Günter Allmaier

Diplomarbeit

Characterisation of proteins derived from *Malus domestica*

Ausgeführt am Institut für

Chemische Technologien und Analytik

der Technischen Universität Wien

unter der Anleitung von

Univ.Prof. Mag.pharm. Dr.rer.nat. Günter Allmaier

durch

Stephanie Steinberger, BSc.

Brigittenauer Lände 30/7, 1200 Wien

Wien, Dezember 2018

Stephanie Steinberger

Acknowledgement

I would like to thank Prof. Günter Allmaier and Prof. Heidi Halbwirth for giving me the opportunity to work on this project and to become part of their working groups. In addition, I would also like to express my gratitude to Austrian Science Fund (FWF,project number P25399-B16) for funding and supporting this research.

A special thanks to Prof. Martina Marchetti-Deschmann and Dr. Victor Weiss for their tremendous help, their support and their know-how they shared with me. Moreover, also a special thanks to Dr. Christian Molitor and Lukas Eidenberger BSc for the isolation, extraction and purification of the protein, for all the data they and for giving me the chance to participate in the process as well.

I would also like to thank the members of the working group, new and old, for their warm welcome, their support and all the times were on hand with help and advice to solve problems. The friendships that formed, the moments we shared, and the inspiring lunch conversations will be cherished forever. I can never express it in words how much I appreciate you: Sam, Vici and Jessi becoming my friends and making me part of your crazy family. I couldn't imagine how boring life would be without you in it.

Furthermore, I would also like to express my appreciation to my friends Johanna, Mira, Chrisi and Alex for being my shoulders to lean on and always having sympathetic ears for me.

Thanks also to all the other student colleagues that started this wondrous journey with me and that I will always remember, when I think back of this time.

Last but not the least, I would like to thank my family, especially my mum and dad, from the bottom of my heart for giving me the opportunity to learn at the university and their endless support and patience. The love I hold for you knows no bounds. I would also wish to thank my cat Lobsie for all the cuddling sessions she gracefully endured.

Thank you all.

„Wenn die Neugier
sich auf ernsthafte Dinge richtet,
dann nennt man sie Wissendrang.“

-Marie von Ebner-Eschenbach (1830 - 1916)

Content

1a. Abstract

Phloretin and phloridzin are currently in the focus of health and cancer research due to their beneficial effects. Both compounds are found to a various degree in the plant family of Rosaceae, but especially in *Malus domestica*, which contains the highest amount of those polyphenols. The biosynthesis pathway of the formation of phloretin is not yet fully elucidated due to the presence of multiple possible substrates. Likewise, the enzyme(s) required for the formation is still not characterized and identified.

For the identification of the specific enzyme(s) (i.e. protein(s)) were isolated and extracted from young apple leaves (*M. domestica* cv. Golden Delicious) using Aqueous Two-Phase System (ATPS) with Triton X-114 as well as PEG 4000. The purification of the protein(s) was achieved with Fast Protein Liquid Chromatography (FPLC) using three successive strong cation exchange columns.

Samples from the different chromatographic steps were taken and analyzed by means of a capillary electrophoresis-on-a-chip system with an integrated laser-induced fluorescence detection to determine the protein amount in the samples and to monitor the purification steps of the protein sample.

For protein identification a Peptide Mass Fingerprint (PMF) approach was performed. Therefore, a 1D SDS-PAGE was done with replicates of the sample to provide reliable results. The gels were stained with a mass spectrometry-compatible silver approach. Subsequently, eight gel bands were excised and *in-gel* digestion using a trypsine/Lys C mixture was performed. Moreover, a second *in-gel* digestion approach with higher amounts of formic acid (FA) and acetonitrile (ACN) was performed to extract more hydrophobic peptides.

The extracted tryptic peptides were purified with ZipTip pipette tips (C_{18} material) and subsequently analyzed with Matrix-Assisted Laser Desorption/Ionization Reflectron Time-Of-Flight (MALDI RTOF) and MALDI TOF/RTOF mass spectrometer (MS) to obtain mass spectra as well as tandem mass spectra (MS^2).

The acquired PMF data and MS^2 data were applied to search with the Mascot Search engine to compare and match the data to *in-silico* digested proteins.

As a result, three out of eight analysed gel bands were identified with a 95 % probability of significance – alpha-xylosidase 1-like, cucumisin-like isoform X1 or X2, beta-glucosidase 12-like or isoforms X1 or X2. However, the originally suggested protein of interest could not be identified.

In the future, further improvement of protein isolation and purification will be required as well as an approach with 2D gel electrophoresis and MS should be performed to be able to identify unambiguously the protein(s) related to phloretin biosynthesis and characterize it (them).

1b. Kurzfassung

Phloretin und Phloridzin sind derzeitig aufgrund ihrer positiven Wirkung auf das menschliche Befinden in den Fokus von Gesundheits- und Krebsforschung gerückt. Beide Komponenten sind in unterschiedlichen Mengen in der Pflanzen der Familie Rosaceae vorhanden, vor allem in *Malus domestica*, der die höchste Menge an diesen Polyphenolen beinhaltet. Der Biosyntheseweg von Phloretin ist aufgrund der Anwesenheit von mehreren möglichen (Ausgangs-)Substraten noch nicht vollkommen aufgeklärt. Ähnlich verhält es sich mit den Enzymen, die für die Bildung von Phloretin nötig sind.

Für die Identifizierung dieser spezifischen Enzyme, wurden Proteine von jungen Apfelblättern (*M. domestica* cv. Golden Delicious) mittels wässriger zwei-Phasen Extraktion sowohl mit Triton X-114 als auch mit PEG4000, extrahiert und isoliert. Die Aufreinigung der Proteine erfolgte mittels Schnell-Leistungs-Flüssig-Chromatographie (fast protein liquid chromatography, FPLC) anhand von drei aufeinanderfolgenden starken Kationenaustauschersäulen.

Proben von den verschiedenen chromatographischen Aufreinigungsschritten wurden genommen und anhand von Gel-Kapillarelektrophorese auf einem Chip mit integriertem Laser-induzierten Fluoreszenz Detektor analysiert, um die Proteinmenge in den Proben zu ermitteln als auch die Aufreinigungsschritte an sich nachzuverfolgen und zu vergleichen.

Für die Protein Identifizierung wurde eine Peptidmassenfingerprint-Analyse (PMF) durchgeführt. Dafür wurde ein 1D SDS-PAGE mit mehreren Replikaten der Proben durchgeführt, um verlässliche Daten zu generieren. Anschließend wurden die Gele mit einer Massenspektrometrie-kompatiblen Silberfärbung eingefärbt, die acht Hauptbanden wurden ausgeschnitten und mit einer Trypsin/LysC Mischung wurde ein *In-gel* Verdau durchgeführt. Des Weiteren wurde eine zweite Herangehensweise des *In-Gel* Verdau mit höheren Konzentrationen an Ameisensäure (FA) und Acetonitril (ACN) durchgeführt, um hydrophobere Peptide zu extrahieren.

