analyzed peptide is indicated as subscript number. These subscripts denote the number of amino acid residues present on either the b- or the y-ion [30]. The experimentally obtained MS/MS spectra are then compared to theoretical MS/MS spectra calculated from databases [18].

For proteome analysis via mass spectrometry, different mass analyzers are available. The most commonly used mass spectrometers are the quadrupole (Q), ion trap, time-of-flight (TOF), orbitrap and fourier transform ion cyclotron resonance (FT-ICR) analyzers [31]. In many cases, hybrid instruments for tandem mass spectrometry are used to take advantage of their different characteristics which include triple quadrupole mass spectrometers, where the first and third quadrupole act as mass filters and the second quadrupole as collision cell, TOF-TOF, Q-TOF or ion trap/FT-ICR instruments [18].

Quadrupole mass analyzers consist of four parallel metal rods in a vacuum cell. A direct current and a radio frequency voltage are applied to the rods in a way that only ions of specific mass-to-charge ratios are allowed to pass through. The remaining ions collide with the rods or are deflected to trajectories not recorded by the detector. The m/z transmission range can be shifted by varying the voltage values [18]. In a Q-TOF instrument, the first quadrupole acts as a mass filter which transmits only a specific ion of interest. This is usually achieved by selecting a mass window with a width of 1 to 3 Th. After acceleration to an energy of 20 to 200 eV, the ion enters the second quadrupole which acts as collision cell where the ion is subjected to collision-induced dissociation. In the next step, the resulting fragment ions as well as the remaining parent ions are re-accelerated and focused into a parallel beam which is then deflected orthogonally into the time-of-flight column. There, the ions are accelerated by a voltage (U) of several keV [32]. Their potential energy (E_p) which depends on the charge of the particle (q) and the strength of the electric field is converted into kinetic energy (E_k) which is proportional to the mass of the ion (m) and its velocity (v). The corresponding relationships are shown in Eq. (1) and (2) [33].

$$E_p = qU \tag{1}$$

$$E_k = \frac{1}{2}mv^2 \tag{2}$$

As the potential energy is fully converted into kinetic energy, Eq. (1) and (2) are equal:

$$qU = \frac{1}{2}mv^2 \tag{3}$$

The charged particle moves in a field-free tube, its velocity will therefore not change after acceleration. The velocity of the ion can be easily determined from the length of the flight path (L) and the time of flight of the ion (t). From the velocity, the ion's mass-to-charge ratio (m/z) can be calculated as shown in Eq. (4) and (5).

$$v = \frac{L}{t} \tag{4}$$

$$qU = \frac{1}{2}m\left(\frac{L}{t}\right)^2 \tag{5}$$

By exchanging the charge of the particle (q) by the product of the elementary charge (e) and the number of charges (z) and rearranging of Eq. (5), Eq. (6) can be obtained which expresses the mass-to-charge ratio in terms of all other factors:

$$\frac{m}{z} = \frac{2eUt^2}{L^2} \tag{6}$$

From Eq. (6), it can be concluded that ions with low m/z-ratios will reach the detector faster [33]. A typical setup for a quadrupole time-of-flight mass spectrometer is shown in Figure 3 [34].

Compared to scanning mass analyzers, TOF instruments benefit from a high transmission efficiency which leads to very high sensitivity, high analysis speed and from the capability of measuring a theoretically unlimited mass range which makes it suitable for soft ionization techniques producing mainly intact molecular ions [33].

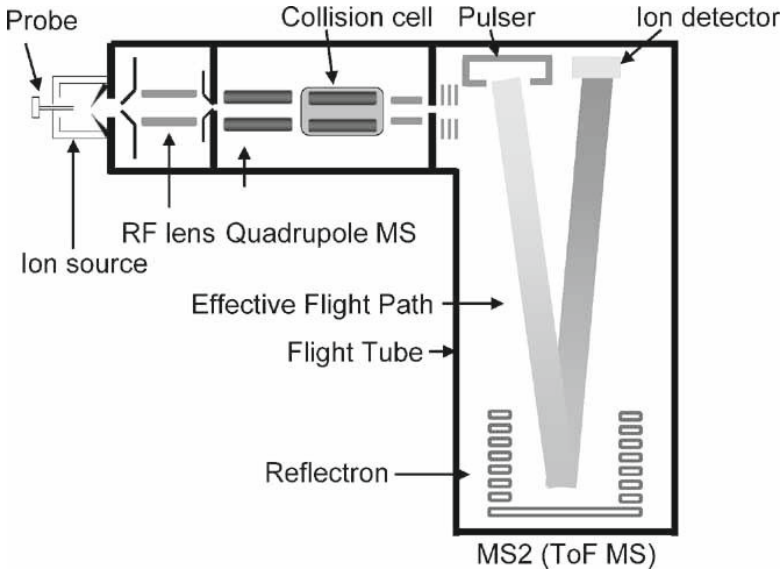

Figure 3: Schematic overview of a quadrupole time-of-flight mass spectrometer [34].

The mass resolution R of a TOF mass analyzer is proportional to the flight time and the length of the flight path, as shown in Eq. (7), where Δm and Δt refer to the peak widths at the 50 % level on the mass and time scales and Δx is the thickness of an ion bundle reaching the detector [33]:

$$R = \frac{m}{\Delta m} = \frac{t}{2\Delta t} \approx \frac{L}{2\Delta x} \tag{7}$$

As the resolution increases with flight time and flight path, one way to improve the resolution of a TOF instrument is to extend the length of the flight tube. Mass resolution also depends on the distribution of flight times of ions with the same mass-to-charge ratio. This distribution is affected by the length of the ion formation pulse (time distribution), the space distribution, i.e. the size of the volume where the ions are formed, and the initial kinetic energy spread among the ions [33].

One possibility to decrease the distribution of initial kinetic energies of ions and therefore increase mass resolution is to use an electrostatic reflector. This so-called reflectron is positioned on the opposite side of the ion source and acts as an ion mirror which deflects the ions and sends them back through the flight tube to the detector. The main advantage of the reflectron is the correction of the kinetic energy dispersions of ions with the same m/z ratio. Ions with higher kinetic energy traverse the tube faster and will travel further in the reflectron than ions with lower kinetic energy. As a result, faster ions will reach the detector at the same time as slower ions with the same m/z

because they take a longer path. Not only is the initial spread of kinetic energies corrected, but the introduction of a reflectron also doubles the flight path which additionally improves the mass resolution of the instrument [33].

1.2.5 Ion mobility spectrometry-mass spectrometry

In the last years, ion mobility spectrometry (IMS) in combination with mass spectrometry for the analysis of biomolecules attracted growing interest [35]. Ion mobility provides an additional dimension of separation which improves the analysis of complex samples [36]. Different types of ion mobility spectrometers are used with mass spectrometers, for example drift time, field asymmetric waveform, travelling wave or trapped ion mobility spectrometers [37].

The traditional drift time ion mobility spectrometer (DTIMS) measures the time an ion takes to migrate through the drift tube filled with an inert stationary gas in the presence of a weak electric field. The ion mobility depends on the differential velocities at which ions travel through the tube under the influence of an electric field. The mobility is therefore a function of the ion's interaction with the gas which depends on many factors including the temperature, pressure and polarizability of the gas as well as the charge, size and shape of the ion [38]. Ion mobility and mass-to-charge ratio can be used to calculate the collision cross section (CCS) of an ion which also describes the interaction of an ion with the drift gas [39].

In traveling wave IMS (TWIMS), a sequence of symmetric potential waves continually propagates through the tube propelling ions along with a velocity which depends on the ion mobility. Consequently, different species migrate through the tube with varying durations [40].

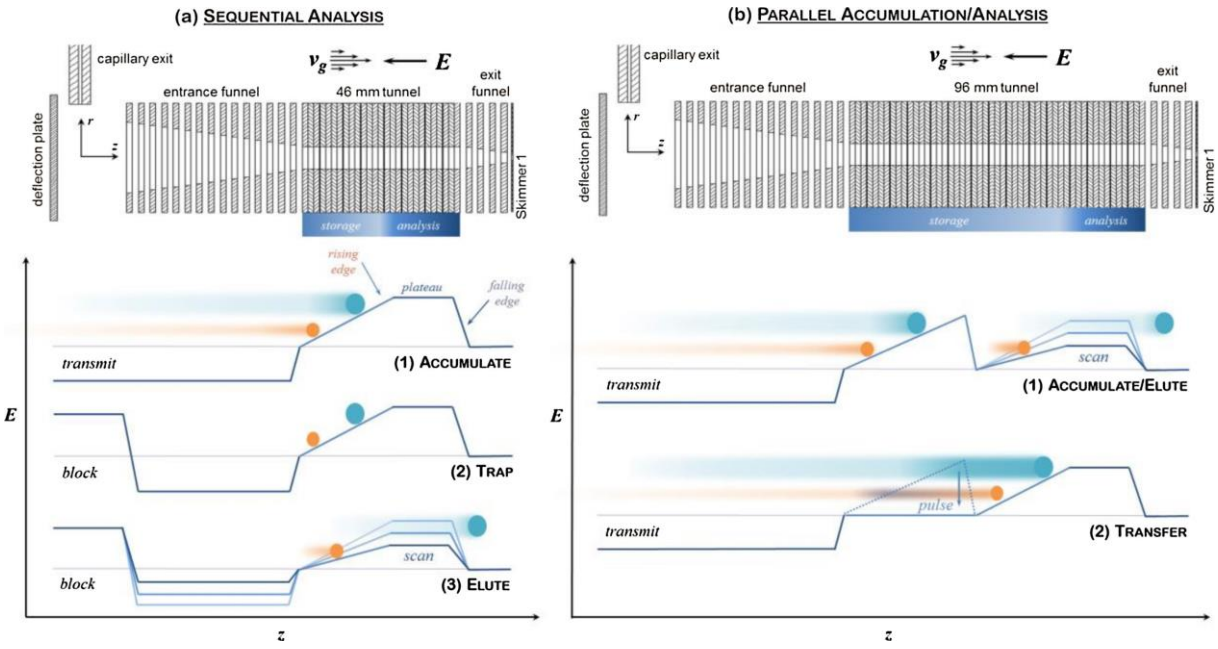

Figure 4: (a) TIMS tunnel for sequential analysis using a single storage region and (b) parallel accumulation with two storage regions. Plots of the electric field as function of the tunnel length are shown below [38].

A trapped ion mobility spectrometer (TIMS) is another example of such an analyzer. Contrary to a conventional IMS instrument, where ions constantly migrate through a stationary gas driven by an electric field, a TIMS experiment is based on using an electric field to hold the ions in place, i.e., to trap them against a moving gas. The concept of a TIMS tunnel for sequential analysis is depicted in Figure 4 (left). It consists of a set of electrodes which are separated into an entrance funnel, the TIMS tunnel and the exit funnel. The second-generation TIMS funnel has the same setup except for an increased tunnel length which allows the addition of a second TIMS stage in the same tunnel. This enables a mode of operation where ions entering the funnel can be accumulated in the first stage, while simultaneously analyzing ion mobility from previously accumulated ions in the second stage [41].

Due to their acquisition speed, TOF instruments are especially suitable mass analyzers which can be employed after the ion mobility separation. Moreover, analytical quadrupoles and collision cells can be included between the TIMS analyzer and the high performance mass spectrometer [38]. The combination of trapped ion mobility spectrometry with TOF instruments can increase speed, sensitivity and selectivity, without compromising ion transmission or resolution [42].

1.2.6 Methods for data acquisition

There are different MS/MS methods for ion selection and measurement to acquire data for proteomics analysis. In data-dependent acquisition (DDA), a subset of the most intense ions is selected in the first stage of tandem mass spectrometry. These ions are individually isolated, fragmented and analyzed in sequential MS2 scans. For very complex peptide mixtures, DDA-based methods select the most abundant precursor ions in a stochastic manner which often results in low analytical reproducibility across technical replicate experiments and problems in quantifying low abundant peptides [43, 44]. Due to its flexibility and simplicity in setup and data analysis, data-dependent acquisition is still a commonly used method for proteomics [45].

Alternatively, data-independent acquisition (DIA) exists where all peptides within a defined mass-to-charge window are fragmented regardless of their intensity. The m/z window is then systematically scanned across the whole m/z range. As a result, reproducibility in protein identification between technical replicates is significantly improved [45]. An important requirement for DIA is the availability of robust data analysis tools because many peptides in a certain m/z window are fragmented together which leads to very complex MS/MS spectra [43]. A schematic overview of data-dependent and data-independent acquisition is shown in Figure 5.

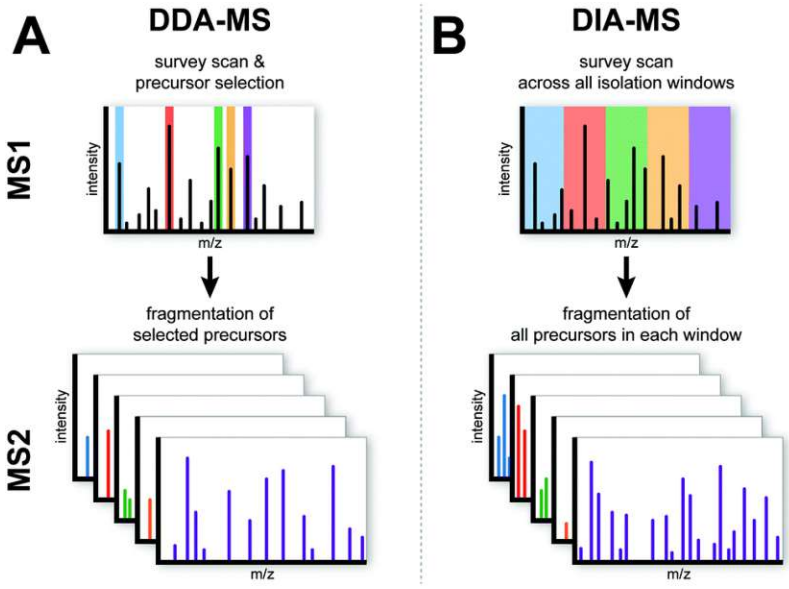

Figure 5: Overview of the principle of DDA and DIA [45].

As described above, precursor peptides are selected and fragmented one at a time in data-dependent acquisition, whereas all other ions are discarded entirely. The

combination of trapped ion mobility with an orthogonal quadrupole time of flight mass spectrometer can mitigate this problem. For this specific instrumentation, a method called parallel accumulation-serial fragmentation (PASEF) has been developed. Figure 6 shows the setup of a timsTOF Pro, consisting of a dual TIMS analyzer followed by a quadrupole mass filter, a high speed collision cell and a time of flight mass analyzer [46].

After chromatographic separation (Figure 6, A), peptides are subjected to ESI and pushed through the TIMS tunnel by a gas flow while simultaneously experiencing a counteracting electrical field. Ions are then held at a position in the tunnel where the two forces are equal. Since the drag exerted by the gas flow is proportional to the collisional cross section, ions are trapped at different positions according to their ion mobility. The TIMS tunnel operates as a dual TIMS and is separated into two sequential parts where the first section performs ion accumulation, that is all ions entering the instrument are stored. The second region is responsible for the actual TIMS analysis (analyzer 1 and 2 in Figure 6, B). Once the analysis is finished, all ions are released into the mass analyzer by decreasing the voltage gradient and the storage part of the tunnel is filled again. Consequently, up to 100 % of all ions which enter the mass spectrometer can be subjected to mass analysis [42, 46].

In TIMS, ions with low mobility will leave the tunnel first, followed by ions with higher ion mobility and therefore smaller collisional cross sections. The TIMS step does not only separate ions according to their size and shape, but it also increases signal-to-noise ratios about 50-fold compared to standard acquisition modes because ions are accumulated in the trapping stage over time and concentrated into narrower IMS peaks during elution resulting in higher signal intensities (Figure 6, C) [41].

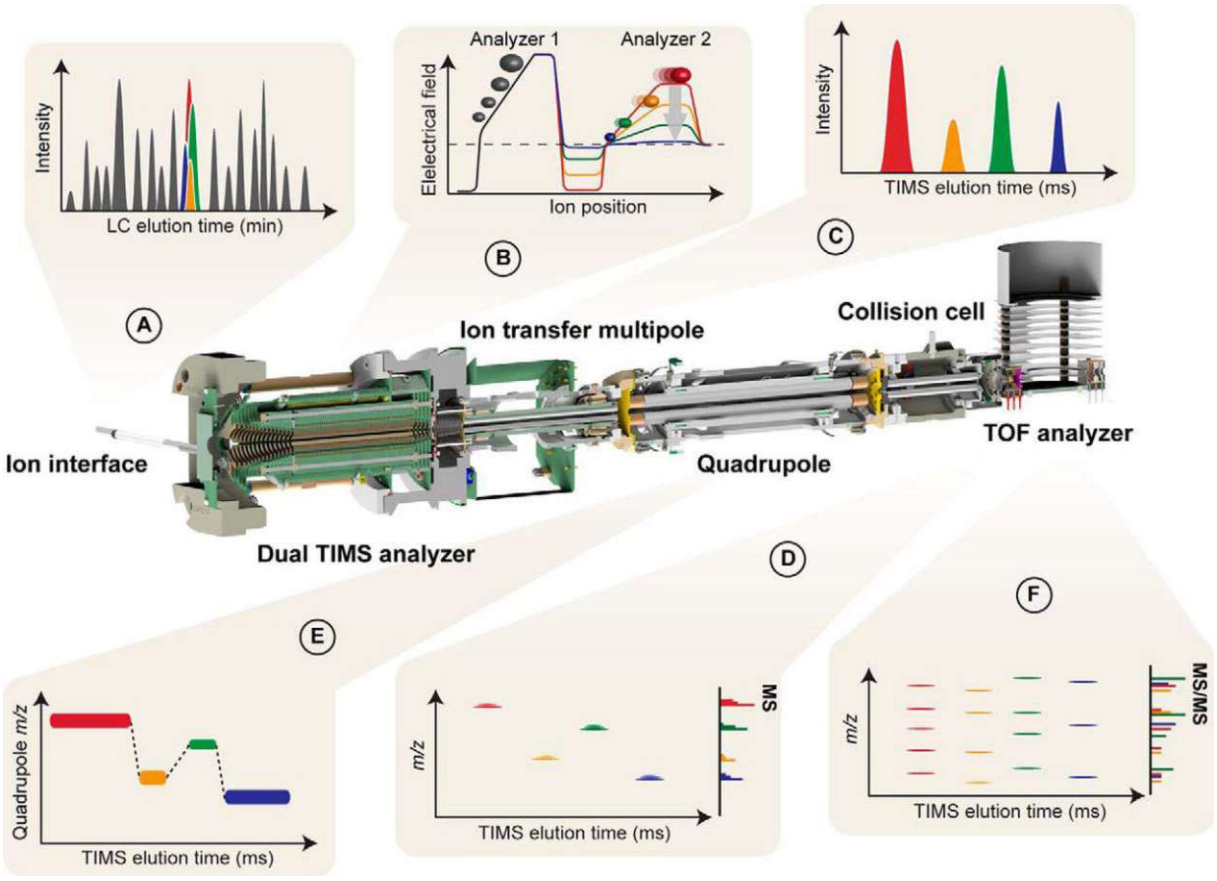

Figure 6: Parallel Accumulation - Serial Fragmentation (PASEF) with the timsTOF Pro [46].

After ion mobility separation, ions leave the TIMS device, pass through the quadrupole mass filter and subsequently enter the collision cell. Either intact (MS scans) or fragment (MS/MS scans) ions are then introduced into the time of flight mass analyzer (Figure 6, D) [46].

PASEF synchronizes the selection of precursor ions by the quadrupole mass filter with the elution of ions from the TIMS instrument (Figure 6, E). The quadrupole can therefore switch within sub-milliseconds between ion mobility resolved precursor ions with different mass-to-charge ratios. For this reason, multiple ions stored in the TIMS device are subjected to fragmentation yielding multiple mobility resolved MS/MS spectra from a single TIMS scan (Figure 6, F). Compared to conventional shotgun acquisition methods, PASEF can be employed in proteomics to increase the sequencing speed severalfold without sacrificing sensitivity [46].

1.2.7 Database search

As already described in chapter 1.2.4, peptide ions selected by the quadrupole dissociate into fragment ions upon collision with an inert gas giving rise to a specific peak pattern in the MS/MS spectrum. Each mass difference between neighboring peaks corresponds to the mass of a certain amino acid residue. As a result, the peptide sequence can be deduced directly from the MS/MS spectrum, an approach referred to as *de novo* sequencing, which is depicted in Figure 7 [47].

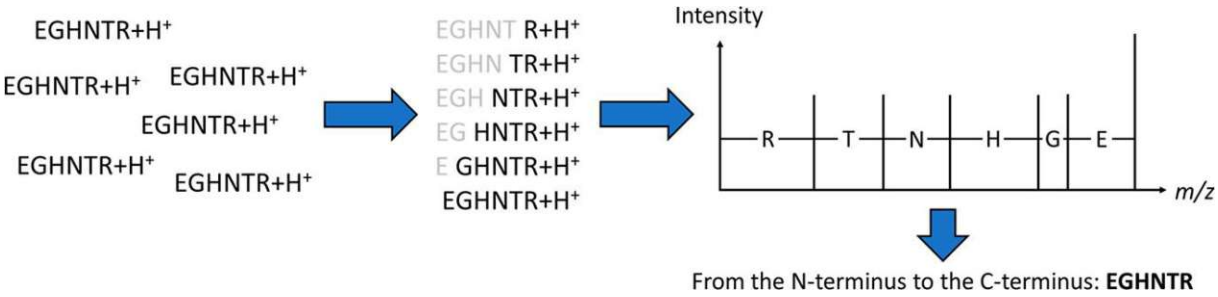

Figure 7: Principle of *de novo* sequencing from an ideal MS/MS spectrum [47].

Alternative methods to *de novo* sequencing are based on database searches. One approach, termed Peptide Mass Fingerprinting (PMF), relies only on MS spectra and has already been described in chapter 1.2.4. PMF is usually applied for the identification of only a single protein [48].

Database searches can also be performed using tandem mass spectrometric data. These approaches are generally more discriminating and specific as MS/MS spectra also contain information on the peptide sequence. One method relying on tandem mass spectra is the peptide sequence tag method [48]. In this approach, an *in silico* digestion of protein sequences from a database is performed with the specificity of the protease used for digestion of the sample. By means of partial *de novo* sequencing, short segments of a peptide sequence (sequence tags) can be obtained from experimental MS/MS spectra. The experimentally determined peptide masses are then compared to predicted masses together with the sequence tags. An unknown peptide can therefore be identified if its molecular weight matches the predicted mass and if its sequence contains the tag [49]. Other approaches are not based on extracting sequence information from the MS/MS spectra. Instead, predicted peptides obtained from an *in silico* digestion are theoretically fragmented. The experimental tandem mass spectrum is then compared to the theoretical fragment masses. An unknown peptide

is identified if its fragment masses match the calculated masses with high significance. The principle of this procedure is illustrated in Figure 8 [47].

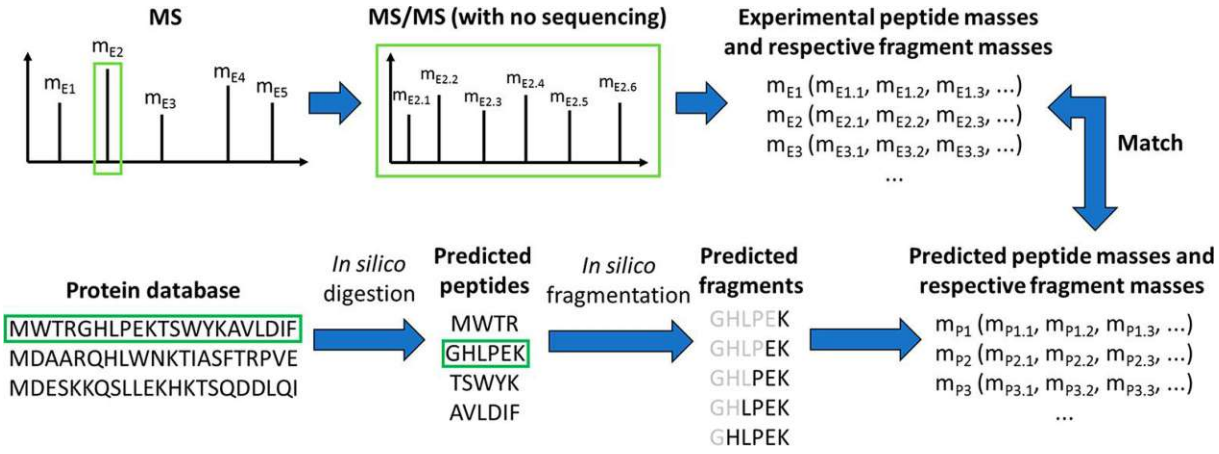

Figure 8: Principle of the MS/MS ion search [47].

An important aspect in database search is that proteins are identified with high confidence which is usually reported in terms of the false discovery rate (FDR). FDR is defined as the ratio of false positives to the number of total identifications, i.e. the sum of false positives and true positives above the score threshold. The number of false positive identifications can be determined by a decoy database which is usually created from the target protein database. However, the number of peptide sequences equal in the target and decoy database should be minimized. This is most often achieved by reversing the target protein sequences. The decoy database can also be derived from the target database in a stochastic way [50].

1.2.8 Quantification

Mass-spectrometry-based proteomics is a powerful method not only for protein identification but also for their quantification. For quantitative analysis, different approaches have been developed which can be divided into stable-isotope based or label-free quantification methods. Isotope based methods rely in on the incorporation of heavy versions of specific molecules into the peptide or the protein and are based on the fact that isotopically labeled peptides differ from unlabeled peptides only with regard to their mass but not in terms of chromatographic separation. The label can be introduced into the peptide or protein either through chemical derivatization or by metabolic labeling [51].

One widely used technique based on chemical labeling are isobaric mass tags which have identical overall mass but the distribution of heavy isotopes within their structure varies. Isobaric tags are in most cases amine-reactive, but cysteine- or carbonyl-reactive tags are also available. They consist of an isotopic reporter group which is linked to a balance group to normalize the total mass of the tag. The balance group is further attached to a peptide reactive group which, in case of an amine reactive group, forms an amide bond with N-termini of peptides and ε-amino groups of lysine. In a typical workflow, samples are labeled with isobaric mass tags which all have the same total mass but only vary in terms of the mass of the reporter group. After labeling, samples are pooled and analyzed by LC-MS/MS. When subjected to collision induced dissociation, the amide linkage fragments in a similar way to the backbone peptide bonds. During fragmentation, the balance group is lost as neutral loss, whereas the reporter group remains charged. On the MS1 level, peptides with the same sequence appear as a single precursor ion, whereas the reporter ions appear as distinct masses in the MS/MS spectrum together with the sequence-specific b- and y-ions. The intensity ratio of the reporter ions indicates the relative abundance of peptide in the mixture. Examples of isobaric mass tags include TMT (short for tandem mass tags) and iTRAQ (isobaric mass tags for relative and absolute quantification), which is shown in Figure 9 as example [52]. Isobaric labeling strongly benefits from the capability of multiplexing up to 16 samples eliminating possible variations between LC-MS/MS runs while simultaneously increasing throughput potential [51, 53].

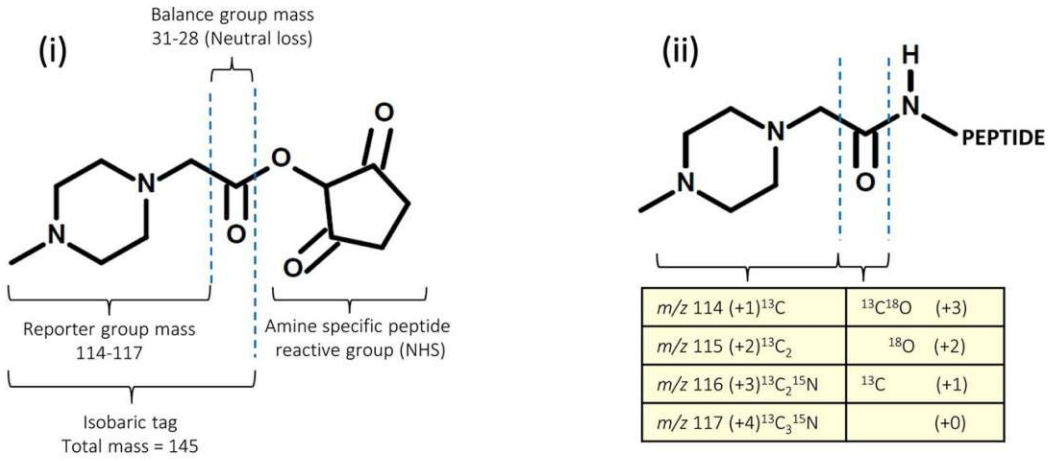

Figure 9: (i) Chemical structure of iTRAQ 4-plex reagents. (ii) The balance (carbonyl) group is lost during CID as neutral loss, whereas the reporter group retains the charge resulting in four different peaks in the MS2 spectrum [52].

In metabolic labeling, cells are cultivated in a medium containing a defined isotope. One commonly used method in mammalian systems is SILAC (stable isotope labeling with amino acids in cell culture), where one cell population is cultured in medium containing light forms of arginine, lysine or other essential amino acids and another population is fed with medium containing the corresponding heavy form of the amino acid. During cell growth in the medium, the light and heavy forms of the specific amino acid are incorporated into each protein which leads to a defined and easily detectable mass difference. Metabolic labeling methods allow the combination of samples directly after cell lysis. The variability in sample processing can therefore not affect measured intensity ratios which leads to very high quantitative accuracy [51].

Although stable isotope-based labeling methods are widely used and the gold standard for quantification, they suffer from additional preparation steps and are usually costly [54]. A further limitation of SILAC is that it cannot be applied to human samples and nondividing cells [55]. Moreover, due to an increase in sample complexity, the number of signals in the mass spectrum is also increased which reduces sensitivity and peptide identification rate [56]. When performing isobaric labeling-based experiments, the accuracy of relative quantification is strongly affected by ratio compression. This effect occurs due to co-elution of peptides within the isolation window used for fragmentation and typically leads to an underestimation of differences in peptide abundances [52, 57].

Alternatively, label-free quantification exist which is the most economical and simplest approach. Furthermore, the number of samples to be compared is not limited in label-free quantification in contrast to label-based methods. One problem of label-free quantification with prefractionation are differences in the fractions to be compared due to separate sample processing. This challenge can be solved by correct normalization of each fraction. In order to normalize LC-MS/MS runs of each fraction, the peptide ion signals can be used which are, however, spread over different runs. It is therefore not possible to sum up the peptide ion signals before knowing the normalization factors for each fraction [54].

To solve this dilemma, a procedure for intensity determination and normalization has been developed called MaxLFQ which is implemented in the MaxQuant proteomics platform. Formally, normalization coefficients have to be determined which multiply all

intensities in a certain LC-MS run. As the fractionation can be slightly irreproducible or the MS responses in a specific run can be different from average, it is not possible to directly adjust the normalization factors for each fraction to obtain an equalized total signal. Thus, it would be beneficial to sum up the peptide ion signals over all fractions in each sample. Doing this, however, would require the determination of the normalization coefficients specific for each run. The MaxLFQ algorithm makes use of the fact that the proteome does not change drastically between different conditions so that the average behavior can be used as a relative standard. After summation of peptide ion signals across fractions, the normalization coefficients are determined in a non-linear optimization model where overall changes for all peptides across all samples are minimized. To now perform relative protein quantification, the protein ratio between any two samples is first calculated using only peptide species occurring in both samples. The pair-wise protein ratio is then determined by first calculating peptide ratios from the intensities of peptides present in both samples. The pair-wise protein ratio is defined as the median of the peptide ratios, as it is less sensitive to outliers. This way, all pair-wise protein ratios are calculated from which protein abundance profiles for each protein can be obtained [54].

1.3 Ribosomally synthesized and post-translationally modified peptides

Natural products have been an important source for the discovery of new drugs throughout human history [58]. Even today, these compounds and derivates thereof still play a major role in the development of pharmaceuticals [59].

Natural products can be divided into different classes based on their structures. The most prominent ones are terpenoids, alkaloids, polyketides, non-ribosomal peptides and ribosomally synthesized and post-translationally modified peptides (RiPPs). RiPPs are produced in all three domains of life and have gained considerable interest in recent years due to their unique biosynthetic pathways, their vast structural diversity and their bioactive properties [60].

RiPP biosynthesis starts with the ribosomal production of a linear precursor peptide which is usually 20 – 110 amino acids long and consists of an N-terminal leader peptide, a core segment which is converted into the mature RiPP and, in some cases, a C-terminal recognition sequence [60]. The leader peptide is usually responsible for

the recognition by modification enzymes which install a variety of post-translational modifications (PTMs) in the core region. These modifications often take place on amino acid side chains (e.g., hydroxylation, γ-carboxylation, acetylation, methylation, etc.) [60, 61]. Further modifications include crosslinks between different amino acid residues [62], N-methylation on the peptide backbone [63, 64], epimerization or C- and N-terminal modifications [65].

The modified core peptide is eventually released by proteolytic removal of the leader peptide and, in some cases, cyclized, to finally yield the mature active RiPP [62]. The optional C-terminal recognition sequence is important for cyclization and excision of the core peptide [66]. Some post-translational modifications are catalyzed only after removal of the leader peptide and are therefore independent of the leader sequence [67]. The general scheme of RiPP biosynthesis is depicted in Figure 10.

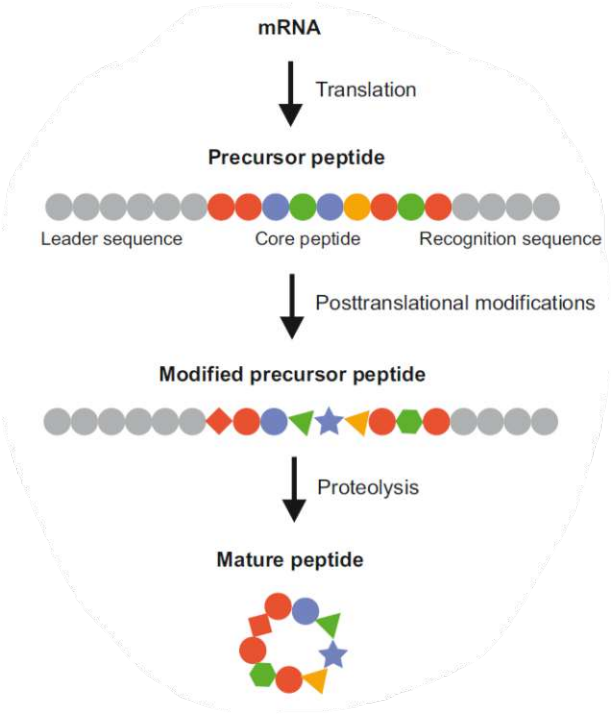

Figure 10: RiPP biosynthetic pathway [66].

Leader peptides play an important role in the biosynthesis of RiPPs as they are recognized by the modifying enzymes mainly by binding interactions which allows subsequent enzymatic processing of the core peptide. It has been shown that when exchanging the core peptide sequences with foreign sequences, the foreign core peptides were still subjected to post-translational modification which indicates that the modification enzymes only recognize the leader peptide but not the core segment [62].

Another interesting feature is that the amino acid sequence of RiPPs is genetically encoded. The genes encoding the modifying enzymes are usually located in close proximity to the gene encoding the RiPP precursor in a so-called biosynthetic gene cluster (BGC) which makes their identification in genomic sequences as well as their modification possible [66, 68].

The promiscuity of the modifying enzymes and the fact that the peptide sequence is genetically encoded makes RiPP biosynthetic pathways attractive for the biotechnological production of novel peptide therapeutics [66]. Due to their structural diversity, RiPPs possess a wide range of activities such as antibacterial [69, 70], antiviral [71] and antifungal [72]. RiPPs have been studied extensively in bacteria and many different classes with various bioactive properties have been identified and described so far [60]. In contrast, fungal RiPPs remain rather understudied. Only four different RiPP families have been described so far which are amatoxins/phallotoxins [73], borosins [63], dikaritins [74] and epichloëcyclins [75].

1.4 Discovery of potential RiPPs in the genome of *Trichoderma reesei*

RiPP discovery can be classified into top-down approaches where a known compound is linked to its biosynthetic gene cluster or bottom-up approaches where a putative RiPP gene cluster is first identified *in silico* by genome mining and then the transcription, translation and synthesis of the RiPP is achieved by different gene manipulation techniques [76]. For a bottom-up approach to be conducted, information about RiPPs has to be available. As a large number of verified RiPPs is available in bacteria, various genome mining tools have been established for the discovery of novel bacterial RiPPs [77, 78]. In fungi, however, only few RiPP classes are known which makes the application of genome mining approaches challenging.

However, the basic structure of bacterial and fungal RiPPs and their biosynthetic pathways are similar. Quite recently, Vignolle *et al.* demonstrated that genome mining tools originally established for the identification of bacterial RiPPs can also be used to discover new classes of fungal RiPPs [79]. All known fungal RiPPs could be identified as RiPPs using this unconventional genome mining approach. Additionally, several potential RiPP BGCs were found in the genome of four *Trichoderma* species.

1.5 Strategies for analyzing and identifying fungal RiPPs

The field of genomics has rapidly developed in the last years which helped to reveal that a variety of organisms exhibit cryptic gene clusters encoding for various secondary metabolites. The majority of these secondary metabolites can only be predicted by bioinformatic analysis of sequenced genomes, their detection and structural elucidation, however, is quite challenging as they are often not produced under laboratory conditions or at concentrations which are too low to be detected by standard methods [80].

RiPPs represent a special class of secondary metabolites as they are usually highly modified and often have very complex, nonlinear structures (e.g. branched or cyclic) [66]. The identification and analysis of mature RiPPs is therefore very difficult by conventional bottom-up proteomics workflows and standard *de novo* sequencing tools [81]. The preferable way of identifying unknown secondary metabolites is through metabolomics-based methods [82], which aim at comprehensively analyzing and quantifying a wide range of metabolites in biological samples [83].

1.5.1 Gene knockout

Fungal natural products are in many cases discovered by genome analysis. Once a gene cluster of interest is identified, the analysis of its product can by carried out by stopping or reducing its expression. As a consequence, a specific signal will be present in the analytic profile of the wild strain of the organism but absent in the knockout strain as shown in Figure 11. This feature can then be attributed to a specific secondary metabolite. The identified natural product can then be purified for characterization and structure elucidation, for example by nuclear magnetic resonance (NMR) or mass spectrometry (MS) based methods. The identification of the natural product is, however, often complicated by the presence of highly abundant primary metabolites or constituents of the cultivation media [82, 84].

By means of the genome mining approach described in chapter 1.4, several highly promising candidates for fungal RiPPs could be identified. mRNAs encoding the corresponding predicted precursors have already been detected. The aim of this work is therefore to verify that the mRNAs are translated into the peptide RiPP precursor.

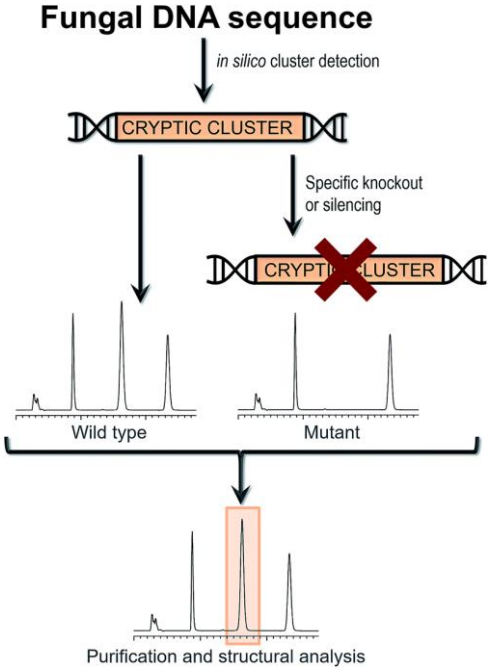

Figure 11: Identification of fungal natural products by comparing wild type and deletion strain metabolic profiles [84].

1.5.2 Extraction of RiPPs

The vast chemical diversity between and within the various classes of secondary metabolites results in different physicochemical properties (e.g. solubility) which makes the quantitative extraction of all secondary metabolites from a fungus impossible. Sample preparation thus plays an important role in secondary metabolite profiling. For extraction, various solvents can be employed such as water, methanol, acetonitrile, dichloromethane, ethyl acetate, etc. [85, 86].

Post-translational modifications on the mature RiPP can range from non-polar to highly polar [60]. As the potential PTMs are unknown, different extraction solvents and sample preparation methods have to be used in order to achieve high metabolite coverage.

Another factor that has to be considered is that secondary metabolites including the final RiPP product may be secreted by the organism into the medium. If this is the

case, extraction can be directly performed from the medium where the organism is cultivated. The medium can be clarified first by filtration or centrifugation to remove unwanted components such as cells or cell debris [86].

1.5.3 Analysis via liquid chromatography coupled to mass spectrometry

For the discovery and analysis of natural products, a variety of analytical platforms exist, among which mass spectrometry is the most widely used technique [87]. Although direct injection of samples on high-resolution mass spectrometers is possible and provides high analytical throughput [88], chromatographic separation is often carried out prior to MS analysis which can reduce ion suppression effects due to matrix compounds and also separate isomers. Separation is usually performed on reversed-phase columns. For very complex samples, conventional reversed-phase HPLC separation is often unsatisfactory as metabolites are not sufficiently resolved. By using smaller particle sizes such as 1.7 µm-particles used in ultra-performance liquid chromatographic (UPLC), chromatographic resolution and peak capacity can be increased even further. Even though reversed-phase stationary materials are the standard tool for separating non-polar and medium polar metabolites, highly polar analytes are not retained on classical RP columns. Hydrophilic interaction chromatography (HILIC) offers an interesting alternative [83]. HILIC employs polar materials as stationary phase. This can either be bare silica or polymeric carrier material modified with polar functional groups. Water-miscible organic mobile phases such as acetonitrile with a small percentage of aqueous solvent are typically employed. HILIC separations can either be carried out in an isocratic way with a high amount of organic solvent or a gradient can be used which starts with a high percentage of organic solvent and increases the content of aqueous solvent, thus eluting polar analytes. The separation mechanism of HILIC is not yet fully understood, however, it is assumed that retention is caused by partitioning. A water-rich layer forms on the hydrophilic stationary phase. The analytes then distribute differentially between the acetonitrile-rich mobile phase and the immobilized water layer. As the partitioning equilibrium is shifted towards the aqueous layer adsorbed on the stationary phase, very polar analytes will be situated more preferably in the water layer and are thus more strongly retained [89].

After chromatographic separation, analytes are introduced into the mass spectrometer, usually via ESI or atmospheric pressure chemical ionization (APCI). As metabolites can either be positively or negatively charged, it is important to perform ionization in positive and negative mode in order to cover a broad range of the metabolome. To perform structural elucidation, hybrid instruments are typically used such as Q-TOF-MS or ultra-high resolution FT-ICR-MS. Precursor ions are selected by the scanning device and fragmented in the collision cell resulting in MS/MS spectra with high mass resolution. The mass information on precursor and fragment ions facilitates the interpretation of mass spectra [83].

In recent years, ion mobility-mass spectrometry has been increasingly applied in the field of natural product research. In this hyphenation, ion mobility spectrometry (IMS) provides an additional but complementary separation step which also enables separation of isomeric analytes which cannot be resolved by mass spectrometry alone. Moreover, the collisional cross section obtained from the ion mobility offers a descriptor for analyte characterization in addition to its retention time, mass-to-charge ratio and fragmentation pattern which increases specificity in structural identification as well as molecular information content. Measured CCS values have been found to be very reproducible, making IMS a robust technique across multiple analytical platforms, samples and timeframes [90, 91].

1.5.4 Total proteome analysis

As already mentioned above, conventional proteomic analysis techniques are not applicable to identification and analysis of mature RiPPs due to their complex structures. However, a total proteome analysis can still give important insights into the pathway of RiPP biosynthesis by identifying tailoring enzymes predicted by genome mining [82]. Even though the final RiPP product cannot be detected by traditional proteomic analysis, the linear, non-modified precursor/leader peptide should in theory be detected by MS/MS analysis and subsequent database search.

2 Experimental Part

2.1 Cell culture

2.1.1 2D cell culture

LX-2 human hepatic stellate cells were cultured at 37 °C and 5 % CO_2 in Dulbecco's Modified Eagle's Medium (DMEM) with phenol-red and supplemented with 10 % fetal bovine serum (FBS) and 2 mM glutamine. Confluent cells were rinsed with PBS, trypsinized and centrifuged at 800 g for 5 min. Media was removed and the cell pellet was washed twice with ice-cold PBS. After centrifuging at 1,000 g for 5 min, PBS was removed and the cell pellet was lysed as described in chapter 2.2.

2.1.2 3D cell culture

For the preparation of spheroids, A549 lung cancer cells were cultured at 37 °C and 5 % CO_2 in Roswell Park Memorial Institute-1640 (RPMI-1640) medium with phenol-red and supplemented with 10 % FBS and 2 mM glutamine. Confluent cells were trypsinized and transferred into an Eppendorf tube. Cell number was determined by EVE automated cell counter (NanoEntek). A cell suspension with a concentration of 10^5 cells/mL was prepared and 10^4 cells (corresponding to 100 µL cell suspension) were pipetted into each well of a 96-well-plate. The plate was centrifuged for 20 min at 311 g at room temperature to achieve spheroid formation on the bottom of each well. The spheroid culture was maintained in an incubator at 37 °C and 5 % CO_2 for 10 days. 50 µL of pre-warmed medium were added on day 4 and day 6. On day 10, spheroids were collected in an Eppendorf tube. Samples containing 1 spheroid, 3 spheroids and 5 spheroids were prepared in triplicates using a 200 µL pipette. The spheroids were then centrifuged for 5 min at 1,200 g at room temperature and washed twice with cold PBS. After the last washing step, the supernatant was removed and the spheroids were lysed in lysis buffer for the preparation according to the in-solution method described in 2.2.1 as well as the optimized SP3 method described in chapters 2.2.2 and 2.2.2.3.

2.2 Proteomics sample preparation

Cell lysis as well as denaturation, reduction and alkylation of proteins was performed by adding either 500 µL of lysis buffer provided by the iST sample preparation kit from PreOmics or 100 mM Tris-HCl at pH 8.5 containing 1 % SDS, 10 mM TCEP and 40 mM chloroacetamide (CAA). Samples were then sonicated for 5 s at 10 % intensity and subsequently placed on a heating block at 95 °C for 10 min.

Protein concentration of the cell lysate was estimated using Bicinchoninic Assay (BCA) from Thermo Fisher Scientific. Bovine serum albumin (BSA) standards with concentrations ranging from 2,000 to 25 µg/mL and an additional blank were prepared. 25 µL of each standard as well as of the unknown sample were pipetted into a microplate well. Then, 200 µL of BCA working reagent (50:1 of BCA Reagent A and B) were added to each well and mixed thoroughly on a plate shaker for 30 s. The plate was covered and incubated at 37 °C for 30 min and cooled to room temperature. Absorbance was measured at 562 nm.

Samples were then subjected to different sample preparation methods which were performed in duplicates or triplicates.

2.2.1 In-solution protocol

The in-solution protocol for sample preparation consists of three steps which are described below.