Die extrahierten trypsin/LycC-generierten Peptide wurde anhand von einem ZipTip (C_{18} Material) Aufreinigungsschritt gereinigt und anschließend mittels Matrix-unterstützter Laser-Desorption/Ionisation (MALDI) Reflektron (RTOF) und MALDI-Linearflugzeit (LTOF)/RTOF Massenspektrometrie (MS) analysiert, um Massenspektren als auch Tandem-Massenspektren (MS^2) zu erhalten.

Die erhaltenen PMF Daten als auch die MS^2 Daten wurden für die Suche in der Mascot Search Engine (ein Suchalgorithmus), in der die Daten mit *in-silico* verdauten Proteinen verglichen und zugeordnet werden, verwendet.

Infolge dessen konnten drei von acht Gelbanden mit einer Signifikanz von 95% identifiziert werden – alpha-xylosidase 1-like, cucumisin-like Isoform X1 und X2 sowie beta-glucosidase 12-like oder deren Isoformen X1 und X2. Das ursprünglich vorgeschlagene Enzym konnte allerdings nicht identifiziert werden.

Für eine eindeutige Identifizierung und Charakterisierung des vorgeschlagenen Enzyms oder der Enzyme bzw. Isoformen des Phloretin-Biosynthesewegs könnte zukünftig eine weitere Verbesserung der Proteinisolierung und -aufreinigung mit anschließender 2D Gelelektrophorese und MS erfolgen.

1.1 List of abbreviations

$(NH_4)_2SO_4$	Ammonium sulfate
ACN	Acetonitrile
AEX	Anion exchange chromatography
$AgNO_3$	Silver nitrate
ATPS	Aqueous two-phase system
CEX	Cation exchange chromatography
CHCA	α-Cyano-4-hydroxycinnamic acid
CID	Collision-induced dissociation
DTT	DL-Dithiothreitol
EtOH	Ethanol
FA	Formic acid
FPLC	Fast protein liquid chromatography
G6P	Glucose-6-phosphate
G6PD	Glucose-6-phosphate-dehydrogenase
HAc	Acetic acid
HCl conc.	Hydrochloric acid, concentrated (37 %)
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HPLC	High performance liquid chromatography
IAA	Iodoacetamide
ISD	In-source decay
KPi	Potassium phosphate buffer
MALDI	Matrix-assisted laser desorption/ionization
MCP	Microchannel plate
MeOH	Methanol
MES	2-(*N*-morpholino)ethanesulfonic acid
$MgCl_2$	Magnesium chloride
MOPS	3-(*N*-morpholino)propanesulfonic acid
MS	Mass spectrometry
MW	Molecular weight

Na_2CO_3	Sodium carbonate
$Na_2HPO_4 \times 2\ H_2O$	Di-sodium hydrogen phosphate
$Na_2S_2O_3 \times 5H_2O$	Sodium thiosulfate -5-hydrate
Na-ascorbate	Sodium ascorbate
NaOAc	Sodium acetate
NaOH	Sodium hydroxide
NH_4HCO_3	Ammonium bicarbonate
PEG	Polyethylene glycol
PMF	Peptide mass fingerprint
PSD	Post-source decay
RTOF	Reflectron time of flight
SDS	Sodium dodecyl sulfate
SEM	Secondary electron multiplier
TFA	Trifluoroacetic acid
TOF	Time of flight

2. Introduction

2.1 Aim of the thesis

Phloretin is a polyphenolic compound found in plants in various degrees. Higher amounts of this polyphenol are found in the family of Rosaceae, which also includes the common apple, *Malus domestica*. Within the family of Rosaceae, the common apple contains the highest amount of phloretin.

In the last few years phloretin and its derivate phloridzin, also termed phlorizin, has been brought back into the research spotlight especially in the context of health and cancer research [1].

But as to date the biosynthesis pathway of phloretin is not yet entirely elucidated, three possible substrates for the conversion to phloretin exist. Nowadays, phloretin is still extracted from apple foliage, but with the knowledge of the biosynthesis pathway, the involved specific enzyme and its substrate, there could be a major future advantage for pharmaceutical applications: With the identified protein (enzyme) and its sequence, phloretin could be produced during a fermentation process and thus be on hand in a higher quantity as well as quality.

The aim of this thesis was the identification of proteins derived from *M. domestica*, with a special focus on one possible enzyme of interest, putative *p*-coumaroyl-CoA reductase being supposedly involved in phloretin biosynthesis.

The project of the characterization of proteins for subsequent X-ray crystallographic investigations and identification of enzymes from the common apple, *Malus domestica*, was a cooperation between Prof. Heidi Halbwirth and Prof. Günter Allmaier and funded and supported by the Austrian Science Fund (FWF, project number P25399-B16). Therefore, the experimental work was split between the working groups of Prof. Halbwirth and Prof. Allmaier. The isolation, extraction and purification of the protein(s) was performed at the Institute of Chemical, Environmental and Bioscience Engineering, while the bioanalytical aspect and analysis via PMF and tandem MS for protein identification was performed at the Institute of Chemical Technologies and Analytics.

An overview of the isolation, extraction and purification of the protein(s), which was performed by a master student in Prof. Halbwirth's working group, is given in this work, but the main focus of this thesis was the identification of the purified target enzyme, i.e. protein(s).

2.2. Strategy for the protein characterization

In order to achieve the goal of the characterization of the protein of interest as well as other proteins, a PMF approach was chosen. At first the proteins were isolated, extracted and purified via FPLC from young *M. domestica* leaves, which were harvested solely for this experiment. The purified proteins were then analyzed with 1D gel electrophoresis, followed by *in-gel* digestion, ZipTip purification and MALDI RTOF MS. Additionally, Tandem MS is used to fragment single peptides and gain information about the peptide sequence tags. A schema of the workflow is shown in figure 1. All the obtained data is then compared to an online database to identify the proteins.