2.2.1.1 Acetone precipitation

For precipitation of the proteins, 4 volume equivalents of 80 % acetone containing 20 mM sodium chloride were added to 1 volume of cell lysate (e.g. 10 µL of cell lysate containing 20 µg proteins were mixed with 40 µL of 80 % acetone) and incubated for 2 min at room temperature. Samples were centrifuged at 20,817 g for 10 min. The supernatant was discarded and proteins were washed with 4 volume equivalents of 80 % acetone followed by another centrifugation step. The supernatant was again removed and the pellet was dried for 5 min at room temperature.

2.2.1.2 Digestion

Digestion was performed by adding 120 µL of digestion buffer (100 mM ammonium bicarbonate, 10 % trifluoroethanol) containing trypsin and Lys-C (1:50) to the protein pellet in a ratio of 1:50 (wt/wt) enzyme to protein, i.e. 0.4 µg of trypsin and 0.4 µg of Lys-C for 20 µg protein input. Samples were digested at 37 °C at 550 rpm overnight.

2.2.1.3 Desalting

After digestion, samples were adjusted to a volume of 200 µL containing 1 % trifluoroacetic acid (TFA) and loaded onto styrenedivinylbenzene- reverse phase sulfonate (SDB-RPS) StageTips (200 µL tip packed with two layers of SDB-RPS material) which were placed onto 2 mL tubes. Tubes were then centrifuged at 1,500 g for 5 min at room temperature. The tips were washed with 200 µL of 0.2 % TFA and centrifuged again. Peptides were eluted with 120 µL of 5 % ammonium hydroxide in 80 % acetonitrile into a glass vial insert by centrifuging at 1,500 g for 5 min. Samples were dried in a vacuum centrifuge at 45 °C until they were completely dry. Finally, peptides were dissolved in running buffer (2 % acetonitrile, 0.1 % formic acid) for nanoLC-IMS-MS/MS analysis. 500 ng of peptides were injected for analysis.

2.2.2 SP3 protocol

For the sample preparation according to the SP3 method [6], proteins were diluted in a total volume of 48 µL of Tris-HCl buffer (50 mM, pH 8.1). SP3 beads were added to the protein solution in a ratio of 10:1 (beads/protein, wt/wt), i.e. when using 10 µg of protein, 100 µg of SP3 beads were added. However, for experiments with low starting amount of protein (below 5 µg), a minimum bead concentration of 0.5 µg of beads per 1 µL of processing volume was maintained.

Homogenization of the solution was achieved by gentle pipetting. Binding of the proteins to the beads was induced by addition of 50 µL ethanol and incubation in a ThermoMixer at 24 °C for 5 min at 1,000 rpm. The tubes were then placed on a magnetic rack and incubated until the beads have formed a pellet on the tube wall. The supernatant was carefully discarded. The beads were then rinsed twice with 180 µL 80 % ethanol by pipetting up and down. After incubation in the magnetic rack, the supernatant was removed.

100 µL of digestion solution (100 mM ammonium bicarbonate) containing trypsin and Lys-C in a ratio of 1:50 (wt/wt) enzyme to protein were added to the beads. Samples were sonicated in a water bath for 30 s before incubation in a ThermoMixer at 37 °C at 1,000 rpm overnight.

After digestion, tubes were centrifuged at 20,000 g for 1 min. The tubes were placed on a magnetic rack to remove the supernatant to a fresh tube followed by evaporation in a vacuum centrifuge. Lastly, peptides were dissolved in running buffer (2 % acetonitrile, 0.1 % formic acid) for nanoLC-IMS-MS/MS analysis.

Moreover, different approaches were chosen to optimize sample preparation which are described in the following sections.

2.2.2.1 Double amount of beads

One approach included the use of twice the amount of SP3 beads. Therefore, 400 µg of SP3 beads were added to the solution containing 20 µg protein.

2.2.2.2 Dilution of the digestion buffer

In order to reduce the salt amount, a 1:4 dilution of the digestion buffer was prepared with a final concentration of 25 mM ammonium bicarbonate, in which the digestion was carried out.

2.2.2.3 StageTip peptide cleanup

Peptide desalting can be performed on StageTips. For this purpose, the supernatant obtained after digestion was acidified by adding 1 µL TFA and loading it onto an SDB-RPS StageTip. Desalting, elution of the peptides and preparation for LC-MS/MS analysis was performed as described in chapter 2.2.1.3. Peptides were dissolved in running buffer (2 % acetonitrile, 0.1 % formic acid).

2.2.2.4 On-bead peptide cleanup

Desalting can be carried out directly on-bead by washing the peptides bound to the SP3 beads with different solvents. Here, acetonitrile and isopropanol were used. To obtain final concentrations of 95 % organic phase, 1.9 mL solvent were added to the samples after digestion. The samples were then centrifuged at 20,000 g for 1 min. The tubes were placed on a magnetic rack and incubated for 5 min. The supernatant was

carefully removed into a fresh tube and kept for LC-MS/MS analysis to see if peptides were lost during the washing step. Beads were then rinsed with 180 µL solvent.

Rinsed beads were reconstituted in 150 µL water by pipetting up and down and sonicating the tubes in a water bath for 30 s. After centrifuging at 20,000 g for 1 min, the tubes were placed on a magnetic rack, the supernatant was transferred into a glass insert and dried in a vacuum centrifuge. Peptides were finally dissolved in running buffer (2 % acetonitrile, 0.1 % formic acid) for nanoLC-IMS-MS/MS analysis.

2.2.2.5 Peptide cleanup on fresh beads

Peptide cleanup was also performed on fresh beads. For this purpose, the supernatant was transferred into a new tube after digestion. After addition of 200 µg of SP3 beads, 1,900 µL acetonitrile were added to obtain a final concentration of 95 % organic phase. Tubes were centrifuged at 20,000 g for 1 min. The supernatant was removed by placing the tubes on a magnetic rack. Beads were rinsed with 180 µL acetonitrile. Rinsed beads were then reconstituted in water as described in 2.2.2.4. After drying the samples, peptides were dissolved in running buffer (2 % acetonitrile, 0.1 % formic acid).

2.2.3 iST sample preparation kit

Sample preparation was carried out according to the instruction of the iST sample preparation kit [92].

20 µg of protein were pipetted into the cartridge which was placed in a waste tube.

210 µL of resuspension buffer were added to the digestion buffer, shaken for 10 min (room temperature, 500 rpm) and pipetted up and down. 50 µL of so-obtained digestion solution was pipetted into the cartridge and placed in a heating block for 3 hours (37 °C, 500 rpm). 100 µL of STOP solution were added to the cartridge which was then shaken for 1 min (room temperature, 500 rpm).

The cartridge was spun in a centrifuge at 3,800 g for 3 min. 200 µL of WASH 1 were added to the cartridge which was centrifuged again before adding 200 µL of WASH 2. After centrifuging, the cartridge was placed into a fresh tube and 100 µL elution solution were added to the cartridge. The flow-through was collected in the tube which was then placed into a vacuum evaporator at 45 °C until completely dry. Lastly, peptides were

dissolved in 20 µL running buffer (2 % acetonitrile, 0.1 % formic acid) for nanoLC-IMS-MS/MS analysis.

2.3 Sample preparation for the identification of RiPPs

Different strains of *T. reesei* (wildtype and knockout strains) were cultivated in medium with and without peptone. For the identification of potential RiPP precursors or leaders, supernatants from the fungal cultures as well as the mycelia were provided.

Identification of ustiloxin B and asperipin-2a from *Aspergillus flavus* was carried out using growth media and mycelial samples obtained from overexpression and knockout strains.

2.3.1 Sample preparation from medium and supernatant

2.3.1.1 Polar extract

Two different ways for sample preparation from the growth medium itself and the supernatant from the fungal culture were performed. First, 200 µL medium or supernatant were mixed thoroughly with 400 µL methanol and 400 µL acetonitrile and centrifuged at 20,817 g for 10 min at room temperature. 200 µL of the supernatant were then used directly for separation on a HILIC column followed by mass spectrometry analysis. Further 200 µL were evaporated in a vacuum centrifuge at 45 °C until dry. The residue was then dissolved in 150 µL running buffer (2 % acetonitrile, 0.1 % formic acid) for separation on a reversed-phase column followed by mass spectrometry analysis.

2.3.1.2 Acetone precipitation

For the second approach, 800 µL acetone and 20 mM sodium chloride were added to 200 µL of growth medium and supernatant of the culture for protein precipitation. Samples were incubated for 2 min at room temperature followed by centrifugation at 20,817 g for 10 min at room temperature. The supernatant was discarded and the pellet was washed once with 800 µL 80 % acetone. After another centrifugation step, the pellet was dissolved in 200 µL 1 % TFA, loaded onto an SDB-RPS stage tip and desalted as described in 2.2.1.3. The residue was finally dissolved in 150 µL running buffer (2 % acetonitrile, 0.1 % formic acid) for analysis on a reversed-phase column.

2.3.2 Sample preparation from the mycelium

For the preparation of mycelial samples, two different approaches were employed as described in the following chapters.

2.3.2.1 Polar extract

Extraction of secondary metabolites was carried out by adding 1 mL of a mixture of water, acetonitrile and methanol in a ratio of 1:2:2 to 100 mg of mycelium and subsequent homogenization on a VWR bead mill MAX using glass beads with a diameter of 0.5 mm. The samples were milled for 2 min at room temperature at a speed of 6 m/s followed by a sonication step at 10 % intensity for 10 s. Cell debris and glass beads were pelleted by centrifugation at 20,000 g for 5 min.

200 µL of the supernatant were directly subjected to LC-MS/MS measurement on a HILIC column. Another 200 µL were evaporated in a centrifugal evaporator before dissolving the residue in 150 µL running buffer (0.1 % formic acid, 2 % acetonitrile) for LC-MS/MS and nanoLC-IMS-MS/MS analysis.

For the digestion of peptides and proteins, the mycelium was extracted as described above. After evaporation of the supernatant, the digestion was performed using pepsin and chymotrypsin, respectively. For the pepsin digestion, the residue was dissolved in 25 µL 1 % TFA and digested for 2 hours at 37 °C. For the digestion with chymotrypsin, the residue was dissolved in Tris-HCl (50 mM, pH 8.1) containing 10 mM calcium chloride and digested over night at 30 °C.

Samples were then adjusted to a volume of 200 µL containing 1 % TFA and desalted as described in 2.2.1.3. Peptides were dissolved in 20 µL running buffer (0.1 % formic acid, 2 % acetonitrile) for nanoLC-IMS-MS/MS analysis.

2.3.2.2 Proteomics workflow

Lysis of 100 mg of mycelium was performed in 1 mL lysis buffer (100 mM Tris-HCl, 1 % SDS, 10 mM TCEP, 40 mM CAA) by bead milling and sonication as described in the previous section. The supernatant was transferred to a new tube and incubated at 95 °C for 10 min to achieve reduction and alkylation of proteins. Protein amount was then estimated using BCA assay.

50 µg of protein were subjected to acetone precipitation and tryptic digestion followed by desalting as described in 2.2.1.1, 2.2.1.2 and 2.2.1.3 respectively. Peptides were finally dissolved in running buffer (2 % acetonitrile, 0.1 % formic acid) and subjected to nanoLC-IMS-MS/MS analysis.

2.4 LC-MS/MS analysis

LC-MS/MS analyses were performed on different LCs equipped with either HILIC- or RP-columns coupled to a Bruker timsTOF Pro mass spectrometer.

2.4.1 LC-IMS-MS/MS method for proteomics

Samples for proteomics were analyzed on an UltiMate 3000 RSLCnano UHPLC system coupled to a timsTOF Pro ion mobility mass spectrometer via a nano-electrospray ion source.

Chromatographic separation of peptides was carried out on an Ionopticks Aurora series 25 cm x 75 µm ID, 1.6 µm C18 column with pulled emitter tip. Solvents were H_2O + 0.1 % formic acid (A) and acetonitrile with 0.1 % formic acid (B) at 50 °C with a flow of 0.4 µL/min and the following gradient: 2 % to 17 % B from 5.5 to 65.5 min, 17 % to 25 % B from 65.5 to 95.5 min, 25 % to 37 % B from 95.5 to 105.5 min, 37 % to 95 % B from 115.5 to 125.5 min, 2 % B from 126.5 to 136.5 min.

Peptides eluting from the column were ionized in positive ESI mode by the TIMS-TOF MS. The mass spectrometer was operated in PASEF mode with a scan range from 100 to 1,700 m/z.

2.4.2 LC-(IMS)-MS/MS methods for the analysis of RiPPs

Samples prepared from fungal media and mycelia were analyzed on an Elute high performance liquid chromatography system. Chromatographic separation in HILIC mode was carried out on a SeQuant ZIC-pHILIC column (5 µm, 150 × 2.1 mm), reversed-phase separation was performed on an Intensity Solo 2 C18 RP (100 × 2.1 mm) column.

HILIC separation was carried out at a flow of 200 µL/min, either at high pH or low pH. For separation at high pH, ammonium bicarbonate (20 mM, pH 9.2) was used as buffer A and acetonitrile as mobile phase B, whereas water with 0.1 % formic acid and

acetonitrile with 0.1 % formic acid were used as solvents A and B at low pH. The gradient for separation on the HILIC column started with 10 % A from 0 to 2 min, followed by a linear increase to 80 % A between 2 and 10 min. 80 % A was maintained until minute 15 and was decreased to 10 % between 15 and 15.2 min. 10 % A was held until minute 20.

Separation on the RP column was performed using H_2O with 0.1 % formic acid (A) and acetonitrile with 0.1 % formic acid (B) as running buffers at a flow rate of 400 µL/min and the following gradient: 0 to 2 min: 2 % B, 2 to 15 min: 2 % to 90 % B, 15 to 18 min: 90 % B, 18 to 18.1 min: 90 % to 2 % B, 18.1 to 20 min: 2 % B.

Analytes were ionized both in positive and negative ESI mode by the timsTOF pro. For detection of smaller analytes, the MS scanned a mass range from 20 to 1,300 m/z in positive and negative polarity.

To cover a wider range of masses and to make use of ion mobility separation, a PASEF method in positive mode for mass spectrometric analysis was employed, scanning a mass range of 100 to 1,700 m/z.

Mass calibration and ion mobility calibration were performed by injecting sodium formate and Agilent Tuning Mix ESI-TOF low concentration as calibrant between 1 and 1.2 minutes of each run.

2.5 Data analysis

Database search and data analysis for the comparison of different sample preparation strategies was carried out using PEAKS X Pro (version 10.6). For the identification of potential RiPP precursors, PEAKS X Pro, Metaboscape (version 2021b) and MaxQuant (version 1.6.17.0) were used. Statistical analysis was performed with Perseus (version 1.6.12.0).

2.5.1 RStudio

Data obtained from the comparison of different sample preparation methods was analyzed using RStudio (4.1.2). One-way ANOVA was used to determine significant differences between the different methods. Statistically different results were then evaluated using Tukey's HSD test. For SP3 optimization and spheroid experiments, paired t-tests were carried out.

2.5.2 PEAKS X Pro

The software package PEAKS X Pro was used for peptide sequencing and protein identification.

For database search, the error tolerance of the precursor mass was set to 15 ppm using monoisotopic mass and error tolerance of fragment ions was set to 0.05 Da.

Database search for samples prepared from LX-2 human hepatic stellate cells, H1299 lung cancer cells or from A549 lung cancer spheroids was carried out using the UniProt human database downloaded on 26.08.2020 (37414 sequences). For database search for samples obtained from *T. reesei*, UniProt Hypocrea Jecorina database was used (9837 sequences, downloaded on 18.03.2021). Contaminant sequences were obtained from cRAP.

The false discovery rate (FDR) for peptide and protein identification was estimated using decoy-fusion and was set to 1 %.

2.5.3 MaxQuant and Perseus

Quantitative analysis of *T. reesei* samples was carried out using MaxQuant. Methionine oxidation and N-terminal acetylation were set as variable modifications, for proteomics samples carbamidomethylation was additionally set as fixed modification. For identification, NCBI *T. reesei* fasta file containing 9,111 protein sequences (downloaded on 21.09.2021) was used. FDR was set to 1 % and match between runs was enabled in a retention time window of 0.7 min and an alignment time window of 20 min. Proteins were quantified by label free quantification using unique and razor peptides.

Statistical analysis was performed using Perseus. First, reverse proteins, proteins only identified by site and contaminants were filtered out to remove these hits from the matrix. The LFQ values were then $\log_2$ transformed, followed by grouping of samples according to replicates (in this case wildtype and knockout strains). To remove identifications with only one reported intensity, rows were filtered based on 3 valid values in at least one group. In the next step, missing values were imputed from normal distribution (downshift 1.8, width 0.3). Finally, a two-sample t-test was performed with a permutation-based FDR of 0.01, S0 was set to 2.

2.5.4 Metaboscape

LC-(IMS)-MS/MS data obtained from fungal samples were processed using Metaboscape (Bruker, Germany). Retention time alignment, deisotoping and feature extraction was performed with the software's T-ReX 3D and 4D algorithm for LC-MS/MS and LC-IMS-MS/MS data, respectively. The minimum number of occurrences for recursive feature extraction was set to 1. The intensity threshold for peak detection was initially set to 1,000 counts but reduced to 100 counts for particular samples. Retention time and mass range were adjusted to the chromatographic and mass spectrometric settings. This way, a bucket table displaying all detected features was obtained for each sample which included the corresponding retention time, molecular weight, measured m/z, detected ions and their intensity.

Annotation was performed using the Bruker HMDB Metabolite Library. Moreover, different groups and attributes were defined according to the sample types (i.e., wild type and knockout, mycelium and supernatant, etc.). For statistical analysis, a t-test was performed to compare features in different sample types. Missing values were substituted by the group mean, p-value limit was set to 0.05 and fold change limit to 2.

2.5.5 Pathway enrichment analysis

Significantly upregulated proteins in wildtype and knockout samples were subjected to pathway enrichment analysis using STRING database (version 11.5) [93].

3 Results and Discussion

3.1 Sample preparation for proteomics

In the first part of this thesis, different sample preparation methods for proteomics were compared. LX-2 human hepatic stellate cells grown in 2D cell culture were prepared according to an already established method for in-solution digestion, a recently published protocol based on paramagnetic beads called SP3 and a commercially available kit for sample preparation obtained from PreOmics.

The SP3 approach was then optimized by employing different strategies to increase the number of identified proteins as described in chapter 2.2.2. After optimizing the SP3 method on cells obtained from 2D cell culture, it was applied to 3D cultured spheroids to minimize the number of spheroids required for proteomic analysis.

3.1.1 Comparison of different methods for sample preparation

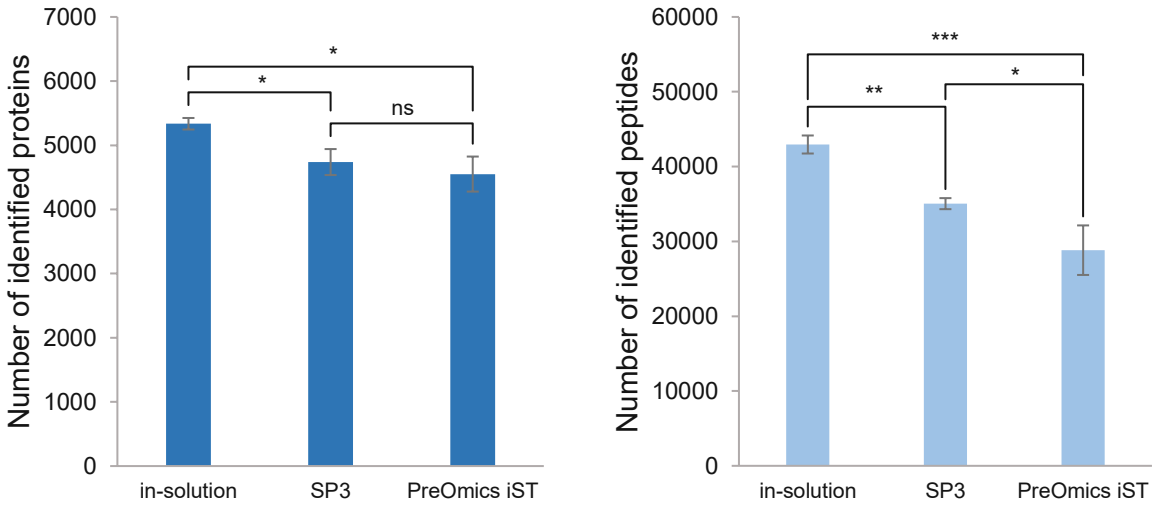

Figure 12: Comparison of the in-solution protocol, the SP3 method and the PreOmics sample preparation kit for 20 µg protein input regarding the number of identified proteins (left) and peptides (right), respectively. Sample preparation for each method was performed in triplicates. One, two and three asterisks indicate the significance level at the 0.05, 0.01 and 0.001 level, respectively, which was determined by ANOVA analysis followed by Tukey's test; ns indicates non-significant results.

For the comparison between different approaches for the preparation of proteomic samples, LX-2 human hepatic stellate cell lysate containing 20 µg of protein was processed by the in-solution protocol, SP3 and iST (PreOmics sample preparation kit).

Figure 12 shows the number of identified proteins and peptides for each of the tested methods. When using 20 µg of starting amount, all three methods showed similar performance. By applying the different protocols, 4,300 – 5,400 proteins and 26,000 – 44,000 peptides could be detected when injecting 500 ng peptides. Concerning the number of protein identifications, the in-solution method outperformed both SP3 and iST which was shown by ANOVA analysis and subsequent Tukey test. On average, 5,335 proteins were identified when processing 20 µg of starting amount using the in-solution protocol compared to 4,738 proteins identified by SP3 and 4,551 proteins detected by iST.

3.1.2 Optimization of the SP3 method

The SP3 method stands out due to its simplicity, efficiency and capability for analyzing samples derived from sub-microgram amounts of starting material [15]. Therefore, it is of interest to optimize this method in order to obtain equivalent results for SP3 as for the in-solution method regarding the number of identified proteins. Toward this purpose, different approaches and strategies were employed, which were based on using twice as many paramagnetic beads to increase protein binding to the beads and

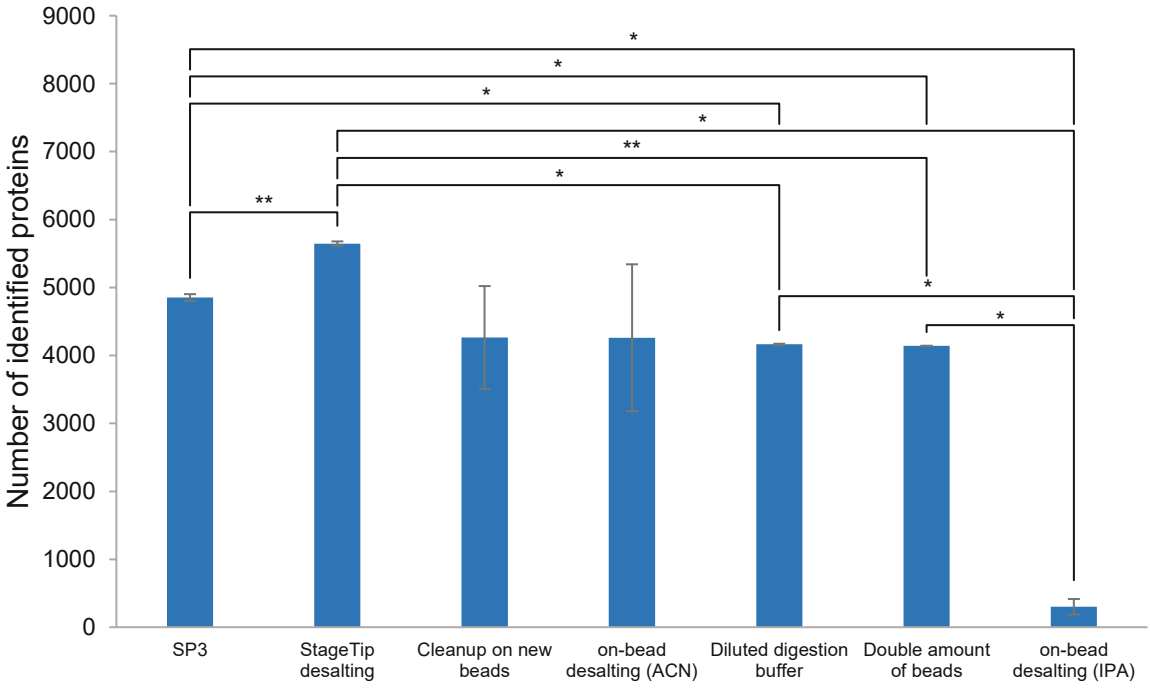

Figure 13: Comparison between different optimization approaches for the SP3 method using 20 µg starting amount. Sample preparation for each optimization strategy was performed in duplicates. One and two asterisks indicate the significance level of 0.05 and 0.01, respectively, which was determined using paired t-tests. Statistically non-significant results are omitted for clarity.

therefore minimize potential protein loss during sample preparation, on-bead peptide cleanup after digestion, peptide cleanup using StageTips, etc. (see chapters 2.2.2.1 to 2.2.2.5 for detailed description).