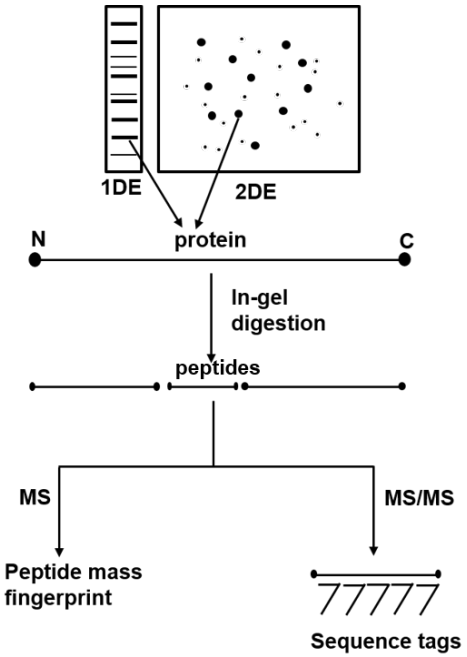


Figure 1: Schema of the workflow of the PMF

2.3 Phloretin

Phloretin is a polyphenol and its derivate phloridzin, a dihydrochalcone, are present in the family of Rosaceae, but especially in *M. domestica* in high amounts, which gives apple a unique status. While many studies have been done on the use of phloretin and phloridzin on human health and their benefits in treatments of diabetes, obesity, neurological diseases and even cancer as well as their use in the food industry as natural sweeting agents and longevity-extending agents, the physiological relevance for the apple is still unknown. Furthermore, also the knowledge of the biosynthesis of dihydrochalcones is limited [2-4].

For the isolation of the protein apple foliage from *M. domestica* cv. Golden Delicious was chosen due to its higher amount of expressed phloretin and phloridzin in the young leaves during the months May and June [5].

The possible biosynthesis pathway of phloretin and phloridzin is displayed in figure 2. Additionally, the possible substrates and ways for the formation of phloretin are shown in figure 3. Phloretin can either be synthesized in the plants from *p*-coumaric acid or *p*-coumaroyl-CoA to dihydro-p-coumaric acid or dihydro-*p*-coumaroyl-CoA, respectively, as intermediates or from the naringenin chalcone directly to phloretin.

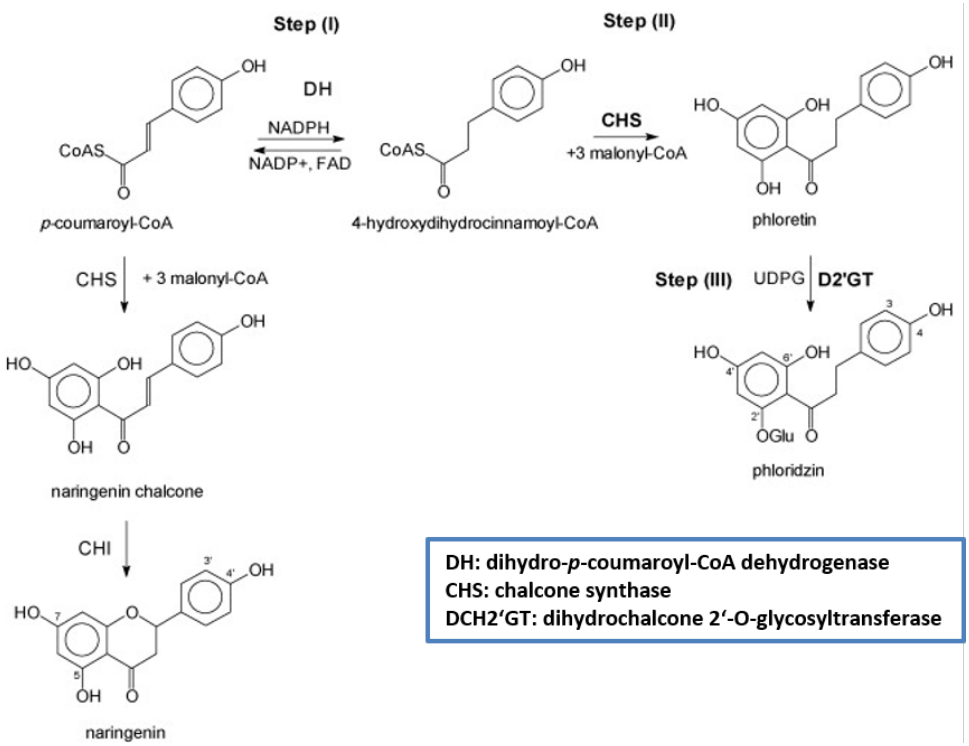


Figure 2: Biosynthesis pathway of phloretin and phloridzin[3]. 4-hydroxydihydrocinnamoyl-CoA is also known as dihydro-p-coumaroyl-CoA.

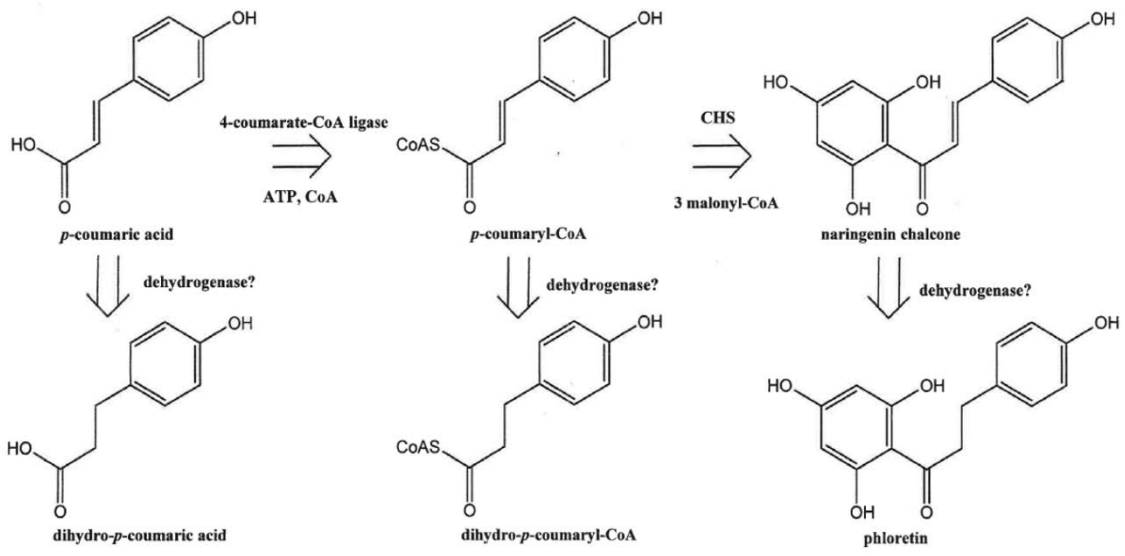


Figure 3: Possible substrates and ways for the formation of phloretin [3].

2.4 Protein extraction

The Aqueous Two-Phase System (APTS) is a liquid-liquid fractionation technique, which is used for the extraction, separation, purification and enrichment of protein, membranes, viruses and nucleic acids [6].

These systems can consist of two polymers or one polymer and one salt component dissolved / suspended in water, respectively. Thus, it forms a gentle environment for biomolecules to separate and the polymers stabilize the structure and biological activity of the analytes in question [6].

The most common biphasic system containing two polymers is usually polyethylene glycol (PEG) and dextran, for a polymer and a salt it is PEG and a phosphate, sulfate or citrate. Polymers are usually not miscible with aqueous solutions and thus the formation of two-phases takes place. In the polymer-salt ATPS, the salt absorbs large amounts of water and the formation of two phases also occurs [6]. A scheme of the phase diagram is displayed in figure 4.

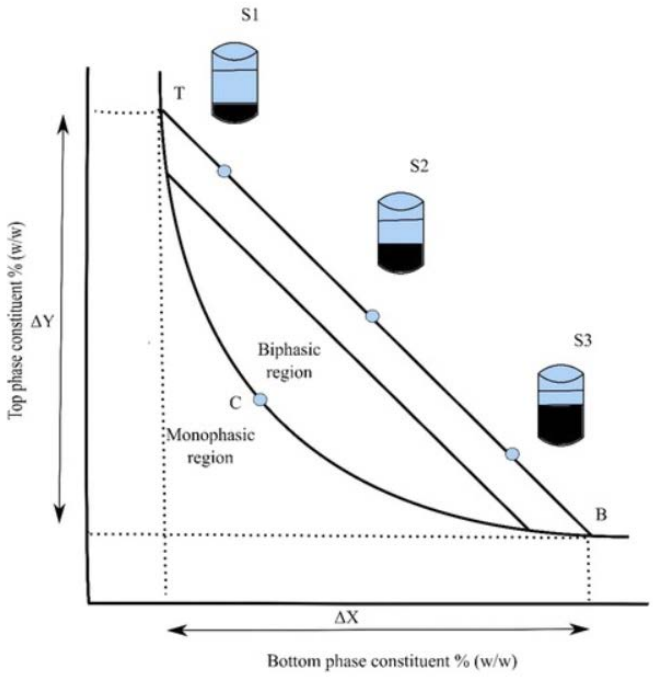


Figure 4: Scheme of an APTS phase diagram. The binodal curve (TCB) is the border and all concentrations above the curve form aqueous two-phase system (e.g. S1, S2, S3). Below the binodal curve the two components form a one-phase system, because they are miscible [6].

Important factors for the separation are the concentration and the molecular weight (MW) of the polymer and the concentration and composition of the used salt. The higher the MW of the polymer, the lower the concentration of polymer required for the phase formation. Additionally, the higher the differences between the MW of the polymers, the higher is the asymmetry of the binodal curve.

An ATPS is very dependent on the pH and the temperature and is sensible to changes. Furthermore, the hydrophobicity and the salting out effect also have an impact on the separation. The Hofmeister series (figure 5) displays the ability of ions to salt out or salt in proteins. Sulfate possesses a high salting out ability and is therefore used for the analytical problem presented in my thesis during the extraction

to precipitate the proteins. PEG is used to eliminate the polyphenol oxidases from the sample, which is visible in the yellow colour of the PEG phase.

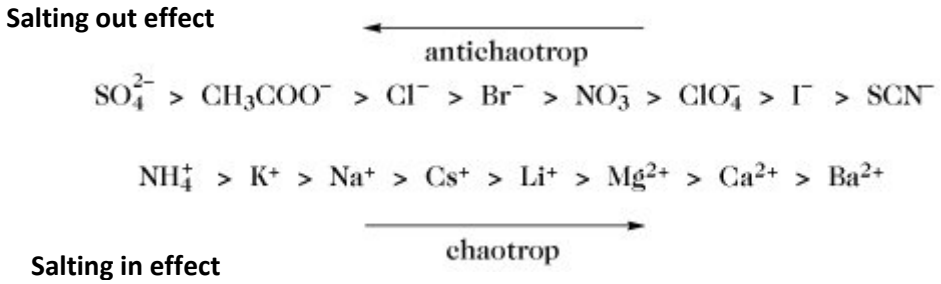


Figure 5: Hofmeister series of antichaotrop and chaotrop ions.

2.5 Liquid Chromatography (LC)

Liquid chromatography is a separation method, where the soluble analytes of a sample are distributed between a stationary phase and a mobile phase, which moves in either axial or radial direction. The analytes migrate through the column with different velocity based on their interaction with the stationary phase.

Chromatographic steps usually include:

- Equilibration step of the stationary phase
- Sample application
- Analyte elution step

The detection of the components can occur e.g. via UV spectroscopy or mass spectrometry [7].

The retention factor k describes the retention of analytes independent form the column dimension or the flow rate (equation 1).

$$k = \frac{t_1 - t_0}{t_0}$$

Equation 1: Retention factor of an analyte with t_1 = retention time of the analyte, t_0 = column void time. [7]

To resolve two analytes their retention factors must be different. The selectivity α describes the ability of the column to separate two analytes in accordance to their retention factors (equation 2).

$$\alpha = \frac{k_2}{k_1}$$

Equation 2: Selectivity of the column with $k_{1/2}$ = retention factor of analyte 1 or 2. [7]

The quality of a separation is not only defined by the selectivity but also the resolution R_S of two neighboring peaks in the chromatogram (equation 3).

$$R_S = \frac{t_{R2} - t_{R1}}{\frac{1}{2}(w_1 + w_2)}$$

Equation 3: Resolution of two peaks with $t_{R1/2}$ = retention time of peak 1 or 2, $w_{1/2}$ = peak width of peak 1 or 2. [7]

The peak broadening is demonstrated in the column efficiency, which is defined as the plate number N or the plate height H (equation 4).

$$N = \frac{L}{H} = 16\left(\frac{t_R}{w}\right)^2$$

Equation 4: Plate number with L = column length, H = plate height, t_R = retention time, w = peak width. [7]

For the comparison of column efficiencies of columns with the same bed dimensions but different stationary material the reduced plate height h can be calculated (equation 5).

$$h = \frac{H}{d_P}$$

Equation 5: Reduced plate height h with H = plate height, d_P = particle diameter.[7]

The peak broadening that is caused by different mass transfer effects can be described with the van Deemter-Knox equation (equation 6). Hence the obtained column efficiency depends on the applied velocity of LC separation.

$$h = A + \frac{B}{u} + Cu$$

Equation 6: van Deemter-Knox equation with h = reduced plate height, A = Eddy diffusion and mobile phase mass transfer effects, B = longitudinal molecular diffusion effect, C = interaction of the analytes with the stationary phase, u = reduced velocity ($u = L/t_0$). [7]

For the purification of the protein samples extracted from *M. domestica* Fast Protein Liquid Chromatography in the Ion Exchange mode was chosen. Therefore, only those methods will be explained in more detail.

2.5.1 FPLC

Fast Protein Liquid Chromatography (FPLC), sometimes also called Fast Performance Liquid Chromatography, was developed by Pharmacia (Uppsala, Sweden) in 1982. The method provides a full range of chromatographic modes like Ion Exchange, Gel Filtration, Hydrophobic Interaction and Reversed Phase, which are all based on particles with average diameter size in the range of HPLC column material. In comparison to conventional HPLC columns, higher protein loadings can be applied on the FPLC. The buffer system consists of a wide range of aqueous and biocompatible substances [8, 9].