The numbers of identified proteins for the different optimization approach are compared in Figure 13. Desalting on StageTips was shown to be the most efficient approach for peptide cleanup, resulting in 5,566 identified proteins, on average when injecting 500 ng peptides. Compared to the conventional SP3 method, this is an increase of more than 800 proteins.

Peptide cleanup on new paramagnetic beads, on-bead desalting with acetonitrile, diluting the digestion buffer prior to proteolytic digestion and using twice as many beads (an SP3 beads/protein-ratio of 20:1 instead of 10:1) did not lead to an increase in the number of identified proteins.

For on-bead desalting with acetonitrile and isopropanol, the sample needs to be adjusted to a final volume of at least 95 % organic solvent prior to washing to promote binding of the tryptic peptides to the paramagnetic beads. When using isopropanol for this step, the number of identified proteins was below 400 indicating substantial peptide loss due to insufficient binding to the beads.

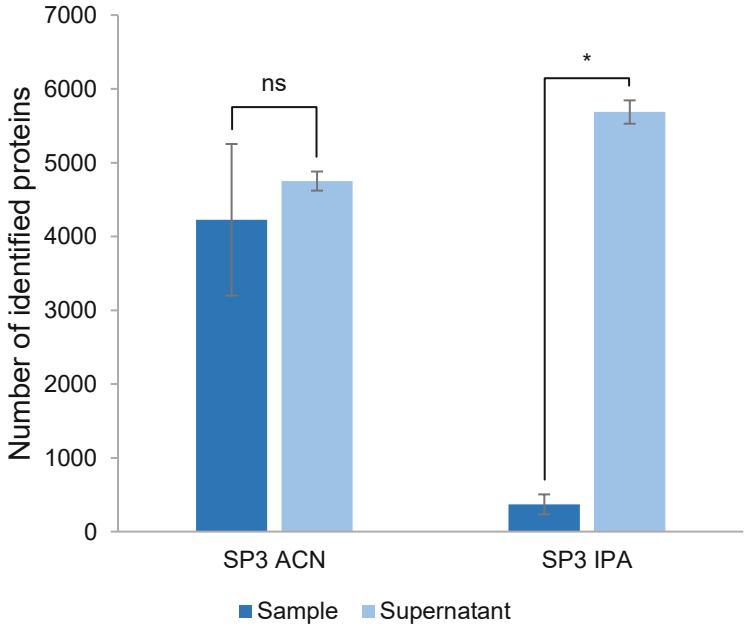

Figure 14: Number of identified proteins in the sample compared to the supernatant for on-bead desalting with acetonitrile and isopropanol, respectively. Two asterisks indicate statistical significance at the 0.05 level, ns indicates no statistical significance which was determined by paired t-tests.

Therefore, the supernatant (95 % acetonitrile and 95 % isopropanol, respectively) was retained and also analyzed by nanoLC-IMS-MS/MS to investigate peptide binding to the SP3 beads during peptide cleanup. To this end, 500 ng of peptides were injected for analysis. In the supernatant obtained from the cleanup with acetonitrile, 4,750 proteins were identified, 525 proteins more than in the original sample, as shown in Figure 14. When washing with isopropanol, 5,749 proteins were detected in the organic phase. These results clearly show that under these conditions, peptides are not bound on the beads but are rather eluted into the organic phase while salts and other contaminants do not dissolve in the organic phase. As a result, salts are removed from the peptides which leads to an increased number of identified proteins during mass spectrometric analysis. This effect is even more pronounced when using isopropanol as organic solvent where an even higher number of proteins can be identified than with the optimized SP3 approach with an additional desalting step on StageTips.

3.1.3 Application of the in-solution method and SP3 to spheroids

It was demonstrated that the SP3 method enables identifying a high number of proteins when peptides are desalted on StageTips prior to mass spectrometry analysis. With this approach, the number of identified proteins is comparable to the conventional in-solution method.

Spheroids obtained from 3D cell culture were therefore subjected to the in-solution method as well as the optimized SP3 protocol followed by StageTip desalting. For the experiments, 1, 3 and 5 spheroids were pooled and processed according to the in-solution protocol as well as the optimized SP3 method to investigate the dependency of the number of identified proteins on the number of spheroids used for analysis. Each experiment was performed in triplicates resulting in a total number of 18 samples. The three remaining spheroids were each subjected to the conventional SP3 protocol.

As the number of spheroids obtained from the 96-well plate used for cultivation was limited, no protein quantification could be performed. However, results from preliminary protein estimations of spheroids have shown that 1 spheroid contains roughly 5-10 µg protein. The protein amount thus ranges from 5-10 µg for one spheroid to 25-50 µg for five spheroids. For LC-IMS-MS/MS analysis, roughly 500 ng of peptides were injected.

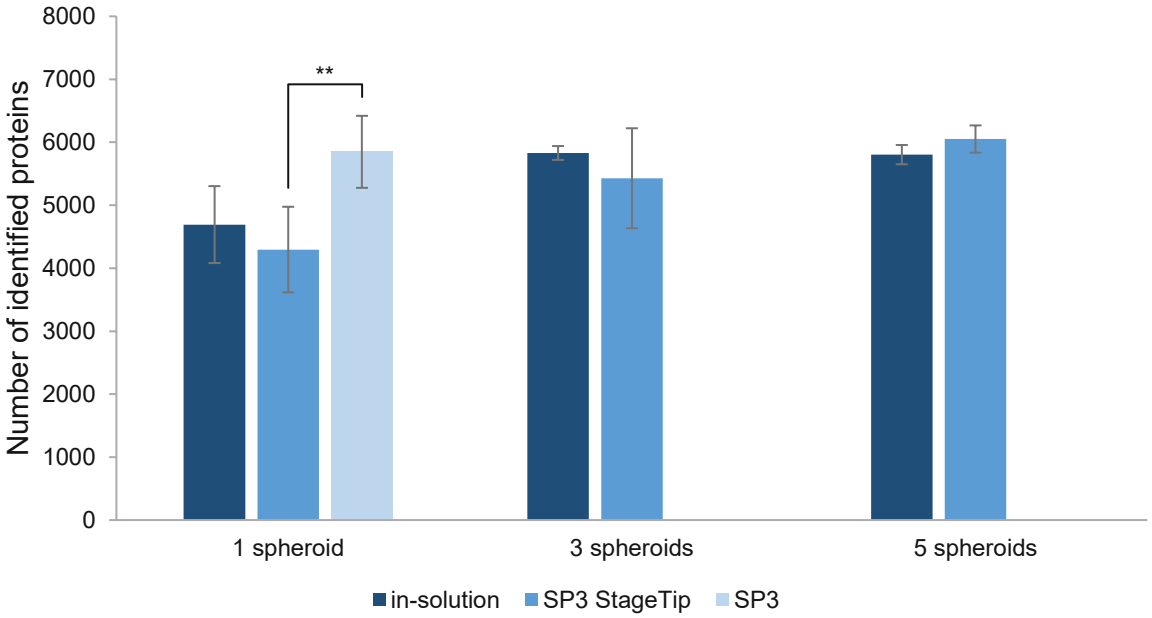

Figure 15: Number of identified proteins obtained from different numbers of spheroids used for both in-solution and SP3 method when injecting 1 µg of peptides for nanoLC-IMS-MS/MS analysis. Two asterisks indicate the significance level of 0.01 which was determined using paired t-tests. Statistically non-significant results are omitted for clarity.

Results obtained from the sample preparation of spheroids are displayed in Figure 15. No significant differences were observed between the in-solution method and the optimized SP3 approach.

Interestingly, the unmodified SP3 method shows significantly higher numbers of identified proteins compared to the optimized SP3 method when using 1 spheroid which corresponds to a protein input of roughly 5-10 µg. This suggests that little protein input also leads to lower salt concentrations in the sample. In this case, the desalting step is dispensable and should be omitted as it leads to peptide loss rather than increasing the number of protein identifications. However, when a higher starting amount of protein is used, the desalting step becomes more relevant, as higher protein input increases the salt amount in the sample which interferes with mass spectrometry analysis.

3.1.4 Performance of SP3 in dependence of protein input

The results shown in the previous chapter suggest, that the non-modified SP3 approach yields better results for low protein amounts whereas the optimized SP3 method is preferable for higher protein inputs. Therefore, the performance of conventional SP3 was compared to the optimized SP3 approach with StageTip

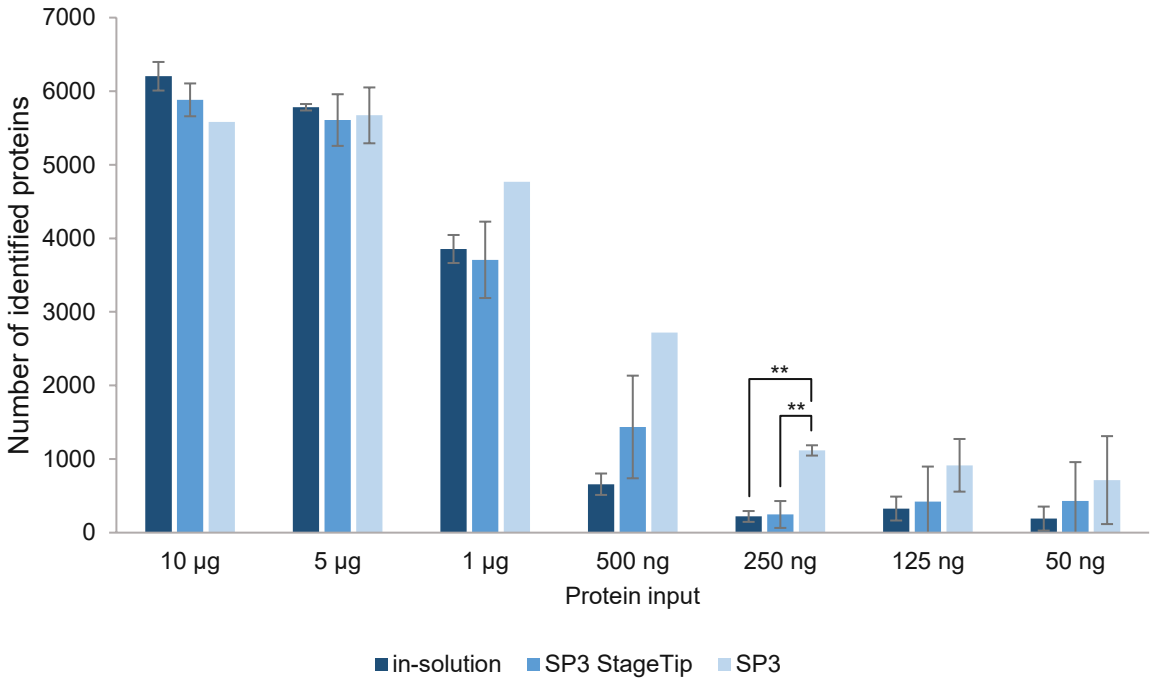

Figure 16: Numbers of identified proteins obtained by the in-solution method, the optimized SP3 protocol which contains a desalting step on StageTips and the conventional SP3 method. Analysis of each concentration was carried out in triplicates for each method, except for the following samples: 10 µg SP3 (single sample), 5 µg SP3 (duplicate), 1 µg SP3 (single sample), 500 ng SP3 (single sample), 250 ng SP3 (duplicate), 125 ng SP3 (duplicate), 50 ng SP3 (single sample), 50 ng SP3 StageTip (duplicate). Two asterisks indicate the significance level of 0.01 which was determined using ANOVA followed by Tukey's test. Statistically non-significant results are omitted for clarity.

desalting and to the in-solution method with protein amounts ranging from 10 µg to 50 ng. From samples obtained with initial protein amounts of 10, 5 and 1 µg, 500 ng of peptides were injected for nanoLC-IMS-MS/MS analysis. In case of initial protein amounts of 500, 250, 125 and 50 ng, column loads were 250, 125, 62.5 and 25 ng, respectively.

The number of identified proteins obtained with the in-solution method, the optimized SP3 approach and the conventional SP3 method in dependence of the protein input used are displayed in Figure 16. When processing 500 ng to 10 µg of protein, the in-solution method and the optimized SP3 approach showed similar performance. The performance of the conventional SP3 method could not be evaluated for protein inputs of 10 µg, 1 µg and 500 ng as only a single sample could be analyzed for these protein amounts. Interestingly, these results contrast with previous results (Figure 13) where higher numbers of identified proteins were yielded due to the additional desalting step. For 250 ng of protein, the conventional SP3 method shows a significantly higher number of identified proteins.

These results suggest that for protein amounts below 500 ng the cleanup step included in the optimized SP3 protocol is not necessary and leads to peptide loss instead of increasing the number of identified proteins.

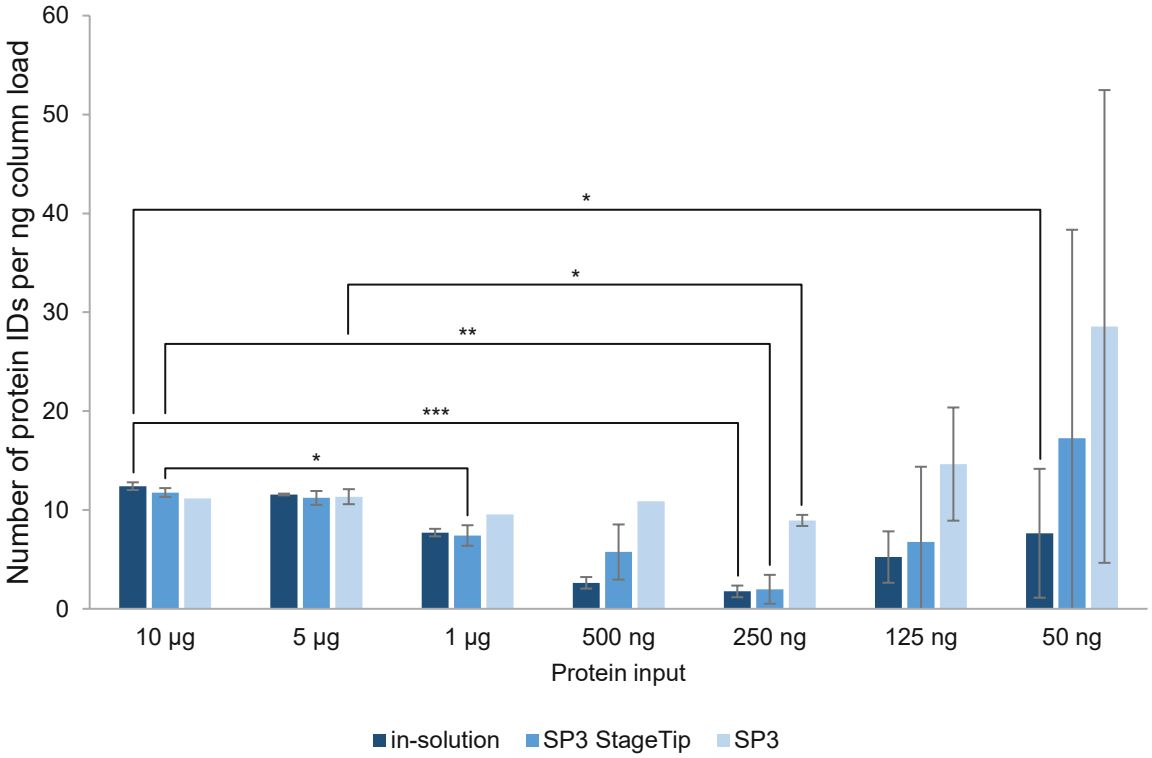

Figure 17: Numbers of identified proteins per nanogram column load for the in-solution method, the optimized SP3 approach and the conventional SP3 method for each protein input. One, two and three asterisks indicate statistical significance at the 0.05, 0.01 and 0.001 level, respectively, which was determined by paired t-tests. Statistically non-significant differences are not shown for reasons of clarity.

Figure 17 depicts the number of identified proteins per nanogram loaded on column. A decrease of protein identifications was observed between 10 µg and 250 ng when performing the in-solution method and the optimized SP3 approach. A reduced number of protein IDs was also determined for the conventional SP3 method when processing 250 ng of proteins compared to 5 µg of proteins. While the number of protein identifications significantly decreases between 10 µg and 50 ng when employing the in-solution protocol, no efficiency loss is observed for the conventional and optimized SP3 method in this concentration range.

3.2 Identification of RiPPs

The second part of this thesis deals with the identification and analysis of RiPPs in the fungus *T. reesei*.

3.2.1 Analysis of supernatants

First, the growth medium itself and the supernatants of the *T. reesei* culture were analyzed using different methods for sample preparation as well as different LC columns and MS acquisition modes to obtain a general idea about their metabolic profiles. Furthermore, supernatants from cultivations of *T. reesei* wild type and knockout strains were compared.

3.2.1.1 Comparison of medium and supernatant

In a first analysis, pure growth medium and supernatant obtained from the *T. reesei* culture were prepared according to the methods described in chapter 2.3.1, followed by separation on a reversed-phase column and MS/MS analysis.

Data processing resulted in a list of 71,275 buckets in total, in which 21 features could be annotated to a specific compound employing the Bruker HMDB database. In Figure 18, a volcano plot of the detected buckets is depicted which shows differences between the growth medium and the supernatant. A large number of compounds are more

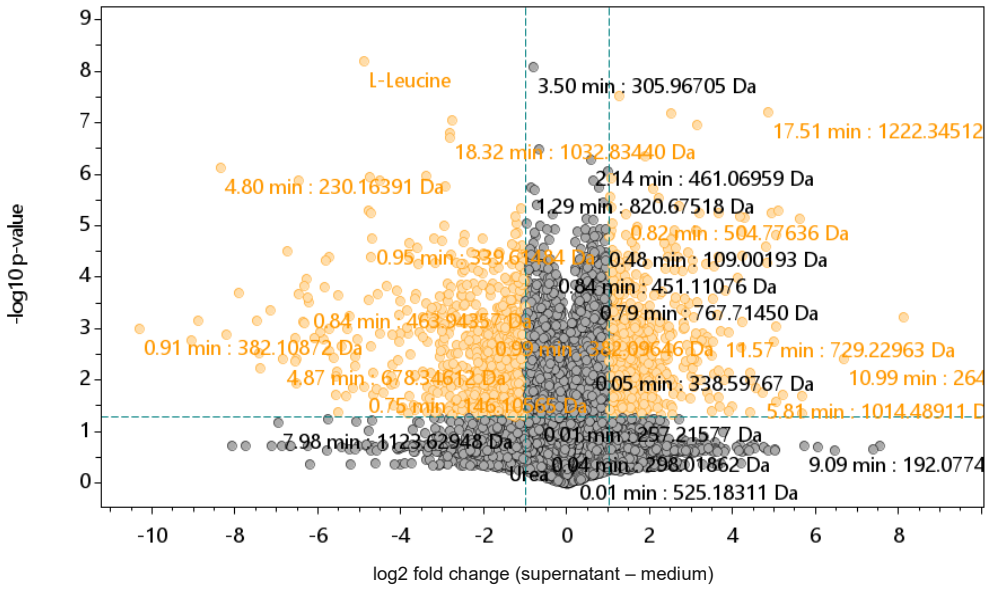

Figure 18: Volcano plot showing differences in detected compounds between pure growth medium and supernatant supplemented with peptone (p-value: 0.05, fold change limit: 2). Samples were separated on a reversed-phase column and mass spectrometric analysis was performed in positive ionization mode without employing ion mobility.

abundant in the medium (left side of the volcano plot) which suggests that they are metabolized by the fungus during growth and are thus less abundant in the supernatant. Moreover, several compounds with molecular masses higher than 500 Da were detected in both medium and supernatant. These compounds are most likely peptides from the peptone which was added to the growth medium. Proteolysis by extracellular proteases might also result in the presence of smaller peptides in the supernatant, in contrast to the pure medium. This shows that the supplementation with peptone leads to a high background which makes finding features corresponding to RiPPs quite challenging.

Therefore, cultivation of *T. reesei* was also performed in medium without peptone to reduce the background. Medium and supernatant were prepared and analyzed as described above. Figure 19 shows the base peak chromatograms obtained from the supernatant and medium without peptone which were prepared according to the acetone precipitation method. A bucket list containing 47,371 buckets was obtained by data processing. The corresponding volcano plot is illustrated in Figure 20.

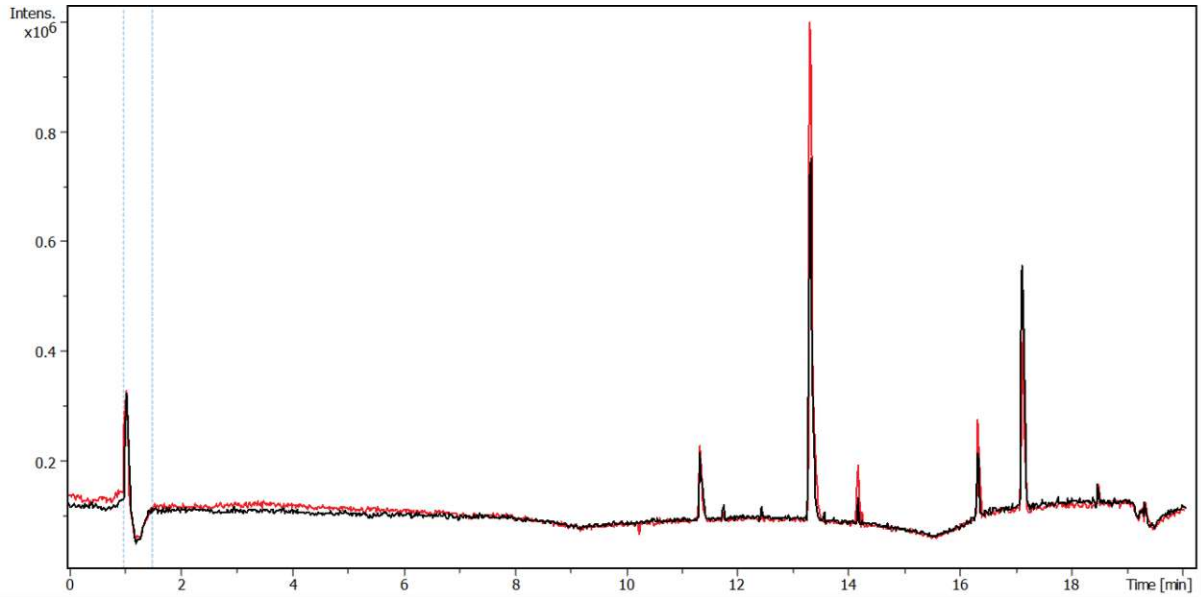

Figure 19: Base peak chromatograms obtained from the supernatant (black) and medium (red) without peptone prepared according to the acetone precipitation method. Separation was carried out on a reversed-phase column and mass spectrometry was performed in positive ionization mode without using ion mobility.