Most often the FPLC is used in combination with an Ion Exchange or Gel Filtration approach. However, before samples can be applied on the FPLC some prior purification steps are usually necessary.

Normally the method is performed at room temperature, but for labile proteins a reduced temperature e.g. at 4 °C is also possible.

2.5.1.1 Ion Exchange Chromatography

The Ion Exchange Chromatography (IEC) was first introduced in the 1960s. Since then the popularity of the method has increased steadily and nowadays it is the preferred technique for the purification of proteins due the fact that the biological activity is not reduced.

Immobilized ligands are bound to the stationary phase and the separation occurs due to the interaction of the charged surface of the analytes with the complimentary charges on the surface of the stationary phase. The analytes are separated according to their negative net charge in the Anion Exchange Chromatography (AEX), while in the Cation Exchange Chromatography (CEX) according to their positive net charge. The retaining mobile phase is aqueous and has a low pH and low salt concentration. The elution buffer, however, is aqueous and possesses a high salt concentration and a low or high pH. A strong CEX can keep the negative charge on the stationary phase over a large pH range without any variations and is therefore ideal for peptide and protein purification. [7, 10]

If the buffer possesses a lower pH than the pI of the protein in the CEX, the protein stays bound to the stationary phase, while for the AEX a buffer of higher pH must be used for the same effect. When the pH of the mobile phase correlates to the pI of the protein, the protein is not retained on the column and is eluted. The elution can be performed with an isocratic, step or gradient profile.

This is a rather simplistic view on the process. Newer studies have shown that the interaction between protein and stationary phase are dependent on charge density of stationary phase, the mobile phase composition, and on the number and distribution of charges sites on the protein itself. Changes in the 3D structure of the protein have a big impact on the protein retention time. [7, 10]

2.5.3 Biological activity tracking

To monitor the purification of the protein or enzyme of interest and to follow the chromatographic steps, often an enzyme assay is performed to track the biological activity. Therefore, a test suitable to the specific protein has to be designed with appropriate values for the parameter's temperature, time, pH, substrate and its concentration and buffer.

For this work a two-step enzyme assay was created by the working group of Prof. Halbwirth (unpublished data).

The first step was the reduction of *p*-coumaroyl-CoA to dihydro-*p*-coumaroyl-CoA with NADPH as a coenzyme, which is the reaction done by the enzyme of interest, a putative *p*-coumaroyl-CoA reductase. Therefore, glucose-6-phosphate (G6P) as the substrate and glucose-6-phosphate-dehydrogenase (G6PD), the correlating enzyme, were added to the enzyme assay. G6PD oxidizes the substrate G6P to 6-phosphogluconolactone and reduces $NADP^+$ to NADPH. This is important to supply the test enzyme reaction with a stable amount of coenzyme.

The second step is the saponification of dihydro-*p*-coumaroyl-CoA and *p*-coumaroyl-CoA to dihydro-*p*-coumaric acid and *p*-coumaric acid, respectively, by hydrolysis. This step is necessary due to the fact that reduction of one double-bond is not enough to distinguish and separate dihydro-*p*-coumaroyl-CoA from *p*-coumaroyl-CoA. However, dihydro-*p*-coumaric acid and *p*-coumaric acid can be separated by means of a Reverse-Phase HPLC and thus can indirectly demonstrate the biological enzyme activity.

2.6 Electrophoresis

Electrophoresis is one of the most frequently applied separation techniques for biological samples. High separation efficiency can be achieved with a relatively simple setup.

The method is based on the principle of migration of charged molecules under the influence of an electric field. Due to their different sizes / shapes and charges, molecules will migrate with different velocities and can thus be separated (figure 6).

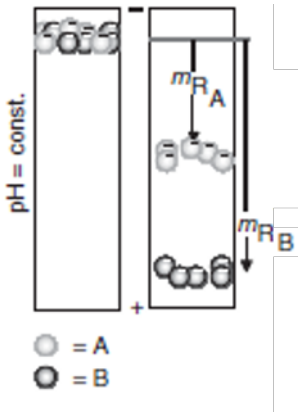


Figure 6: Electrophoretic separation principle for a zone electrophoresis. The molecules A and B are both part of the sample. They differ in their sizes / shapes as well as their charges and can be separated in an electric field [11].

Electrophoretic separation methods can be carried out in solutions or employing stabilizing media like thin-layer plates, films and gels. The zone electrophoresis, simply called electrophoresis, is performed in a homogenous buffer system over the whole separation time and range and thus supplies a constant pH value. The migration itself is limited to the electrophoretic mobility of the analytes. It can be applied for both amphoteric and non-amphoteric molecules.

In addition to the sample molecules, also the supporting media, i.e. the surface of the equipment like capillaries or glass plates carry reactive groups. These groups get ionized in the buffers and therefore they get attracted toward the anode during the separation process. However, they are fixed in place and cannot migrate themselves, thus an electroosmotic effect is created. This defines the counterflow of H_3O^+ ions or equal charged buffer species toward the cathode. Especially in the gel electrophoresis, this effect is observed visually. In capillary electrophoresis the effect is termed electroosmotic flow. The electroosmotic flow is opposite the electrophoretic migration as displayed in figure 7. [11]

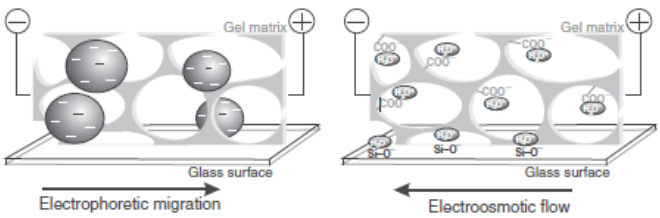


Figure 7: Principle of the electroosmotic flow in comparison to the electrophoretic migration [11].

In the course of this thesis gel electrophoresis and protein separation with a microchip-based gel electrophoretic approach were used, therefore only those methods are described in detail in the following section.

2.6.1 Gel electrophoresis

Gel electrophoresis is based on separation of analytes according to their size in a porous gel matrix. The gel matrix can be made of starch, dextran, agarose or polyacrylamide, which differ in their pore sizes, cross-linking and the cross-linking reagents. The gel can be run in a flatbed or vertical system, depending on the sample and the separation requirements. Due to the better resolution gel layers ≤1 mm are preferred. The advantages of thinner gels are the better heat dissipation, faster separation, better defined bands, faster staining and higher sensitivity.