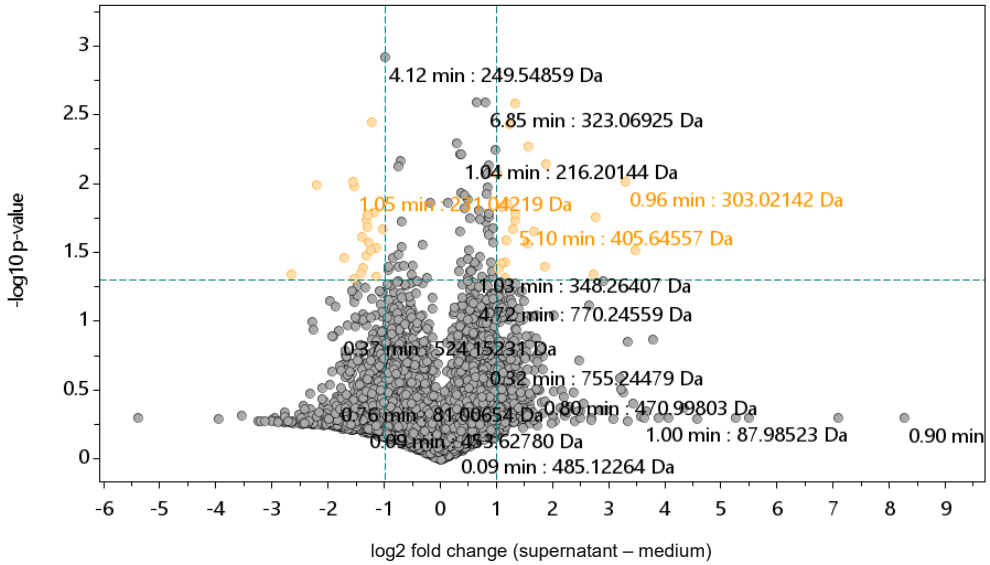

Figure 20: Volcano plot showing differences in detected compounds between pure growth medium and supernatant without peptone (p-value: 0.05, fold change limit: 2). Samples were separated on a reversed-phase column and mass spectrometric analysis was carried out in positive ionization mode without employing ion mobility.

When omitting peptone during cultivation, still a high number of features could be detected in both growth medium and supernatant. Due to the high number of detected features, it was not possible to find a potential RiPP candidate. Moreover, it is possible

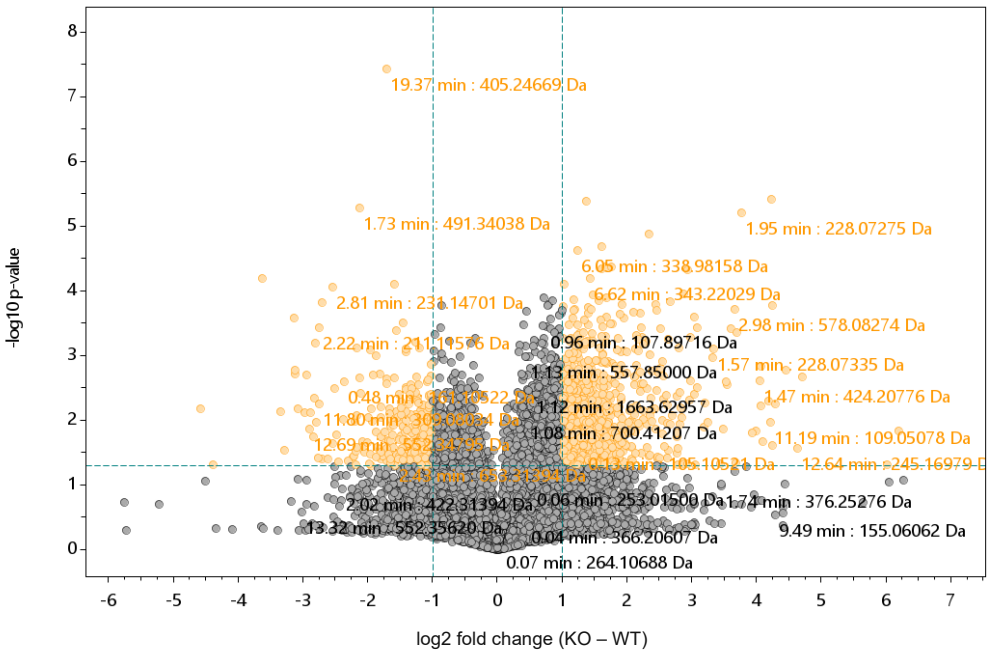

Figure 21: Volcano plot showing differences in detected features between wildtype and 120,311 knockout mycelial samples cultivated in growth medium with peptone. Samples were separated on a HILIC column and mass spectrometric analysis was carried out in positive ionization mode without using ion mobility.

that the RiPP is not secreted into the growth medium and can therefore not be detected.

Furthermore, mycelial wildtype samples and samples where the gene 120,311 which encodes for a potential RiPP was knocked out were analyzed after preparation according to chapter 2.3.2.1. Again, a large number of features was detected in wildtype and knockout samples (Figure 21). As it was at this point of time not possible to retrieve any information on successful RiPP extraction from this data, *Aspergillus flavus* strains which produce the RiPPs ustiloxin B and asperipin-2a were subjected to the same sample preparation methods and LC-MS/MS methods used for the analysis of *T. reesei* samples to find out which of these methods are suitable to extract and detect fungal RiPPs as described in chapter 3.2.4.

3.2.2 Analysis of RiPP precursor peptides

Several biosynthetic gene clusters containing putative RiPPs were already detected by genome mining. In this work, the putative RiPP precursors 120,311 and 106,662 shall be investigated. It was demonstrated by quantitative reverse transcription PCR that these RiPP precursor sequences are transcribed into mRNA. Sequence 120,311 is expressed in medium with peptone as well as in medium without peptone. Sequence 106,662 is only expressed when being cultured with peptone. To show that the detected precursor mRNAs are translated into peptides, different extraction methods and digestion strategies were employed, as described in 2.3.2.1 and 2.3.2.2, respectively. In total, 5 different samples were analyzed and compared (Table 1).

Table 1: *T.* reesei samples used for the identification of RiPP precursor peptides

Sample name	Sample type
Δtmus +P	Mycelium of wild type *T. reesei* cultivated in medium with peptone
Δtmus -P	Mycelium of wild type *T. reesei* cultivated in medium without peptone
Δ120311 +P	Mycelium of *T. reesei* where sequence 120,311 was knocked out, cultivated in medium with peptone
Δ120311 -P	Mycelium of *T. reesei* where sequence 120,311 was knocked out, cultivated in medium without peptone
Δ106662 +P	Mycelium of *T. reesei* where sequence 106,662 was knocked out, cultivated in medium with peptone

Figure 22 shows the sequence coverage of precursor 120,311 for each digestion strategy. When no digestion is performed and the extract is directly analyzed, the highest sequence coverage of 91 % is obtained. Digestion of extracted proteins with pepsin and chymotrypsin resulted in sequence coverages between 63 and 67 %.

For samples prepared according to the proteomics workflow which included protein extraction, acetone precipitation, digestion with trypsin followed by StageTip desalting and subsequent nanoLC-IMS-MS/MS analysis, the sequence coverage was only 13 %. Sequence coverages for wildtype and knockout samples obtained for different digestion strategies are summarized in Table 2.

These results clearly show that the precursor mRNAs of precursor 120,311 are translated into peptides. A simple extraction of proteins with a mixture of water, methanol and acetonitrile is sufficient to detect precursor peptides with high statistical significance. Database search can be performed easily with the database downloaded from UniProt which contains more than 9,000 proteins from the fungus *T. reesei* as well as the sequences of the predicted precursor peptides.

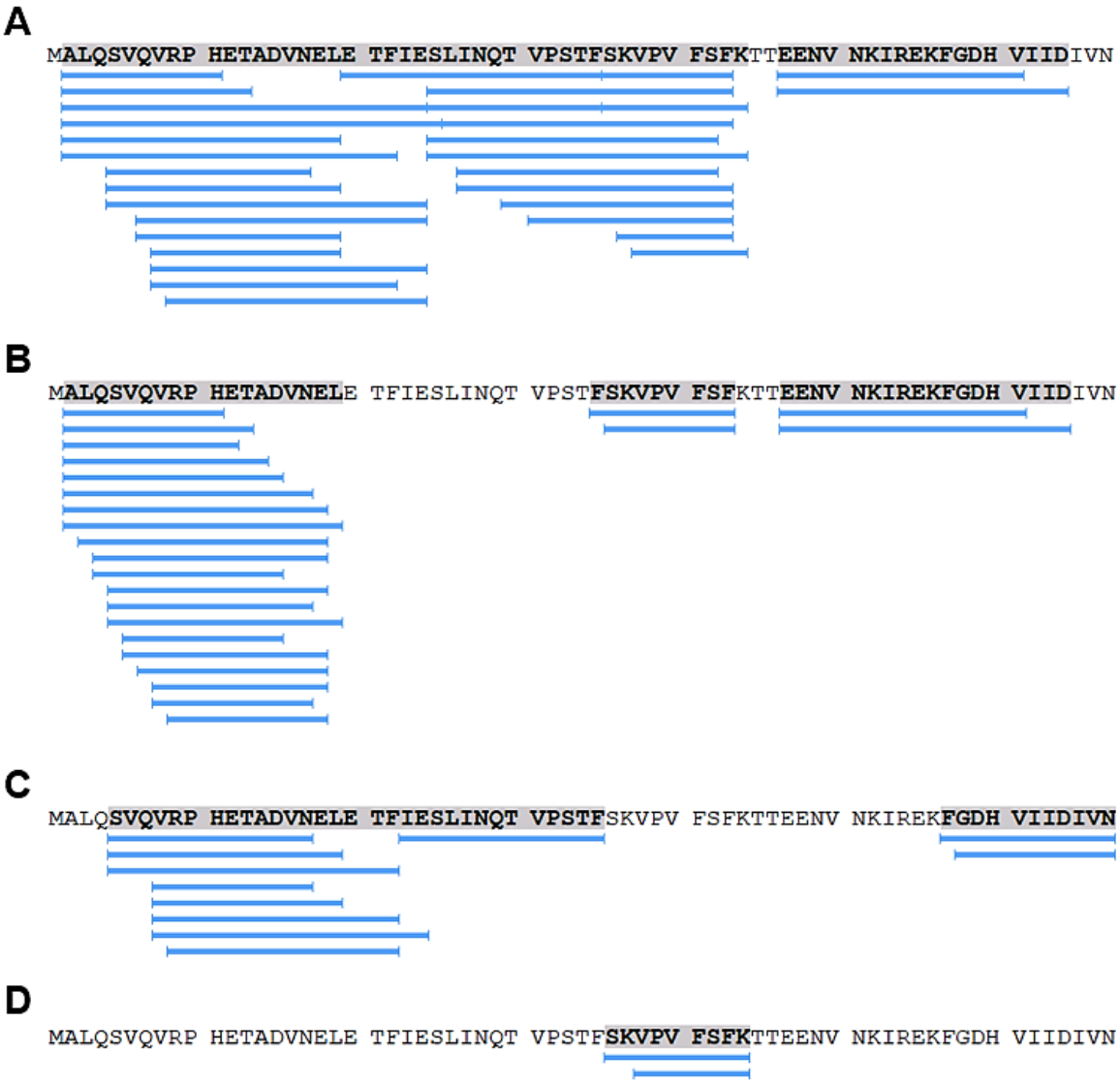

Figure 22: Sequence coverages for the precursor 120,311 obtained from (A) undigested samples, (B) digestion with pepsin, (C) digestion with chymotrypsin and (D) digestion with trypsin. Samples A – C were obtained from the mycelium by extraction with water, methanol and acetonitrile, sample D was prepared according to the proteomics workflow.

Table 2: Sequence coverages in wildtype and knockout samples of the precursor peptide 120,311 for different proteases and corresponding -log(p)-values

Protease	-10logP	Coverage (%) wildtype	Coverage (%) knockout
None	271.31	91	0
Pepsin	246.65	67	0
Chymotrypsin	195.10	63	0
Trypsin	48.49	13	0

3.2.3 Pathway enrichment analysis

Samples obtained by extraction with water, methanol and acetonitrile from the mycelium of wildtype and knockout strains were also subjected to label free quantification. Principal component analysis is displayed in Figure 23 and shows the separation of clusters of wildtype and knockout samples which indicates differences in protein expression between both strains.

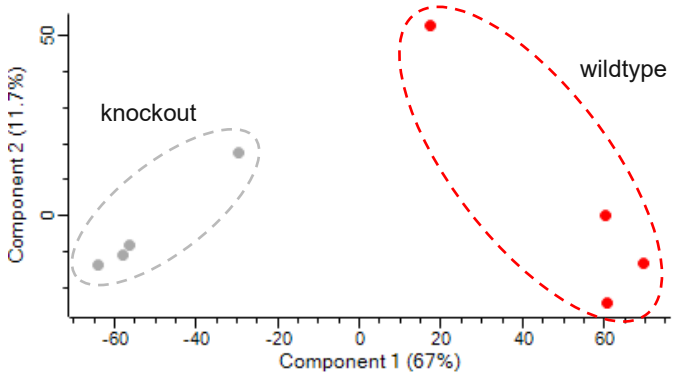

Figure 23: Principal component analysis of 120,311 knockout and wildtype samples.

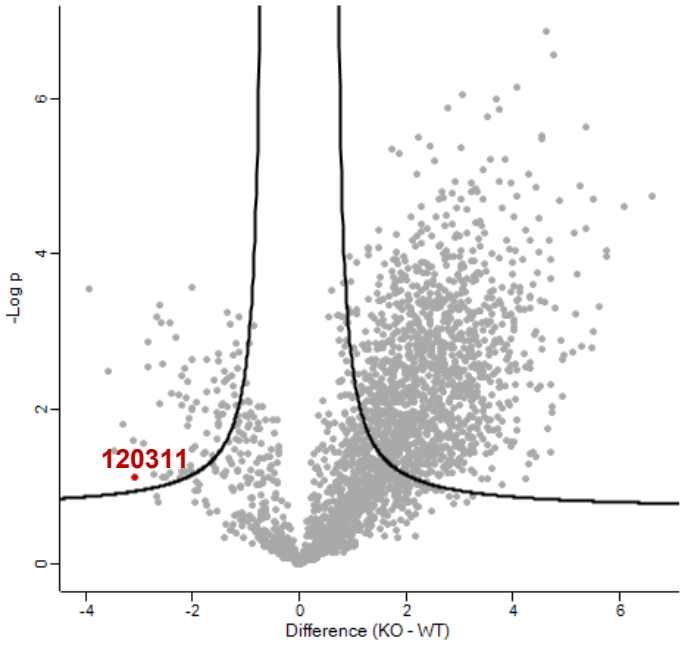

Figure 24: Volcano plot of LFQ data obtained from polar extraction of mycelial samples; FDR 1 %, S0 0.5. The potential RiPP precursor peptide 120,311 is colored in red.

As can be seen in the volcano plot displayed in Figure 24, more proteins could be identified in 120,311 knockout samples compared to the wildtype strain, although identical amounts of mycelium of each sample were used. Subsequent protein estimation revealed that protein concentration in knockout samples was roughly 1.5-times as high as in wildtype samples. After normalization of protein concentration and re-measurement of samples, the volcano plot shown in Figure 25 was obtained.

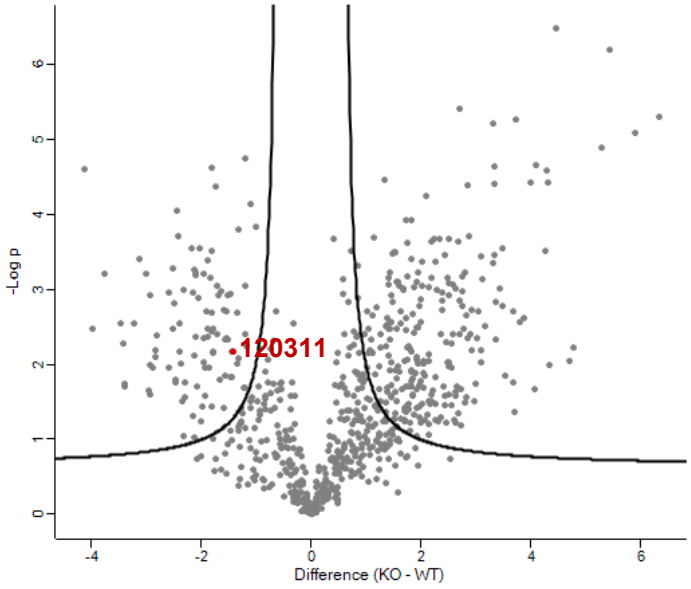

Figure 25: Volcano plot of LFQ data obtained from polar extraction of mycelial samples after normalization of protein concentration; FDR 1 %, S0 0.5. The potential RiPP precursor peptide 120,311 is colored in red.

In total, 244 of all identified proteins were significantly higher expressed in knockout samples, as displayed on the right side of the volcano plot, whereas 103 proteins are more expressed in the wildtype, as shown on the left side. Significantly upregulated proteins were then subjected to STRING analysis.

Figure 26 displays the interaction of proteins which are upregulated in 120,311 knockout strains obtained from polar extraction of mycelial samples. Four prominent clusters were detected. Proteins from cluster 1 are involved in the regulation of the actin cytoskeleton and include several actin-binding proteins such as coronin and profilin as well as components of the Arp2/3 complex which is implicated in actin polymerization and in the formation of branched actin networks.

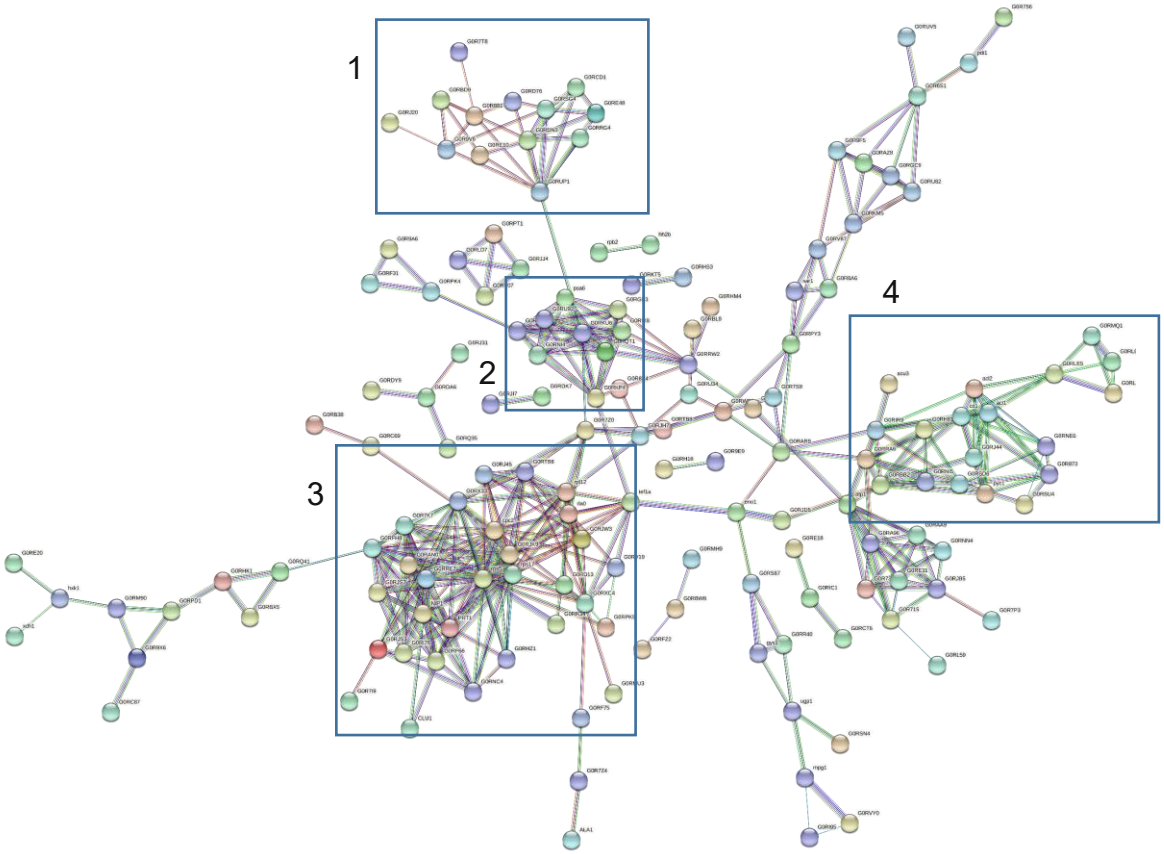

Figure 26: Interaction network of proteins significantly more expressed in 120,311 knockout samples of *T. reesei* obtained from polar extraction. Disconnected nodes are omitted for clarity, interaction score was set to highest confidence (0.9).

Proteins from cluster 2 are mainly subunits of proteasomes such as 26S proteasome regulatory subunit RPN2 which is involved in the ATP-dependent degradation of ubiquitinated proteins. In addition, this cluster contains proteins from the AAA ATPase family which play an important role in many different processes such as protein degradation or protein refolding [94].

Moreover, increased expression of proteins involved in translation could be detected in the knockout strain. These proteins are summarized in cluster 3 and are mainly translation initiation factors but also ribosomal proteins such as the 60S acidic ribosomal protein P0 and proteins which belong to the families uL11, uS7 or eS25.

Many enzymes involved in the TCA cycle could also be identified to be higher expressed in 120,311 knockout samples (cluster 4) such as malate dehydrogenase, pyruvate carboxylase, citrate synthase or aconitate hydratase.