In addition, discontinuous electrophoresis can be performed and hence prevents protein aggregation and precipitation during the liquid sample introduction and it boosts the formation of well-defined protein bands. The gel is divided into two parts: the stacking gel and the resolving gel. The stacking gel has larger pores compared to the resolving gel and is therefore excellent to prevent protein aggregations. When the electric field is applied, the proteins migrate and form stacks according to their mobilities in the stacking gel and migrate toward the resolving gel. There they are slowed down by the smaller pore diameter and zone becomes higher concentrated. In the resolving gel they are then separated according to their sizes and charges and the separation increases in speed. [11]

Another commonly used technique is the gradient gel electrophoresis. A pore gradient gel, either linear or exponential, is obtained due to the continuously changing of the acrylamide concentration along the gel. When the acrylamide concentration as well as the cross-linking are high enough in the small pore area, even small protein molecules can be driven until their end point and be trapped within the gel matrix. [11]

SDS electrophoresis is the most used electrophoresis and was first introduced by Shapiro, Viñuela and Maizel in 1967 [12]. This method separates the molecules exclusively by their molecular mass. Therefore, the protein surface is loaded with the anionic detergent SDS, which forms anionic micelles around the proteins and conceals their charges. In addition, the secondary and tertiary structures are unfolded due to the disruption of the hydrogen bonds. For the cleavage of disulfide bonds between cysteines, a reducing agent like dithiothreitol or 2-mercapotoethanol is required. To prevent the re-folding of the proteins a subsequent alkylation with iodoacetamide, iodoacetic acid or vinyl pyridine is done. Continuous, discontinuous and gradient gels can be used for SDS gel electrophoresis. The advantages of SDS electrophoresis are the rapid separation, the high resolution, the staining efficiency, the solubilization of even very hydrophobic and denaturated proteins and the separation solely on the molecular weight. [11]

2.6.2 Microchip electrophoresis

The employed microchip electrophoresis is essentially a miniature gel electrophoresis in a planar device. It allows for quicker analysis and a certain degree of automation. The separation, sample application and fluorescence staining of analytes / destaining of the sample background is done via narrow channels on glass slides. Etching of glass slides is accomplished with a photolithographic process. Those narrow channels are usually positioned in a cross-like pattern, depending on the number of sample slots. The reservoirs are equipped with electrolyte solutions for sample injection and all the necessary separation steps. In addition, the sample volume needed for the technique is in the nanoliter range, which is of utmost importance for bioanalytical processes where high sample volumes are only rarely obtained.

2.6.3 Staining methods

In general, staining methods rely on covalent and non-covalent binding of a dye molecule to the protein or the staining of the gel, but not the proteins. Fluorescence staining is an example for covalent binding of a dye molecule to an analyte, while Coomassie Blue and silver staining are for non-covalent binding. For a fast and inexpensive quantification Coomassie Blue staining is the method of choice. If more sensitivity is needed, silver or fluorescence staining are chosen.

2.6.3.1 Coomassie Brilliant Blue Staining

Coomassie Brilliant Blue staining is a steady state, fast and efficient staining method. Additionally, the method yields good quantification of the proteins and is inexpensive as well as MS compatible [10]. The improvement of Coomassie Brilliant Blue staining was established in 1988 by Neuhoff et. al. [13]. Colloidal Coomassie G-250 was introduced as a dye molecule, which reduced the detection limit down to 50 – 100 ng. The dye molecule binds to positively charged amino acids like arginine, tyrosine, lysine and histidine.

2.6.3.2 Silver Staining

Silver staining excels due to its high sensitivity and a detection limit of 0.1 - 0.2 ng, which is therefore 100 times more sensitive than Coomassie Brilliant Blue staining. Silver staining is a multistep procedure, thus more laborious than Coomassie Brilliant Blue staining, and additionally the quantification is limited. Due to the replacement of glutaraldehyde with formaldehyde (very low amount 0.04 %) it is also MS compatible. The theory of the mechanism is based on the agglomeration of Ag^+ ions at the amino acids' glutamine, asparagine and cysteine and their reduction by formaldehyde to elemental silver. This process is often compared to the one of black and white photography [10, 14-16].

2.6.3.3 Fluorescence Staining

Fluorescence staining is the preferred technique next to Coomassie Brilliant Blue and silver staining due to its high sensitivity. The most used fluorescence dyes for SDS-gels are SYPRO stains. The SYPRO dyes bind covalently to the protein-SDS complex. The method is very effective and also exhibits very good results for quantification. Compared to the other two staining methods fluorescence staining it is just as laborious as silver staining and also a multistep staining method. The limit of detection is around 1 ng and can be even lower, depending on the protein. All fluorescence dyes are light sensitive and therefore have to be stored in the dark [10, 17].

2.6.3.3.1 Laser-induced Fluorescence (LIF)

The commonly used technique for detection in microchip electrophoresis is the Laser-induced fluorescence (LIF) method, also known as Laser-stimulated fluorescence (LSF). Herein, fluorescent dyes, which bind either covalently to the proteins or non-covalently to the SDS-protein complexes, are used. The detection of proteins occurs during the migration of the fluorophore-system along the capillary window with a laser that emits at a wavelength, corresponding to the specific dye [18].

2.7 Mass spectrometry

Mass spectrometry (MS) is an analytical technique to identify unknown analytes, verify the chemical composition of the molecules and to identify post-translational modifications like phosphorylation or glycosylation. Even quantification of analytes is possible. The method is based on the ionization of molecules and the separation of the charge molecules according to their mass-to-charge ratio (m/z) in a high vacuum. The first MS instrument was invented in the 1900s by J. J. Thomson, after he discovered the mass of electrons. Since then MS was steadily improved and new ionization techniques, mass analyzers and hybrid instruments were created [19].

MS is based on the generation of gas-phase ions, which happens in the ion source of the instrument. These ions are afterwards separated in the mass analyzer according to their m/z values, followed by their detection in the ion detector. The data is then collected by the system and is converted into a mass spectrum. The schema of a typical mass spectrometer is shown in figure 8.

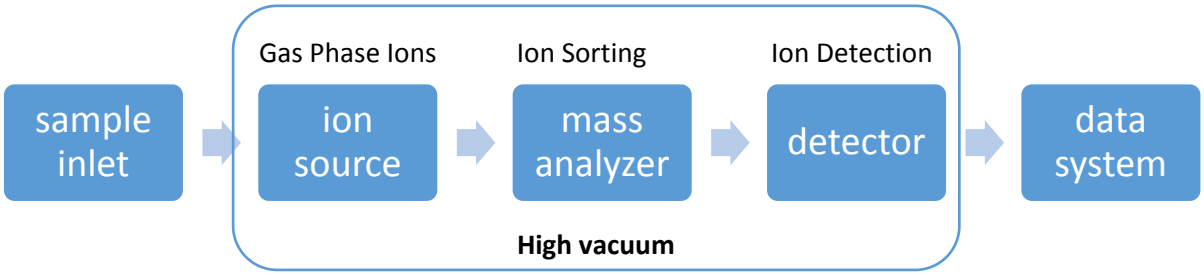


Figure 8: Schema of a typical mass spectrometer.

2.7.1 Ion source

In general, there is a distinction between soft and hard ionization methods. Hard ionization techniques, like electron ionization (EI), are very energetic and thus cause strong analyte fragmentation. Soft ionization techniques, however, are less energetic and cause lower amount of analyte fragmentation, which is more suitable for biological samples. Depending on the stability and characteristics of the sample the appropriate ionization technique has to be chosen.

The most important ionization techniques are listed below [19]:

- Chemical Ionization (CI)
- Electron Ionization (EI)
- Field Ionization (FI)
- Fast Atom Bombardement (FAB)
- Liquid Secondary Ion Mass Spectrometry (LSIMS)
- Field Desorption (FD)
- Plasma Desorption (PD)
- Laser Desorption (LD)
- Matrix-assisted Laser Desorption Ionization (MALDI)
- Thermospray (TS)
- Secondary Ion Mass Spectrometry (SIMS)
- Electrospray Ionization (ESI)
- Atmospheric Pressure Ionization (API)
- Atmospheric Pressure Photoionization (APPI)
- Atmospheric Pressure Secondary Ion Mass Spectrometry (APSIMS)

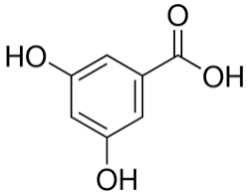
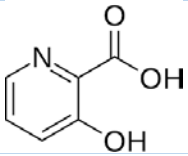
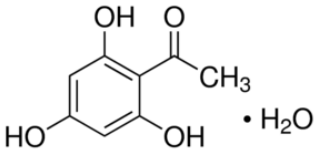
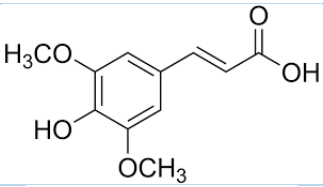
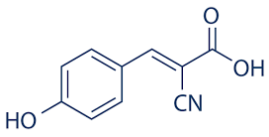
The ionization technique applied during the thesis was MALDI, therefore only this method will be described in detail.

2.7.2 MALDI

MALDI is defined as a soft ionization technique used to produce gas-phase ion, and is applied for the ionization of biological, organic and inorganic samples. The application of a matrix, which is usable for desorption and ionization purposes also in terms of the employed laser of the instrument, is very important. The sample preparation itself is simple and not very time-consuming. Even though MALDI has a high tolerance for salts, buffers and detergents an additional sample purification has to be performed for *in-gel* digested and extracted peptides (ZipTip purification).

At first, the sample preparation has to be done. Therefore, the analyte is dissolved and co-crystallized with a matrix that possesses strong light absorption qualities at the laser wavelength and thus can transfer energy to the analyte. Different matrices are listed in table 1.

Table 1: Common MALDI matrices.

Compound	**Structure**	**Wavelength**	**Major applications**
2,5-dihydroxy benzoic acid (DHB)		UV 337 nm, 353 nm	Proteins, peptides, adduct formation
3-hydroxy-picolinic acid (HPA)		UV 337 nm, 353 nm	Nucleic acids
K,m,n-di(tri)hydroxy-acetophenone (THAP)		UV 337 nm, 353 nm	Proteins, peptides, non-covalent complexes; near neutral pH
Sinapinic acid (SA)		UV 337 nm, 353 nm	Proteins, peptides
α-Cyano-4-hydroxycinnamic acid (CHCA)		UV 337 nm, 353 nm	Peptides, fragmentation

In the second step (figure 9) the laser ablates layers of the matrix under high vacuum conditions. The irradiation by the laser induces rapid heating of the matrix crystals, which leads to a localized

sublimation and release of analytes into the gas phase, where the ionization process occurs. Nevertheless, the exact mechanism of the MALDI ionization process is not entirely understood [7, 19, 20].

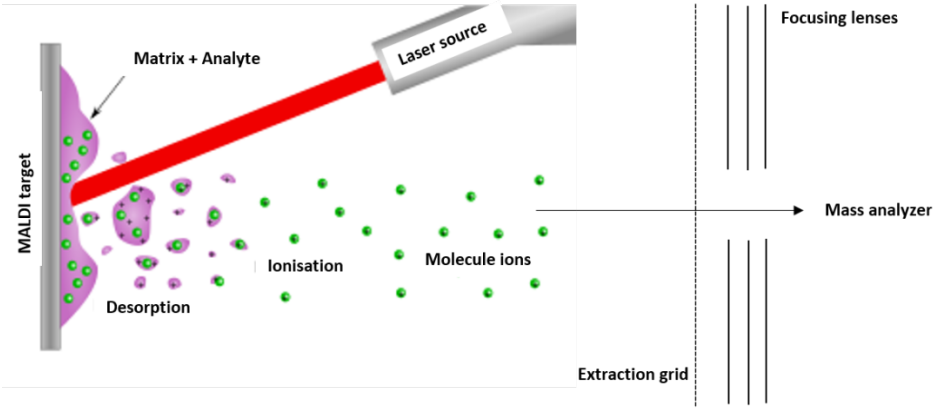


Figure 9: Schema of the MALDI process.

2.7.3 Sample preparation

The MALDI technique requires the mixing of the sample with the matrix for the absorption of the laser energy and its transfer to charge the analytes. Next to the choice of matrix also the sample preparation technique is important. For this thesis only the dried-droplet method was used.

A few sample preparation techniques are listed here [7]:

- Dried-droplet method: This method is the most common used, due to its simple application and the reproducibility it offers. The sample is directly mixed with the matrix-solution e.g. in a 1:1 ratio and spotted on the target plate. The droplet is left to dry at room temperature with or without a soft stream of neutral gas.
- Thin-layer method: At first the matrix solution is spotted onto the target plate and dried completely resulting in a very thin homogenous layer of matrix crystals. Subsequently, the samples solution is spotted on top of the matrix layer and left to dry. The difficulty hereby is the application of the sample drop. The spotted sample solution is not allowed to dissolve the whole matrix layer, which would lead to mixing of the layers, but just the top part of the matrix layer.
- Sandwich method: For this method a thin layer of matrix is spotted on the target and left dry. Then the sample solution is spotted onto the thin matrix layer and dried, just like in the thin-layer method. Afterwards, an additional layer of matrix is applied to the top of the sample layer and left to dry at room temperature. Hence, the sample layer is surrounded by matrix layers.

2.7.4 Mass analyzer

After the gas-phase ions have been generated and passed through the focusing lenses, they are separated in the mass analyzer according to their mass-to-charge ratios. This separation happens in the mass analyzer, which measures the m/z ratios of the ions due to their defined kinetic energy. To obtain this static or dynamic fields are used in the instrument. The mass analyzer of the instrument is permanently under high vacuum (10^{-3} – 10^{-9} mbar). Different mass analyzers are listed in table 2.

Table 2: Comparison of different mass analyzers [19].