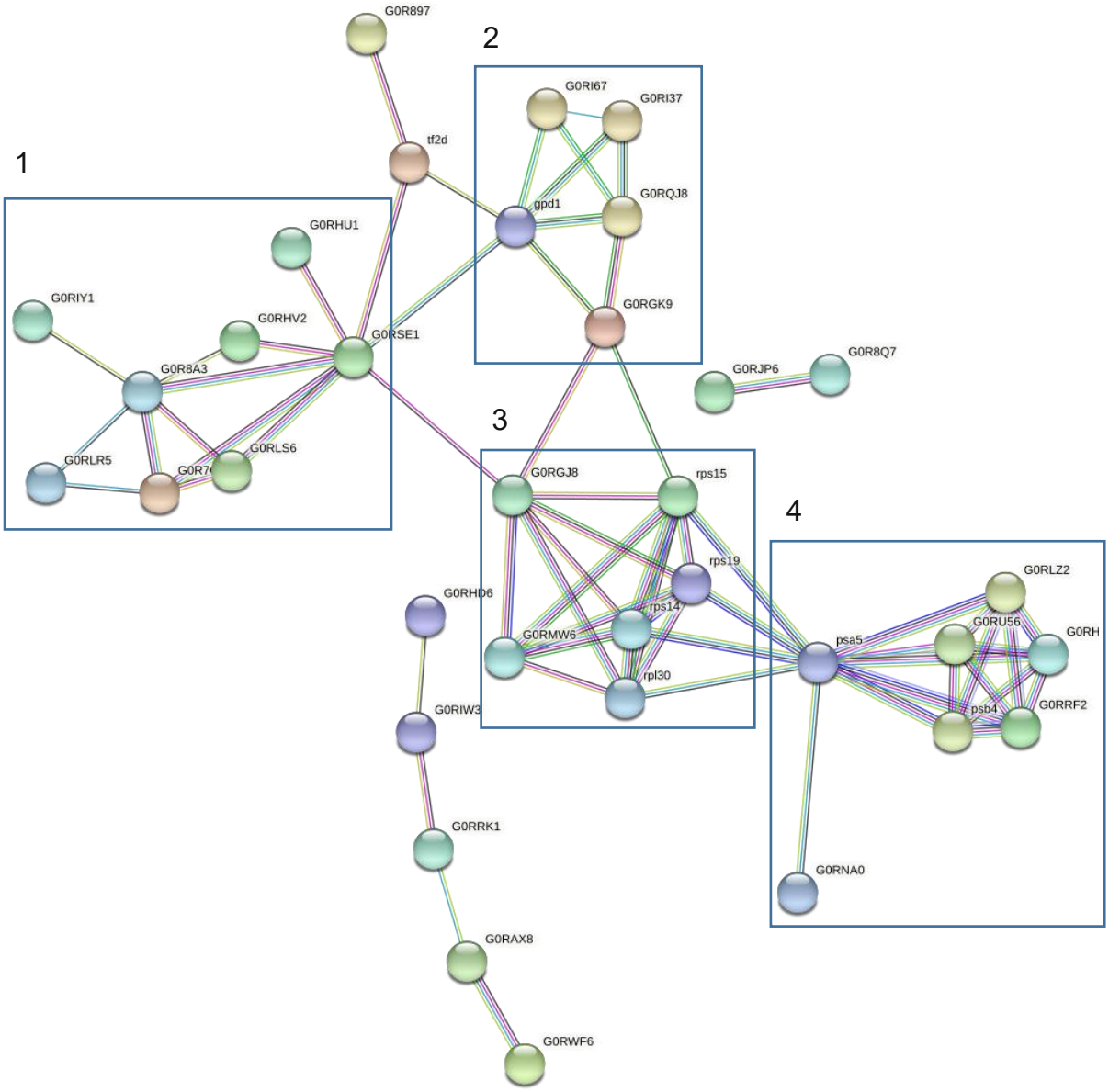

Figure 27: Interaction network of proteins higher expressed in the wildtype samples obtained by polar extraction. Disconnected nodes are hidden for clarity, interaction score was set to high confidence (0.7).

Proteins higher expressed in the wildtype are involved in the electron transport chain. These proteins include subunits of cytochrome c oxidase and electron carrier proteins and are shown in cluster 1 in Figure 27.

Moreover, several enzymes involved in glycolysis such as glyceraldehyde-3-phosphate dehydrogenase or phosphoglycerate mutase were among the proteins more highly expressed in the wildtype (cluster 2).

Additionally, four ribosomal proteins were identified to be higher expressed in the wildtype such as the 40s ribosomal protein S15 or the ribosomal protein L30 (cluster 3).

Another group of detected proteins are alpha and beta subunits of the proteasome (cluster 4).

Mycelial samples were also subjected to a typical proteomics sample preparation method as described in 2.3.2.2. The volcano plot depicted in Figure 28 shows that 16 proteins are significantly higher expressed in the wildtype whereas 18 proteins are more expressed in the knockout. Among these proteins, the precursor peptide 120,311 could not be identified in contrast to samples obtained by polar extraction as described in the section above.

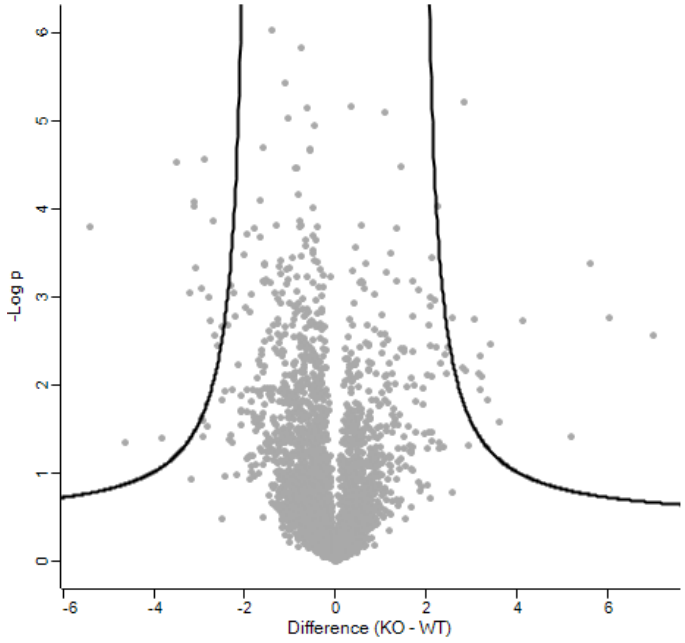

Figure 28: Volcano plot of LFQ data obtained from proteomics sample preparation of mycelial samples; FDR 1 %, S0 0.5.

Due to the low number of significantly upregulated proteins STRING analysis was not possible. Proteins which are significantly differently expressed in wildtype and knockout samples are summarized in Table 3 and Table 4.

Table 3: Significantly higher expressed proteins in *T. reesei* wildtype samples

Difference	Protein IDs	Protein name
-5.40	A0A024SKM2	OMP decarboxylase
-4.62	A0A024RZ08	PAN2-PAN3 deadenylation complex catalytic subunit PAN2
-3.82	A0A024SKK2	Uncharacterized protein
-3.49	A0A024RYT2	Alcohol dehydrogenase class-3
-3.22	A0A024S1D9	Aa_trans domain-containing protein
-3.12	A0A024SE46	FMN-linked oxidoreductase
-3.10	A0A024SHA7	Uncharacterized protein
-3.08	A0A024S2S3	Cytochrome P450 52A12
-2.94	A0A024SM95	General substrate transporter
-2.89	A0A024S054	NADH:flavin oxidoreductase/NADH oxidase
-2.80	A0A024RZU8	Uncharacterized protein
-2.75	A0A024SNL4	NAD(P)-binding protein
-2.71	A0A024RZI6	DSBA oxidoreductase
-2.66	A0A024RZ87	ATP-binding cassette transporter ABC1
-2.59	A0A024S0A2	Uncharacterized protein
-2.49	A0A024S4V2	Beta-galactosidase

Table 4: Significantly higher expressed proteins in *T. reesei* 120,311 knockout samples

Difference	Protein IDs	Protein name
2.26	A0A024S039	Nudix hydrolase domain-containing protein
2.58	A0A024S1N2	Heme peroxidase
2.79	A0A024SHL9	CVNH domain-containing protein
2.83	A0A024RYB5	Tripeptidyl-peptidase 1
2.87	A0A024RXK9	Uncharacterized protein
3.07	A0A024RYY0	EthD domain-containing protein
3.16	A0A024S5L4	p-loop containing nucleoside triphosphate hydrolase protein
3.20	A0A024SCE6	MARVEL domain-containing protein
3.20	A0A024SM41	AAA domain-containing protein
3.23	A0A024RYX7	MFS domain-containing protein
3.36	A0A024SDG2	Cytochrome P450
3.41	A0A024SI67	Uncharacterized protein
3.61	A0A024SEU8	Uncharacterized protein
4.13	A0A024SNG6	MFS general substrate transporter
5.19	A0A024SKK5	Uncharacterized protein
5.62	A0A024RXP3	Zn(2)-C6 fungal-type domain-containing protein
6.02	A0A024S064	Uncharacterized protein
7.00	A0A024S3F7	Putative agmatine deiminase

3.2.4 Analysis of RiPPs in *Aspergillus flavus*

One potential way for the analysis of natural products is to extract metabolites from the mycelium with polar solvents followed by chromatographic separation and mass spectrometric analysis. Moreover, the growth medium can be analyzed to determine whether specific natural products are secreted into the surrounding medium. To show that this strategy is also suitable for the analysis of unknown RiPP products, this sample preparation method was applied to the fungus *Aspergillus flavus*.

A. flavus is known to produce a variety of secondary metabolites including aflatoxins, cyclopiazonic acid and aflatrem. Recently, it has been shown to also produce ustiloxin B which is a toxic cyclic peptide containing the tetrapeptide Tyr-Ala-Ile-Gly. This tetrapeptide is cyclized at the side chains of Tyr and Ile and carries a methyl group, a hydroxyl group and norvaline, a non-protein-coding amino acid, as modifications on the tyrosine (Figure 29 A) [95]. Another interesting compound which has been discovered recently is asperipin-2a, a bicyclic peptide derived from the sequence FYYTGY shown in Figure 29 [96, 97].

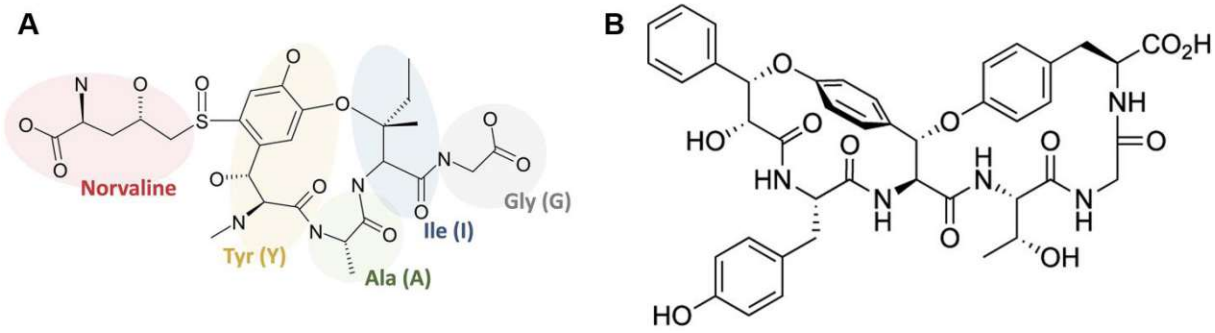

Figure 29: (A) Structure of ustiloxin B [95] and (B) asperipin-2a [97].

To analyze the expression of ustiloxin B and asperipin-2a in *A. flavus*, growth media and mycelia obtained from ustiloxin B overexpression strains and ustiloxin B and asperipin-2a knockout strains were used. Each strain was either cultivated in pure Mandels-Andreotti medium (MAM) or in MAM supplemented with amino acid mix. Table 5 shows an overview of the samples obtained from the three *A. flavus* strains. For each condition and strain, three biological replicates were used.

Table 5: *A. flavus* samples used for the analysis of ustiloxin B and asperipin-2a. Samples were prepared from the growth medium as well as the mycelium.

Sample name	Sample type
Δ041400 -AA	Asperipin-2a knockout, cultivated in MAM
Δ041400 +AA	Asperipin-2a knockout, cultivated in MAM supplemented with amino acid mix
ustR^{OE} -AA	Ustiloxin B overexpression strain, cultivated in MAM
ustR^{OE} +AA	Ustiloxin B overexpression strain, cultivated in MAM supplemented with amino acid mix
Δust -AA	Ustiloxin B knockout, cultivated in MAM
Δust +AA	Ustiloxin B knockout, cultivated in MAM supplemented with amino acid mix

For sample preparation and LC-(IMS)-MS/MS analysis, the same methods as for samples obtained from *T. reesei* were employed to investigate if these methods are in theory applicable for the discovery and analysis of RiPPs. One method was based on extracting (secondary) metabolites from the medium and mycelium with a mixture of water, acetonitrile and methanol, whereas for the other approach, proteins and small peptide natural products were extracted from the mycelium or growth medium using Tris-HCl buffer, acetone precipitated and finally analyzed by LC-(IMS)-MS/MS.

For chromatographic separation, reversed-phase and HILIC columns were employed to investigate the retention behavior of mature RiPPs on different column materials. Mass spectrometry analysis was performed in positive and negative ionization mode. Moreover, the effect of ion mobility on the analysis was determined.

Reversed-phase and HILIC chromatograms obtained from mycelial samples of ustiloxin B overexpression strains prepared by the extraction method using polar solvents (chapter 2.3.2.1) are displayed in Figure 30. The sum formula of ustiloxin B ($C_{26}H_{39}N_5O_{12}S$) [98] was used to create extracted ion chromatograms (EIC) to identify the peak corresponding to ustiloxin B. For reversed-phase separation, ustiloxin B elutes at a retention time of 2.0 min. When separation is carried out on a HILIC column, elution happens at 11.2 min.

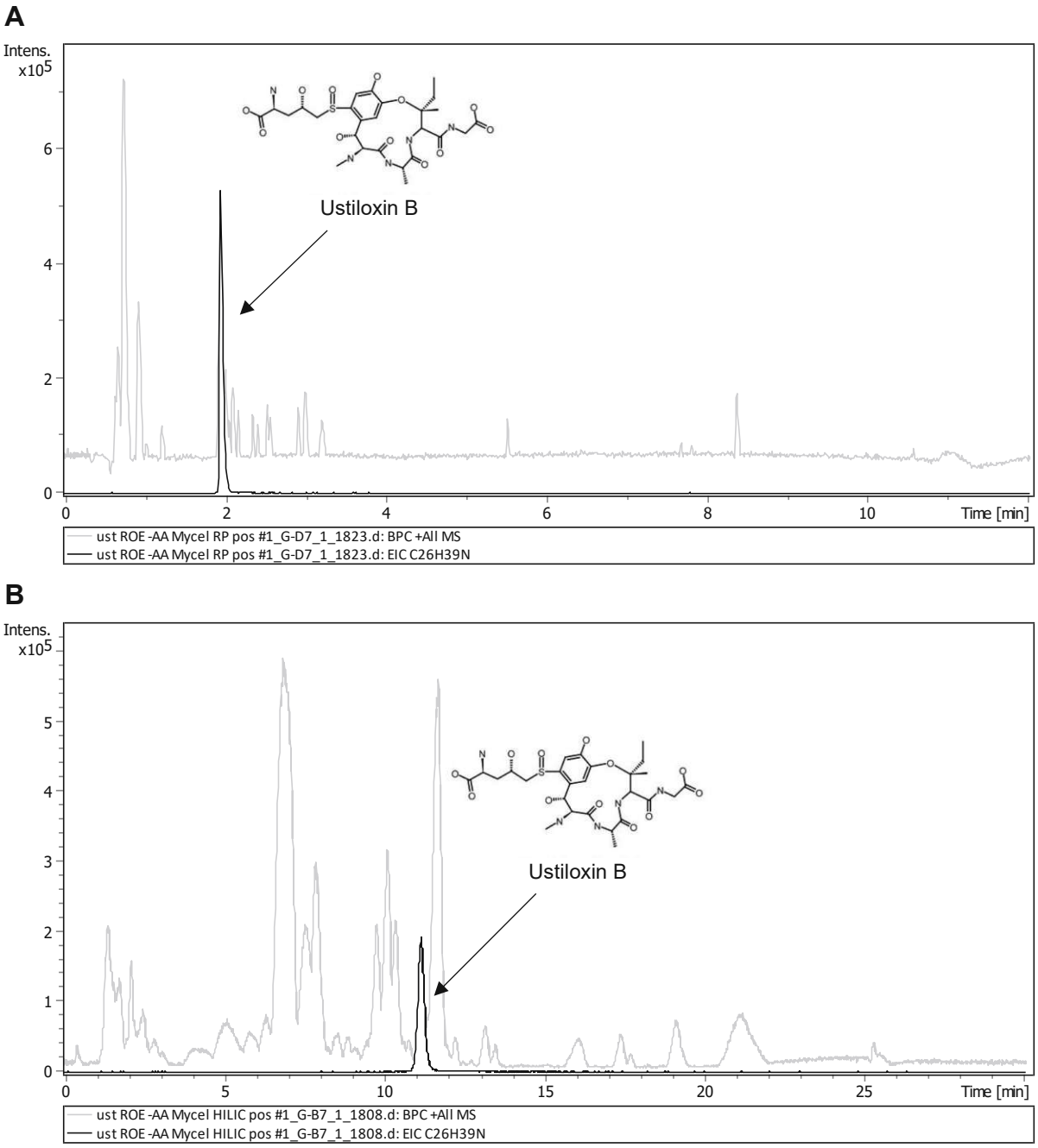

Figure 30: Base peak chromatogram obtained from the mycelium of *A. flavus* ustiloxin B overexpressing strain (grey) and extracted ion chromatogram of m/z 646.24 ± 0.01 $[M+H]^+$ corresponding to ustiloxin B (black). Sample preparation was performed according to chapter 2.3.2.1. Chromatographic separation was carried out on a reversed-phase column (A) and on a HILIC column (B). Mass spectrometric detection was performed in positive ionization mode.

The extracted ion chromatogram obtained from the growth medium of ustiloxin B overexpressing strains is illustrated in Figure 31 together with the corresponding EIC obtained from the mycelial sample. This shows that ustiloxin B is secreted from the fungus into the surrounding medium.

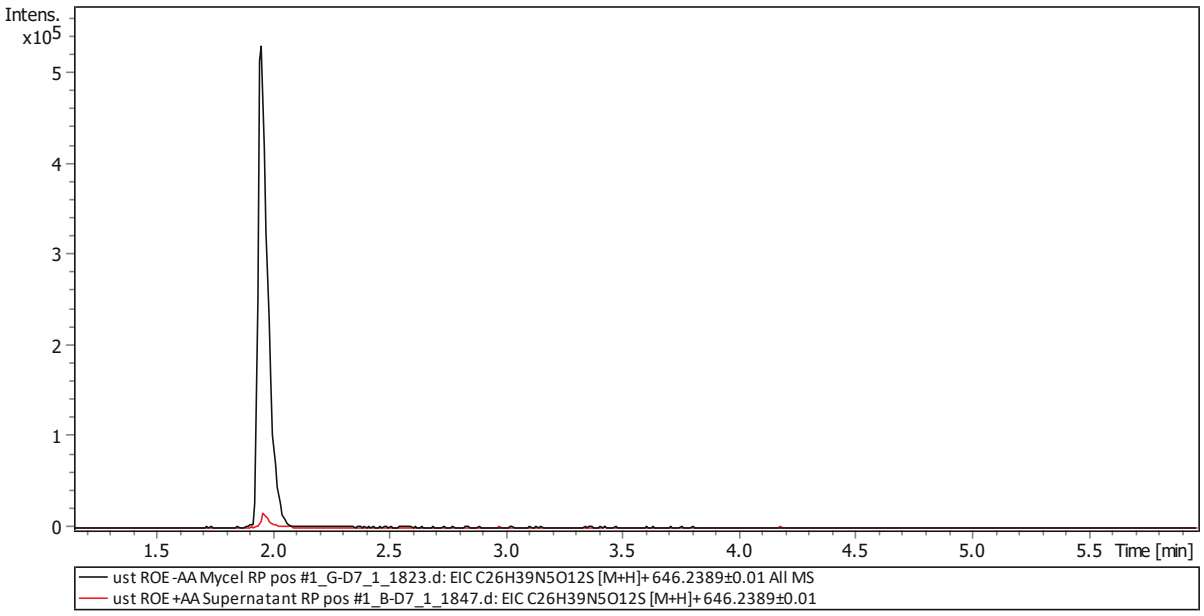

Figure 31: EICs of m/z 646.24 ± 0.01 $[M+H]^+$ obtained from mycelial sample (black) and growth medium (red) of ustiloxin B overexpressing strains for reversed-phase separation and mass spectrometric detection in positive ionization mode

While ustiloxin B could be successfully detected in positive ion mode, asperipin-2a could not be detected. Therefore, samples were also analyzed in negative ionization mode.

Figure 32 displays the base peak chromatogram obtained from the ustiloxin B overexpressing strain in negative ionization mode as well as the extracted ion chromatograms of ustiloxin B and asperipin-2a. When ionization is carried out in negative mode, asperipin-2a can be detected at retention times of around 6.1 min for reversed-phase separation and 2.8 min in case of separation on a HILIC column, albeit with a smaller peak area compared to ustiloxin B. For a successful identification of an unknown compound, it is therefore crucial to use overexpressing strains for analysis as differences in chromatograms between strains can then easily be detected.

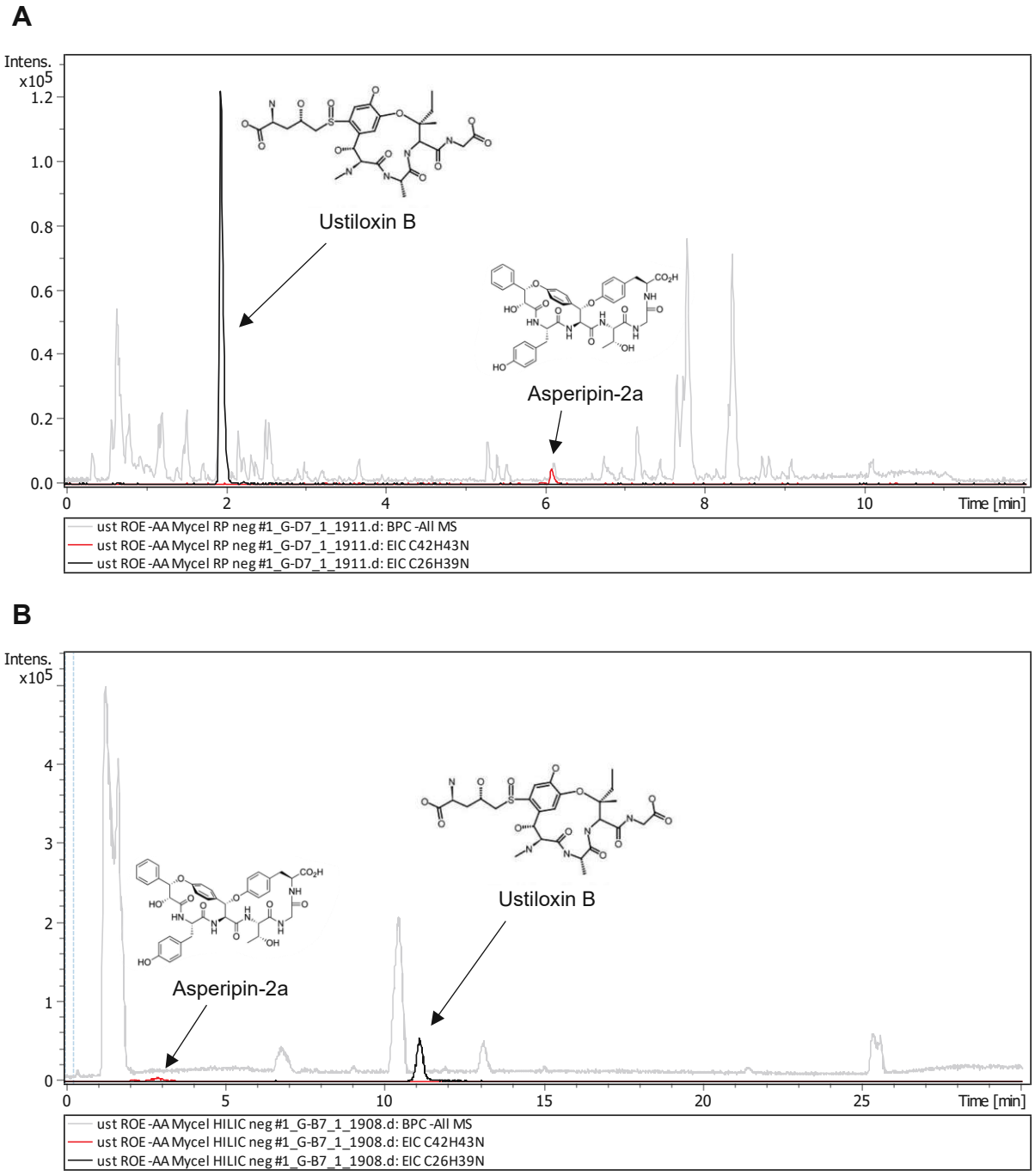

Figure 32: Base peak chromatogram obtained from the mycelium of *A. flavus* ustiloxin B overexpressing strain (grey), extracted ion chromatogram of m/z 644.21 ± 0.02 $[M-H]^-$ corresponding to ustiloxin B (black) and extracted ion chromatogram of m/z 808.28 ± 0.02 $[M-H]^-$ corresponding to asperipin-2a (red). Sample preparation was performed according to chapter 2.3.2.1. Chromatographic separation was carried out on a reversed-phase column (A) and on a HILIC column (B). Mass spectrometric detection was performed in negative ionization mode.

Moreover, the effect of employing ion mobility for analysis was investigated. Figure 33 and Figure 34 show the chromatograms obtained in positive and negative ionization mode, respectively, when using PASEF as MS scan mode.

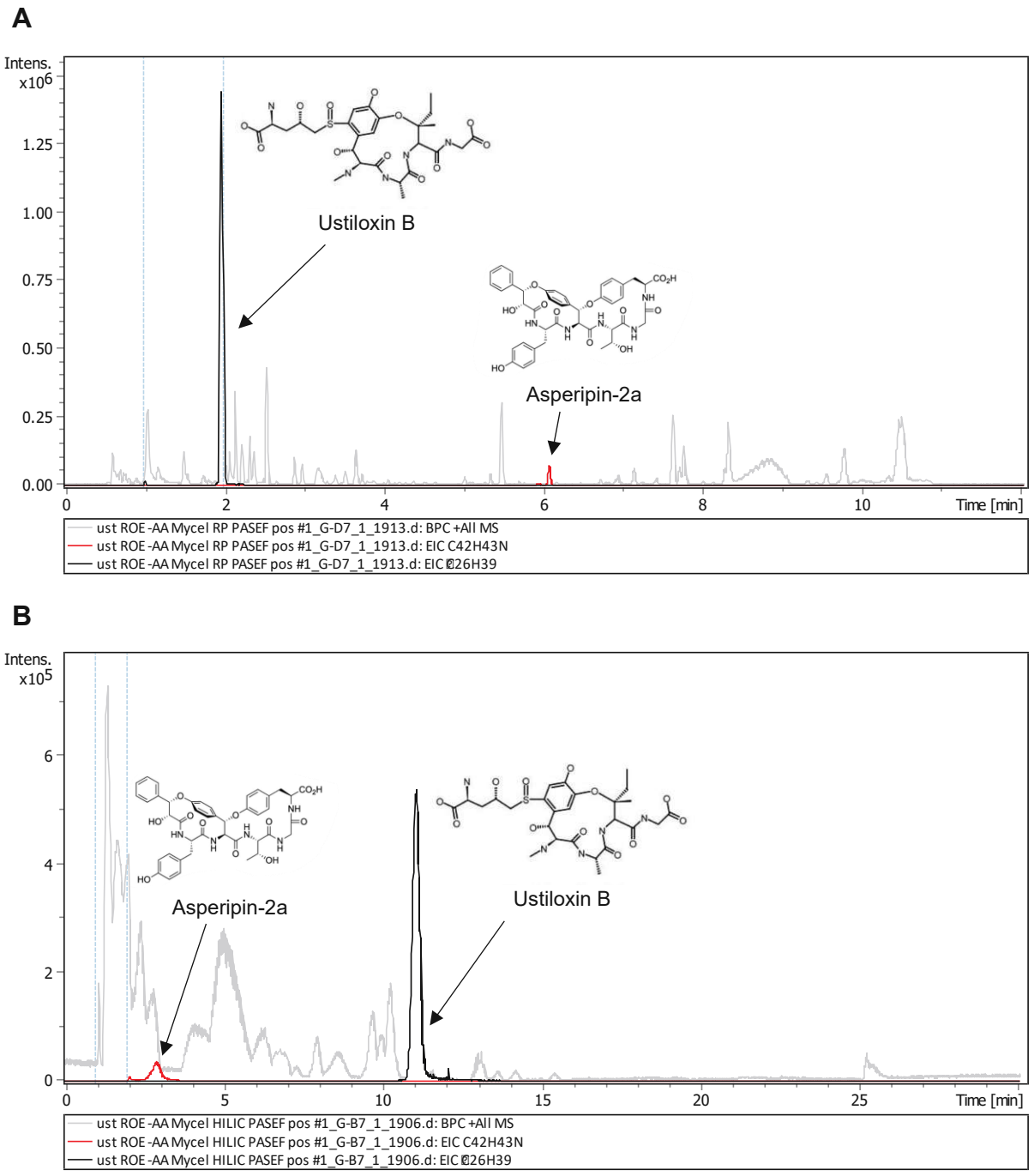

Figure 33: Base peak chromatogram obtained from the mycelium of *A. flavus* ustiloxin B overexpressing strain (grey), extracted ion chromatogram of m/z 646.24 ± 0.02 $[M+H]^+$ corresponding to ustiloxin B (black) and extracted ion chromatogram of m/z 810.28 ± 0.02 $[M+H]^+$ corresponding to asperipin-2a (red). Sample preparation was performed according to chapter 2.3.2.1. Chromatographic separation was carried out on a reversed-phase column (A) and on a HILIC column (B). Mass spectrometric detection was performed in positive ionization mode using PASEF.

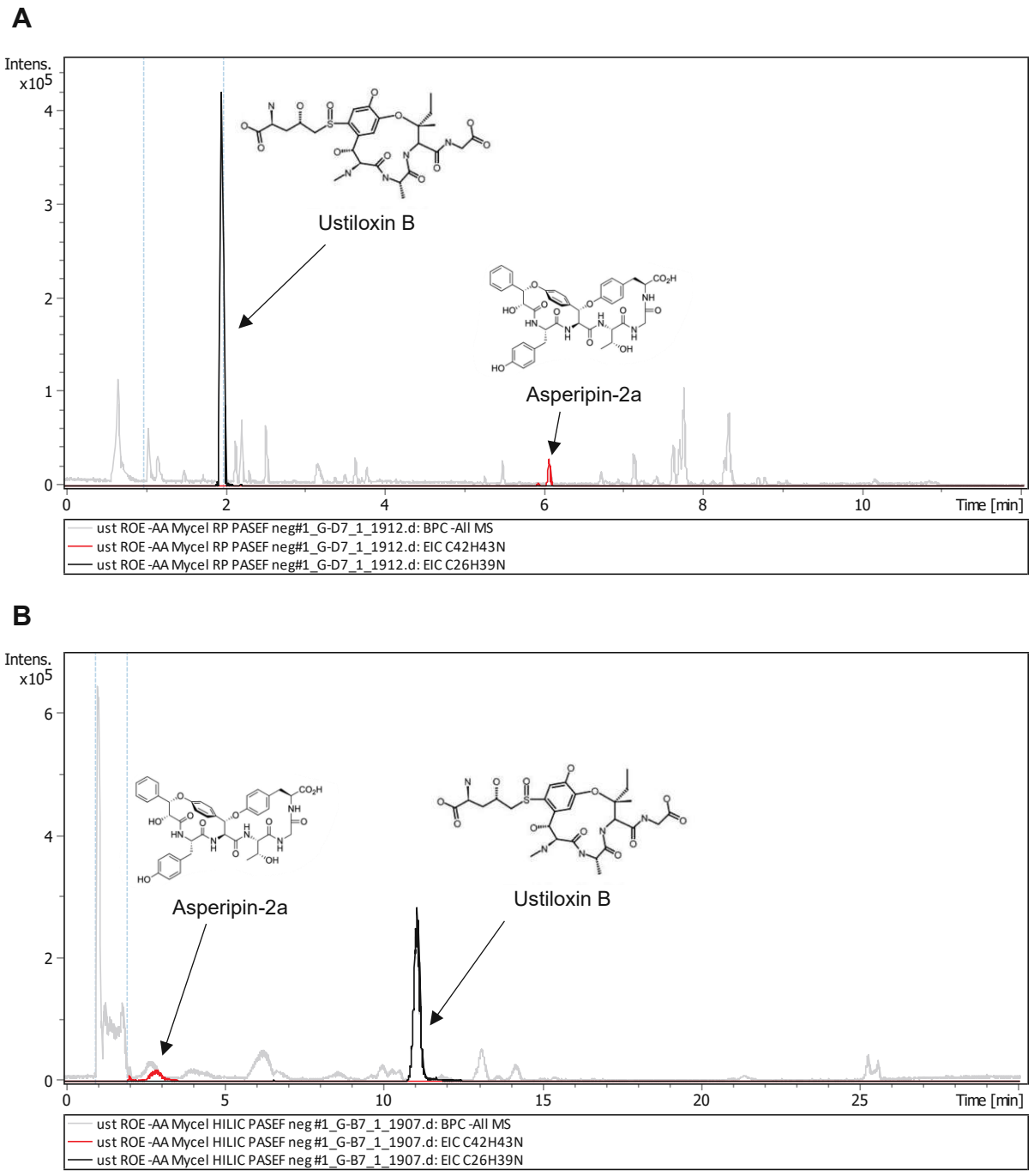

Figure 34: Base peak chromatogram obtained from the mycelium of *A. flavus* ustiloxin B overexpressing strain (grey), extracted ion chromatogram of m/z 644.21 ± 0.02 $[M-H]^-$ corresponding to ustiloxin B (black) and extracted ion chromatogram of m/z 808.28 ± 0.02 $[M-H]^-$ corresponding to asperipin-2a (red). Sample preparation was performed according to chapter 2.3.2.1. Chromatographic separation was carried out on a reversed-phase column (A) and on a HILIC column (B). Mass spectrometric detection was performed in negative ionization mode using PASEF.

Analysis in PASEF mode has shown that detection of both substances is possible with increased sensitivity not only in negative ionization mode but also when positive ionization mode is employed, in contrast to analysis without ion mobility where asperipin-2a could not be detected. Figure 35 shows the increase in sensitivity by more than the 2.5-fold.

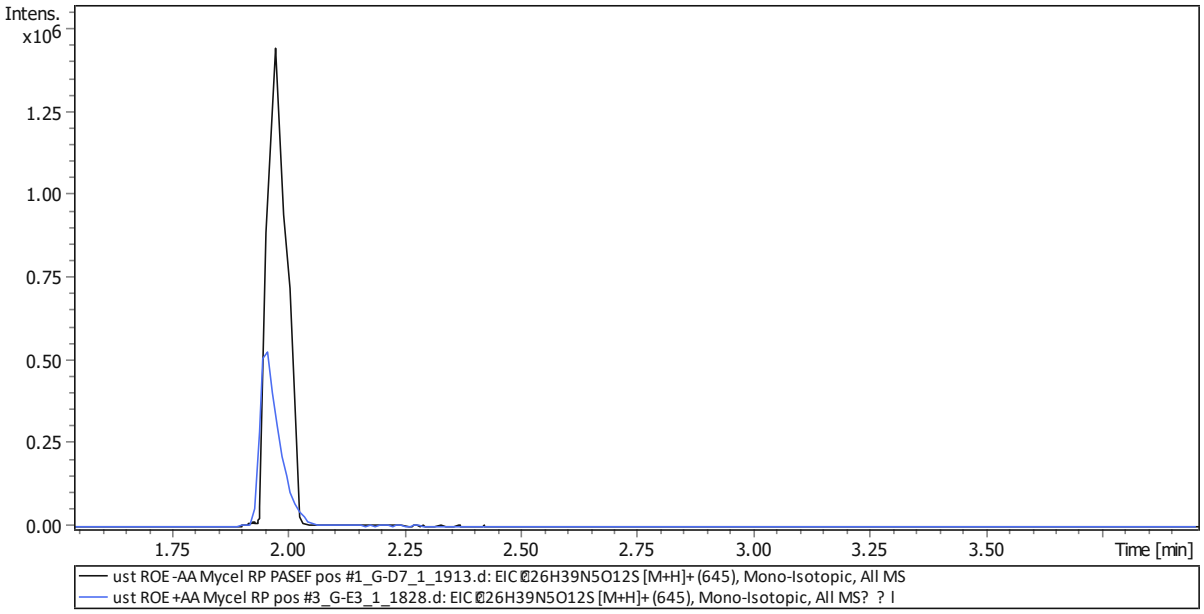

Figure 35: Extracted ion chromatograms of m/z 646.24 ± 0.02 $[M+H]^+$ corresponding to ustiloxin B in PASEF mode (black) and without ion mobility (blue).

Fragmentation by collision-induced dissociation is vital for structural elucidation of an unknown compound. Fragmentation of ustiloxin B was thus analyzed using different collision energies between 25 and 60 eV. Fragment spectra at 25.0 eV, 40.0 eV and 60.0 eV are shown in Figure 36. Increasing the collision energy results in an increased number of produced fragments which can be used for structure determination. In case of ustiloxin B, a prominent fragment peak is visible at a mass-to-charge ratio of around 181 which might be the norvaline fragment bound to the sulfoxide group.

These results indicate that a collision energy of 40 eV is optimal for fragmentation of ustiloxin B. If the collision energy is too low, only few fragments are obtained in the MS2 spectrum which is not sufficient for structural elucidation. High collision energies, on the other hand, result in very complex MS2 spectra with many fragments.

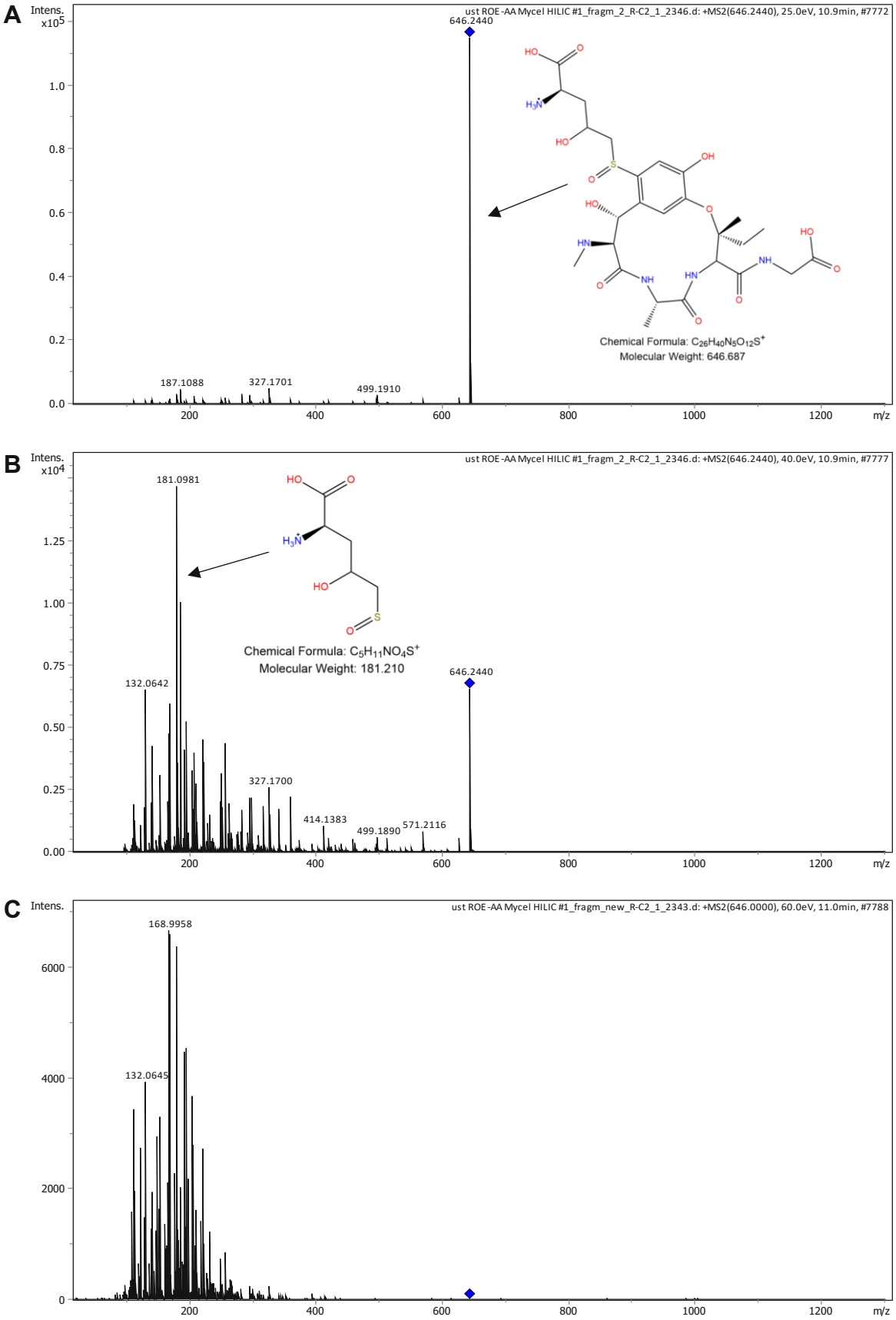

Figure 36: MS2 spectra of ustiloxin B fragments obtained at collision energies of 25.0 eV (A), 40.0 eV (B) and 60.0 eV (C).

Finally, data obtained from *A. flavus* samples was processed using Metaboscape in the same way as described in chapter 2.5.4 to obtain volcano plots showing features with altered abundance.

When comparing ustiloxin B expression in overexpressing and knockout samples, no p-value was calculated for the feature corresponding to ustiloxin B. Therefore, the corresponding feature is not displayed in the volcano plot (Figure 37).

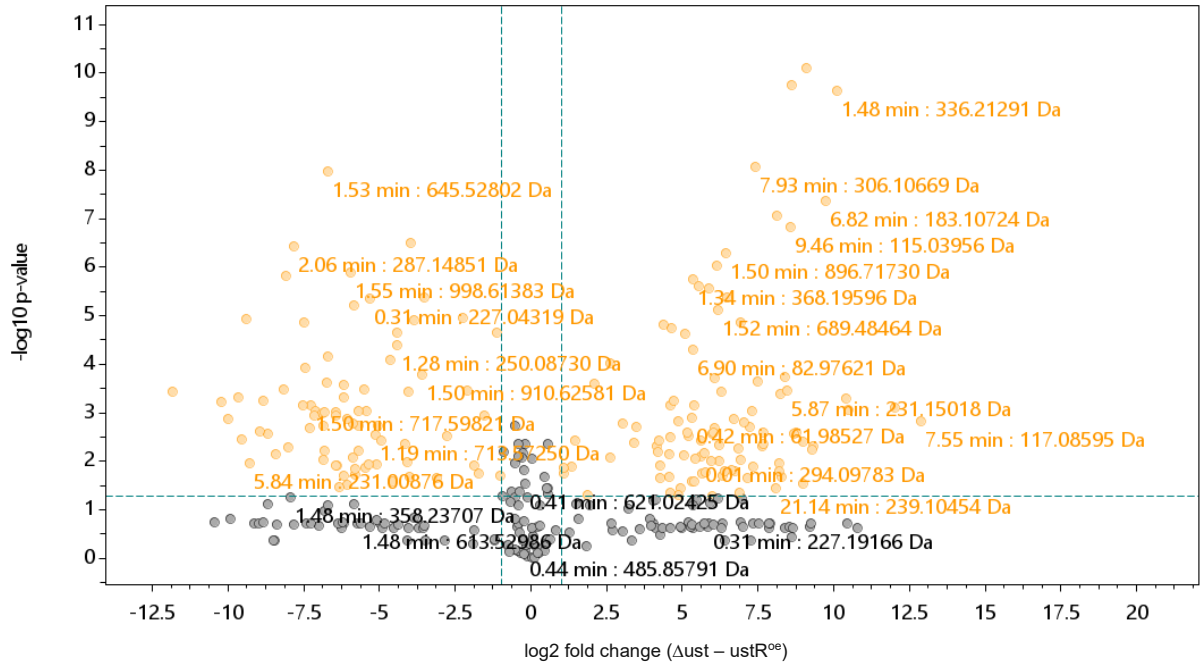

Figure 37: Volcano plot showing features with different abundances in samples obtained from Δust and ustR^{oe} strains which were cultivated in MAM supplemented with amino acid mix. Samples were separated on HILIC columns and mass spectrometric analysis was carried out in positive ionization mode without using ion mobility.

Moreover, samples obtained from ustiloxin B overexpressing strains were compared to Δ041400 strains where ustiloxin B is expressed but to a much smaller extent. The volcano plot obtained by this comparison is displayed in Figure 38 and clearly shows overexpression of ustiloxin B. Interestingly, two features were extracted from the raw data with a mass-to-charge ratio of 646.3 which both elute at 11.18 min and therefore correspond to ustiloxin B.

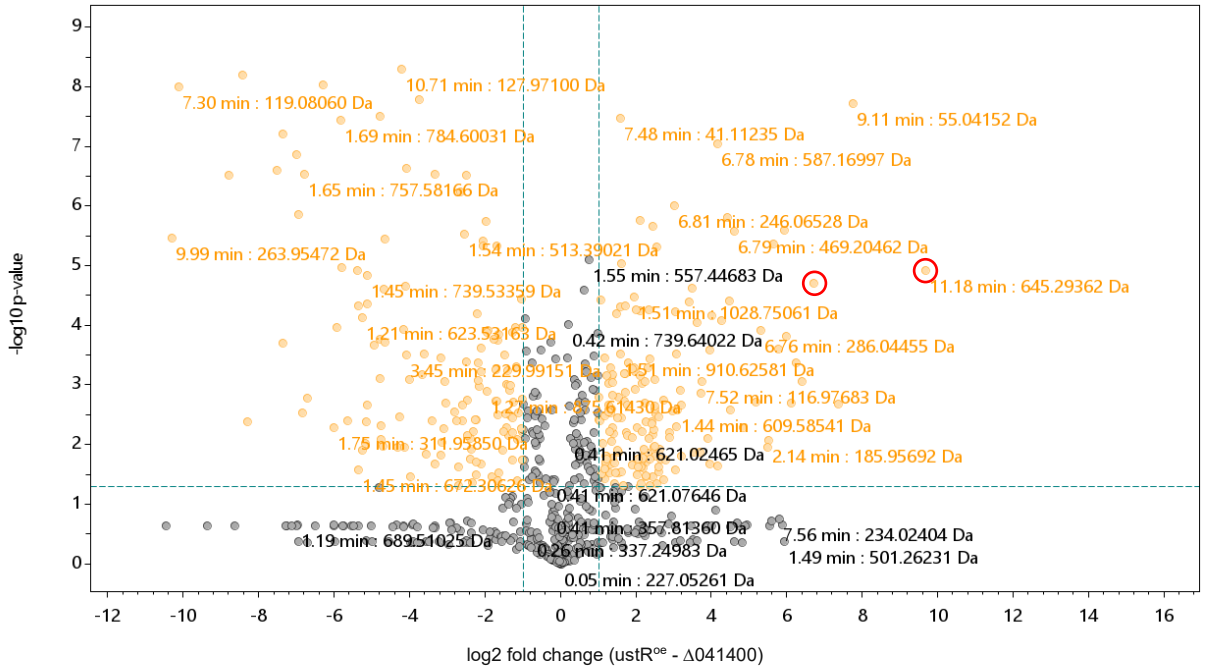

Figure 38: Volcano plot showing features with different abundances in samples obtained from Δ041400 and ustR^{oe} strains which were cultivated in MAM supplemented with amino acid mix. Samples were separated on HILIC columns and mass spectrometric analysis was carried out in positive ionization mode without using ion mobility. Features marked in red correspond to ustiloxin B.

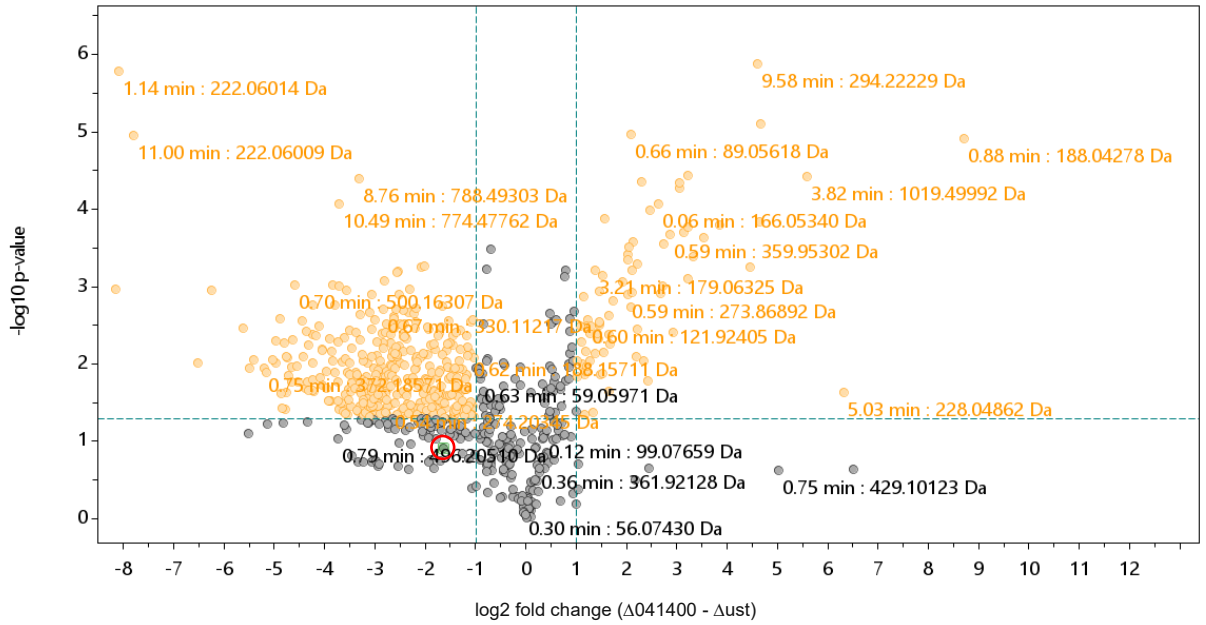

Figure 39: Volcano plot showing features with different abundances in samples obtained from Δ041400 and Δust strains which were cultivated in MAM supplemented with amino acid mix. Samples were separated on RP columns and mass spectrometric analysis was carried out in positive ionization mode without using ion mobility. The feature marked in red corresponds to asperipin-2a.

A comparison of samples obtained from asperipin-2a knockout and ustiloxin B knockout strains was also carried out to investigate asperipin-2a expression (Figure 39). Significantly higher expression of asperipin-2a in ustiloxin B knockout samples could not be detected, which is surprising as asperipin-2a should not be expressed in Δ041400 samples at all. The reason for this lies in the presence of an asperipin-2a peak eluting at 6.12 min in one of the Δ041400 samples. This is most probably due to carryover during LC-MS/MS analysis or contamination during sample preparation.

4 Summary and Conclusion

In the first part of this thesis, an already established protocol for in-solution digestion was compared to the SP3 approach as well as a commercially available kit for sample preparation from PreOmics. It was demonstrated that the in-solution method yields a significantly higher number of protein identifications compared to the other two approaches. As the SP3 approach stands out due to its simplicity and can be performed with low sample quantities, several clean-up methods were carried out in order to optimize the SP3 method.

A significantly higher number of protein identifications was observed when an additional desalting step on StageTips after digestion of proteins was included in the SP3 protocol. In a next step, this modified SP3 approach was applied to spheroids obtained from 3D cell culture together with the in-solution method. Comparable number of identified proteins were observed for both methods. However, in the case of one spheroid used, the conventional SP3 method was additionally performed and yielded significantly higher numbers of identified proteins compared to the optimized SP3 approach. This led to the assumption that the desalting step is not necessary for low protein inputs as the overall salt concentration in the sample is low. An additional desalting step would therefore one lead to higher peptide loss.

In order to test this hypothesis, the in-solution method, the conventional as well as the modified SP3 approach were performed on samples with different protein inputs ranging from 10 µg to 50 ng of protein. A general decrease in protein identifications was observed for decreasing protein amounts due to lower column loads. While the efficiency was comparable for all three methods for protein inputs of 10 µg, 5 µg, 1 µg and 500 ng, a significantly higher number of protein identifications was observed for a protein input of 250 ng. This confirms the assumption that the desalting step included in the SP3 method is dispensable for low protein inputs. For lower protein amounts of 125 ng and 50 ng, however, no significant difference in efficiencies was observed between the three methods.

Moreover, at a later point in time, supernatants obtained from washing SP3 beads with acetonitrile and isopropanol, respectively, were analyzed by nanoLC-IMS-MS/MS analyzed. Results showed that high numbers of proteins were identified in the

supernatants suggesting that peptides are eluted into the organic phase during the washing steps while salts and other contaminants do not dissolve in the organic solvent. When the desalting step was performed with isopropanol, more protein identifications could be observed in the supernatant than with the optimized SP3 approach. This method shows comparable efficiency to the optimized SP3 approach and therefore requires a more profound investigation in the future.

The second part of this thesis focuses on the identification of RiPPs the fungus *T. reesei*. Initial experiments on supernatants obtained from *T. reesei* culture and the corresponding mycelial samples were performed using different methods for sample preparation as well as different LC-MS/MS methods to obtain general information on metabolic profiles. Due to the vast number of features detected in the samples, it was not possible to retrieve any information on RiPP extraction from this data. Therefore, *A. flavus* strains which are known to produce the RiPPs asperipin-2a and ustiloxin B were subjected to the same extraction methods and LC-MS/MS methods used for the analysis of *T. reesei* samples in order to determine suitable methods for RiPP extraction and detection. Asperipin-2a and ustiloxin B were successfully extracted from *A. flavus* mycelial samples using water, acetonitrile and methanol in a ratio of 1:2:2. While asperipin-2a showed better retention behavior on the reversed-phase column and was only detected in negative ionization mode, ustiloxin B is better retained on a HILIC column and can be detected in both positive and negative ionization modes. The chromatogram obtained from the extract of ustiloxin B overexpressing strains showed a prominent peak whereas asperipin-2a only led to a low signal. This clearly shows that for successful identification of unknown RiPPs overexpressing strains have to be used for analysis.

Furthermore, RiPP production in *T. reesei* was analyzed on the peptide level. Preliminary experiments have demonstrated the transcription of the RiPP precursors 106,662 and 120,311 under various conditions. In order to show that mRNAs detected in the transcript analysis, several extraction, LC-MS/MS and data processing methods were tested. Peptide extraction with water, methanol and acetonitrile (1:2:2) followed by nanoLC-IMS-MS/MS analysis and database searches have proved to be suitable for the identification of RiPP precursors. Precursor 120,311 could be successfully detected with this approach.

The methods developed in the course of this thesis form an important basis for the future identification and analysis of RiPPs in *T. reesei*. It was demonstrated that overexpressing strains are necessary for RiPP detection. The development of these strains is thus crucial for further RiPP investigation in *T. reesei.* To facilitate identification of mature RiPPs, spectral network approaches can be employed where RiPPs are classified based on similar fragmentation patterns [99].

5 References

[1] K. K. Hixson, D. Lopez-Ferrer, E. W. Robinson, L. Paša-Tolić, in *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)* (Eds.: J. C. Lindon, et al.), Academic Press, Oxford, **2017**, pp. 766-773.

[2] R. M. Tubaon, P. R. Haddad, J. P. Quirino, Sample Clean-up Strategies for ESI Mass Spectrometry Applications in Bottom-up Proteomics: Trends from 2012 to 2016, **2017**, *17*, 1700011.

[3] S. Andrecht, J. v. Hagen, in *Proteomics Sample Preparation* (Ed.: J. v. Hagen), Wiley-VCH, **2008**, pp. 5-9.

[4] P. Feist, A. B. Hummon, Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples, **2015**, *16*, 3537-3563.

[5] W. Weiss, A. Görg, in *Proteomics Sample Preparation* (Ed.: J. v. Hagen), Wiley-VCH, **2008**, p. 136.

[6] C. S. Hughes, et al., Single-pot, solid-phase-enhanced sample preparation for proteomics experiments, *Nature Protocols* **2019**, *14*, 68-85.

[7] L. Switzar, M. Giera, W. M. A. Niessen, Protein Digestion: An Overview of the Available Techniques and Recent Developments, *Journal of Proteome Research* **2013**, *12*, 1067-1077.

[8] H. Wang, et al., Development and Evaluation of a Micro- and Nanoscale Proteomic Sample Preparation Method, *Journal of Proteome Research* **2005**, *4*, 2397-2403.

[9] J. Rappsilber, Y. Ishihama, M. Mann, Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics, *Analytical Chemistry* **2003**, *75*, 663-670.

[10] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, *Nature Protocols* **2007**, *2*, 1896-1906.

[11] I. R. León, V. Schwämmle, O. N. Jensen, R. R. Sprenger, Quantitative Assessment of In-solution Digestion Efficiency Identifies Optimal Protocols for Unbiased Protein Analysis*, *Molecular & Cellular Proteomics* **2013**, *12*, 2992-3005.

[12] A. Zougman, P. J. Selby, R. E. Banks, Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis, *Proteomics* **2014**, *14*, 1006-1000.

[13] W. Chen, et al., Simple and Integrated Spintip-Based Technology Applied for Deep Proteome Profiling, *Analytical Chemistry* **2016**, *88*, 4864-4871.

[14] L. L. Manza, et al., Sample preparation and digestion for proteomic analyses using spin filters, *Proteomics* **2005**, *5*, 1742-1745.

[15] C. S. Hughes, et al., Ultrasensitive proteome analysis using paramagnetic bead technology, *Molecular Systems Biology* **2014**, *10*, 757.

[16] T. C. Walther, M. Mann, Mass spectrometry–based proteomics in cell biology, *Journal of Cell Biology* **2010**, *190*, 491-500.

[17] Y. V. Karpievitch, et al., Liquid chromatography mass spectrometry-based proteomics: Biological and technological aspects, *The Annals of Applied Statistics* **2010**, 1797-1823, 1727.

[18] T. Fröhlich, G. J. Arnold, Proteome research based on modern liquid chromatography – tandem mass spectrometry: separation, identification and quantification, *Journal of Neural Transmission* **2006**, *113*, 973-994.

[19] P. Xu, D. M. Duong, J. Peng, Systematical optimization of reverse-phase chromatography for shotgun proteomics, *Journal of proteome research* **2009**, *8*, 3944-3950.

[20] J. Fenn, et al., Electrospray ionization for mass spectrometry of large biomolecules, *Science* **1989**, *246*, 64-71.

[21] M. Wilm, Principles of Electrospray Ionization, *Molecular & Cellular Proteomics* **2011**, *10*.

[22] C. K. Meng, M. Mann, J. B. Fenn, Of protons or proteins, *Zeitschrift für Physik D Atoms, Molecules and Clusters* **1988**, *10*, 361-368.

[23] J. B. Fenn, et al., Electrospray ionization–principles and practice, *Mass Spectrometry Reviews* **1990**, *9*, 37-70.

[24] S. Banerjee, S. Mazumdar, Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte, *International Journal of Analytical Chemistry* **2012**, *2012*, 282574.

[25] R. A. Zubarev, The challenge of the proteome dynamic range and its implications for in-depth proteomics, *Proteomics* **2013**, *13*, 723-726.

[26] J. P. C. Vissers, Recent developments in microcolumn liquid chromatography, *Journal of Chromatography A* **1999**, *856*, 117-143.

[27] S. R. Wilson, T. Vehus, H. S. Berg, E. Lundanes, Nano-LC in proteomics: recent advances and approaches, *Bioanalysis* **2015**, *7*, 1799-1815.

[28] A. Schmidt, M. Karas, T. Dülcks, Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI?, *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 492-500.

[29] B. Cañas, et al., Mass spectrometry technologies for proteomics, *Briefings in Functional Genomics* **2006**, *4*, 295-320.

[30] R. Aebersold, D. R. Goodlett, Mass Spectrometry in Proteomics, *Chemical Reviews* **2001**, *101*, 269-296.

[31] X. Han, A. Aslanian, J. R. Yates, Mass spectrometry for proteomics, *Current Opinion in Chemical Biology* **2008**, *12*, 483-490.

[32] I. V. Chernushevich, A. V. Loboda, B. A. Thomson, An introduction to quadrupole–time-of-flight mass spectrometry, *Journal of Mass Spectrometry* **2001**, *36*, 849-865.

[33] E. d. Hoffmann, V. Stroobant, in *Mass Spectrometry Principles and Applications*, Wiley, **2007**, pp. 126-143.

[34] S. Lacorte, A. R. Fernandez-Alba, Time of flight mass spectrometry applied to the liquid chromatographic analysis of pesticides in water and food, *Mass Spectrometry Reviews* **2006**, *25*, 866-880.

[35] Y. M. Ibrahim, et al., Development of a new ion mobility time-of-flight mass spectrometer, *International Journal of Mass Spectrometry* **2015**, *377*, 655-662.

[36] M. E. Ridgeway, C. Bleiholder, M. Mann, M. A. Park, Trends in trapped ion mobility – Mass spectrometry instrumentation, *TrAC Trends in Analytical Chemistry* **2019**, *116*, 324-331.

[37] A. B. Kanu, et al., Ion mobility–mass spectrometry, *Journal of Mass Spectrometry* **2008**, *43*, 1-22.

[38] M. E. Ridgeway, et al., Trapped ion mobility spectrometry: A short review, *International Journal of Mass Spectrometry* **2018**, *425*, 22-35.

[39] V. Gabelica, et al., Recommendations for reporting ion mobility Mass Spectrometry measurements, *Mass Spectrometry Reviews* **2019**, *38*, 291-320.

[40] A. A. Shvartsburg, R. D. Smith, Fundamentals of Traveling Wave Ion Mobility Spectrometry, *Analytical Chemistry* **2008**, *80*, 9689-9699.

[41] J. A. Silveira, et al., Parallel accumulation for 100% duty cycle trapped ion mobility-mass spectrometry, *International Journal of Mass Spectrometry* **2017**, *413*, 168-175.

[42] F. Meier, et al., Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device, *Journal of Proteome Research* **2015**, *14*, 5378-5387.

[43] A. Doerr, DIA mass spectrometry, *Nature Methods* **2015**, *12*, 35-35.

[44] C. Fernández-Costa, et al., Impact of the Identification Strategy on the Reproducibility of the DDA and DIA Results, *Journal of Proteome Research* **2020**, *19*, 3153-3161.

[45] L. Krasny, P. H. Huang, Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology, *Molecular Omics* **2021**, *17*, 29-42.

[46] F. Meier, et al., Online Parallel Accumulation–Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer*, *Molecular & Cellular Proteomics* **2018**, *17*, 2534-2545.

[47] V. Marquioni, F. M. F. Nunes, M. T. M. Novo-Mansur, Protein Identification by Database Searching of Mass Spectrometry Data in the Teaching of Proteomics, *Journal of Chemical Education* **2021**, *98*, 812-823.

[48] M. Mann, R. C. Hendrickson, A. Pandey, Analysis of Proteins and Proteomes by Mass Spectrometry, *Annual Review of Biochemistry* **2001**, *70*, 437-473.

[49] M. Mann, M. Wilm, Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags, *Analytical Chemistry* **1994**, *66*, 4390-4399.

[50] J. E. Elias, S. P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, *Nature Methods* **2007**, *4*, 207-214.

[51] J. Cox, M. Mann, Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology, *Annual Review of Biochemistry* **2011**, *80*, 273-299.

[52] N. Rauniyar, J. R. Yates, Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics, *Journal of Proteome Research* **2014**, *13*, 5293-5309.

[53] A. Thompson, et al., TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set, *Analytical Chemistry* **2019**, *91*, 15941-15950.

[54] J. Cox, et al., Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ, *Molecular & Cellular Proteomics* **2014**, *13*, 2513-2526.

[55] H.-T. Lau, H. W. Suh, M. Golkowski, S.-E. Ong, Comparing SILAC- and Stable Isotope Dimethyl-Labeling Approaches for Quantitative Proteomics, *Journal of Proteome Research* **2014**, *13*, 4164-4174.

[56] C. E. Bakalarski, D. S. Kirkpatrick, A Biologist's Field Guide to Multiplexed Quantitative Proteomics, *Molecular & Cellular Proteomics* **2016**, *15*, 1489-1497.

[57] M. M. Savitski, et al., Measuring and Managing Ratio Compression for Accurate iTRAQ/TMT Quantification, *Journal of Proteome Research* **2013**, *12*, 3586-3598.

[58] J. W.-H. Li, J. C. Vederas, Drug Discovery and Natural Products: End of an Era or an Endless Frontier?, *Science* **2009**, *325*, 161-165.

[59] D. J. Newman, G. M. Cragg, Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019, *Journal of Natural Products* **2020**, *83*, 770-803.

[60] P. G. Arnison, et al., Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature, *Natural Product Reports* **2013**, *30*, 108-160.

[61] T. J. Oman, W. A. van der Donk, Follow the leader: the use of leader peptides to guide natural product biosynthesis, *Nature Chemical Biology* **2010**, *6*, 9-18.

[62] Manuel A. Ortega, Wilfred A. van der Donk, New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products, *Cell Chemical Biology* **2016**, *23*, 31-44.

[63] N. S. van der Velden, et al., Autocatalytic backbone N-methylation in a family of ribosomal peptide natural products, *Nature Chemical Biology* **2017**, *13*, 833-835.

[64] S. Ramm, et al., A Self-Sacrificing N-Methyltransferase Is the Precursor of the Fungal Natural Product Omphalotin, *Angewandte Chemie International Edition* **2017**, *56*, 9994-9997.

[65] M. A. Funk, W. A. van der Donk, Ribosomal Natural Products, Tailored To Fit, *Accounts of Chemical Research* **2017**, *50*, 1577-1586.

[66] E. Vogt, M. Künzler, Discovery of novel fungal RiPP biosynthetic pathways and their application for the development of peptide therapeutics, *Applied Microbiology and Biotechnology* **2019**, *103*, 5567-5581.

[67] S. Luo, S.-H. Dong, Recent Advances in the Discovery and Biosynthetic Study of Eukaryotic RiPP Natural Products, *Molecules* **2019**, *24*, 1541.

[68] J. A. McIntosh, M. S. Donia, E. W. Schmidt, Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds, *Natural Product Reports* **2009**, *26*, 537-559.

[69] A. Essig, et al., Copsin, a Novel Peptide-based Fungal Antibiotic Interfering with the Peptidoglycan Synthesis*, *Journal of Biological Chemistry* **2014**, *289*, 34953-34964.

[70] G. Férir, et al., The Lantibiotic Peptide Labyrinthopeptin A1 Demonstrates Broad Anti-HIV and Anti-HSV Activity with Potential for Microbicidal Applications, *PLOS ONE* **2013**, *8*, e64010.

[71] Y. Fu, A. H. Jaarsma, O. P. Kuipers, Antiviral activities and applications of ribosomally synthesized and post-translationally modified peptides (RiPPs), *Cellular and Molecular Life Sciences* **2021**, *78*, 3921-3940.

[72] K. I. Mohr, et al., Pinensins: The First Antifungal Lantibiotics, *Angewandte Chemie International Edition* **2015**, *54*, 11254-11258.

[73] H. E. Hallen, H. Luo, J. S. Scott-Craig, J. D. Walton, Gene family encoding the major toxins of lethal Amanita mushrooms, *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 19097-19101.

[74] W. Ding, et al., Biosynthetic investigation of phomopsins reveals a widespread pathway for ribosomal natural products in Ascomycetes, *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 3521-3526.

[75] R. D. Johnson, et al., A novel family of cyclic oligopeptides derived from ribosomal peptide synthesis of an in planta-induced gene, gigA, in Epichloë endophytes of grasses, *Fungal Genetics and Biology* **2015**, *85*, 14-24.

[76] K. J. Hetrick, W. A. van der Donk, Ribosomally synthesized and post-translationally modified peptide natural product discovery in the genomic era, *Current Opinion in Chemical Biology* **2017**, *38*, 36-44.

[77] P. Agrawal, et al., RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links, *Nucleic Acids Research* **2017**, *45*, W80-W88.

[78] A. J. van Heel, et al., BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides, *Nucleic Acids Research* **2013**, *41*, W448-W453.

[79] G. A. Vignolle, R. L. Mach, A. R. Mach-Aigner, C. Derntl, Novel approach in whole genome mining and transcriptome analysis reveal conserved RiPPs in Trichoderma spp, *BMC Genomics* **2020**, *21*, 258.

[80] R. Breitling, A. Ceniceros, A. Jankevics, E. Takano, Metabolomics for Secondary Metabolite Research, *Metabolites* **2013**, *3*, 1076-1083.

[81] H. Mohimani, et al., NRPquest: Coupling Mass Spectrometry and Genome Mining for Nonribosomal Peptide Discovery, *Journal of Natural Products* **2014**, *77*, 1902-1909.

[82] C. Du, G. P. van Wezel, Mining for Microbial Gems: Integrating Proteomics in the Postgenomic Natural Product Discovery Pipeline, *Proteomics* **2018**, *18*, 1700332.

[83] K. Dettmer, P. A. Aronov, B. D. Hammock, Mass spectrometry-based metabolomics, *Mass Spectrometry Reviews* **2007**, *26*, 51-78.

[84] T. Hautbergue, et al., From genomics to metabolomics, moving toward an integrated strategy for the discovery of fungal secondary metabolites, *Natural Product Reports* **2018**, *35*, 147-173.

[85] K. F. Nielsen, T. O. Larsen, The importance of mass spectrometric dereplication in fungal secondary metabolite analysis, *Frontiers in Microbiology* **2015**, *6*.

[86] V. Seidel, in *Natural Product Isolation* (Ed.: S. D. Sarker), Humana Press, **2005**.

[87] J. L. Markley, et al., The future of NMR-based metabolomics, *Current Opinion in Biotechnology* **2017**, *43*, 34-40.

[88] T. Fuhrer, D. Heer, B. Begemann, N. Zamboni, High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection–Time-of-Flight Mass Spectrometry, *Analytical Chemistry* **2011**, *83*, 7074-7080.

[89] B. Buszewski, S. Noga, Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique, *Analytical and Bioanalytical Chemistry* **2012**, *402*, 231-247.

[90] K. Masike, M. A. Stander, A. de Villiers, Recent applications of ion mobility spectrometry in natural product research, *Journal of Pharmaceutical and Biomedical Analysis* **2021**, *195*, 113846.

[91] G. Paglia, et al., Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications, *Analytical Chemistry* **2014**, *86*, 3985-3993.

[92] PREOMICS, 2021, iST Sample Preparation Kit, accessed 20.08.2021, https://www.preomics.com/resources

[93] D. Szklarczyk, et al., STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Research* **2018**, *47*, D607-D613.

[94] P. A. Tucker, L. Sallai, The AAA+ superfamily—a myriad of motions, *Current Opinion in Structural Biology* **2007**, *17*, 641-652.

[95] M. Umemura, et al., Characterization of the biosynthetic gene cluster for the ribosomally synthesized cyclic peptide ustiloxin B in Aspergillus flavus, *Fungal Genetics and Biology* **2014**, *68*, 23-30.

[96] N. Nagano, et al., Class of cyclic ribosomal peptide synthetic genes in filamentous fungi, *Fungal Genetics and Biology* **2016**, *86*, 58-70.

[97] S. Shabani, J. M. White, C. A. Hutton, Total Synthesis of the Putative Structure of Asperipin-2a and Stereochemical Reassignment, *Organic Letters* **2020**, *22*, 7730-7734.

[98] T. Shan, et al., Determination and Analysis of Ustiloxins A and B by LC-ESI-MS and HPLC in False Smut Balls of Rice, **2012**, *13*, 11275-11287.

[99] M. Wang, et al., Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking, *Nature Biotechnology* **2016**, *34*, 828-837.