Mass analyzer	Principle of separation	Accuracy
Quadrupole (Q)	m/z (trajectory stability)	100 ppm
Ion Trap (IT)	m/z (resonance frequency)	100 ppm
Time-of-flight (TOF)	Velocity (flight time)	200 ppm
Time-of-flight Reflectron (RTOF)	Velocity (flight time)	10 ppm
Magnetic sector (B)	Momentum	<10 ppm
Fourier transform ion cyclotron resonance (FTICR)	m/z (resonance frequency)	<5 ppm

Although each mass analyzer is important, only the ones used during the thesis, which are TOF and RTOF, are described in more detail.

2.7.4.1 Time-of-flight (TOF) Analyzer

A TOF analyzer is a flight tube, where the ions, which were generated in the ion source, drift in a high vacuum, field free zone. The acceleration of the ions occurs in the ion source itself, afterwards there is no additional acceleration. Therefore, every ion starts with the same amount of kinetic energy and the separation takes place due to their different m/z values. Smaller ions will have a higher velocity then larger ions and hence, the time they need to reach the detector differs. The exact m/z values are determined with the measurement of the exact amount of time needed from the source to the detector [19, 20].

A schema of a linear TOF is displayed in figure 10. The ions are generated in the ion source and are accelerated towards the mass analyzer with a constant acceleration U (20 keV). The electric potential energy (E_{el}) is transformed into kinetic energy (E_{kin}), which is displayed in equation 7.

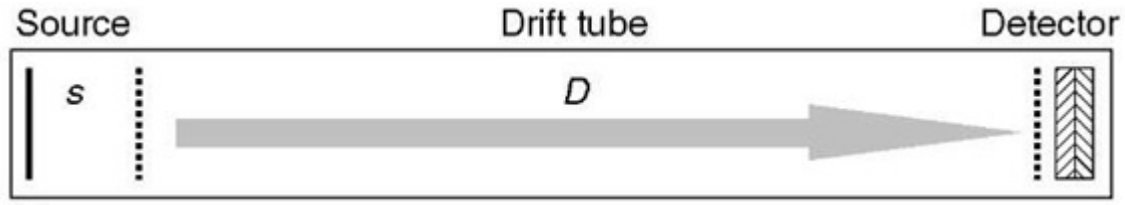


Figure 10: Schema of a linear TOF. [21]

$$E_{kin} = \frac{1}{2}mv^2 = zeU = qU = E_{el}$$

Equation 7: Kinetic energy of the ions in the TOF analyzer. m = mass of an ion, v = speed of the ion after the acceleration distance, z = charge state, e = elementary charge, U = acceleration voltage.[7]

The speed of the ion in the flight tube remains constant and therefore it can be calculated with equation 8.

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

Equation 8: Velocity of the ion in the TOF analyzer. v = speed of the ion after the acceleration distance, L = Length of the flight tube, t = time needed to travel the distance of the flight tube.[7]

Combining equation 7 and 8 results in equation 9 and subsequently equation 10, which is the calculation of the m/z values.

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

Equation 9: Velocity inserted into the kinetic energy and the electric potential energy. m = mass of an ion, L = Length of the flight tube, t = time needed to travel the distance of the flight tube, z = charge state, e = elementary charge, U = acceleration voltage.[7]

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

Equation 10: Calculation of m/z values. m = mass of an ion, L = Length of the flight tube, t = time needed to travel the distance of the flight tube, z = charge state, e = elementary charge, U = acceleration voltage.[7]

The mass resolution capacity R can be calculated from equation 11.

$$R = \frac{m}{\Delta m} = \frac{m_1}{m_2 - m_1}$$

Equation 11: Resolution capacity of a mass analyzer. m = mass of the ion; m_1 = mass of ion 1, m_2 = mass of ion 2, $m_1 < m_2$.[7]

The acceleration U, the elementary charge and the length of the flight tube are constant. Therefore, the m/z values can be accurately calculated, because they only depend on the exact time the specific ion needs to travel through the flight tube.

However, not all ions are desorbed and ionized at the same time, which results in differences in the energy, place and time and therefore leads to wrong m/z values or peak broadening and poor mass resolution [7, 19, 20].

To improve the poor mass resolution, two techniques have been developed: delayed extraction and the reflectron.

2.7.4.1.1 Delayed pulsed extraction

The delayed extraction is a method to reduce the kinetic energy differences among ions with the same m/z values leaving the source. A delay between the ion formation and the extraction is introduced.

At first the ions are allowed to expand in the field-free region in the source and after a certain delay a voltage pulse is applied to extract the ions. The ions of same m/z ratio with more kinetic energy move further towards the detector than the less energetic ones. The voltage pulse then transmits more energy to the ions, which remain longer in the flight tube. Therefore, the initial energy loss of the ions is compensated, and they join the energetically higher ions at the same time at the detector. The

method corrects the energy dispersion of the ions leaving the source and improves the mass resolution of the linear TOF analyzer. Unfortunately, this method also has a drawback, which is noticeable in the complication of the mass calibration procedure [22].

2.7.4.1.2 Time-of-flight Reflectron (RTOF)

Another wax to compensate the kinetic energy dispersion of ions with the same m/z ratio and thus to improve the mass resolution is the reflectron. A retarding field, which acts like an ion mirror, is positioned at the end of the flight tube and focusses the ions and reverses their flight direction back towards the detector, which is positioned next to the source. Figure 11 shows a schema of a reflectron TOF mass analyzer. The reflectron is a series of ring electrodes that create a constant electric field near their center and thus slows down the ions and redirects them [19].

The reflectron not only compensated the initial kinetic energy difference of ions with the same m/z ratio, but it also has a positive effect on the mass resolution.

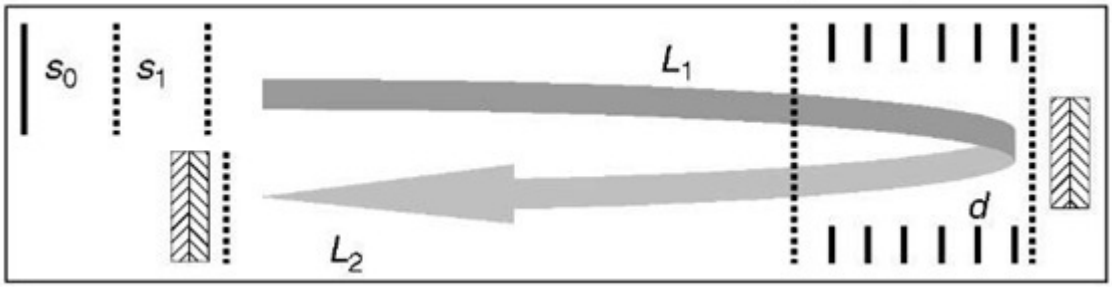


Figure 11: Schema of a reflectron TOF analyzer [21].

In addition to a single-stage reflectron a dual-stage reflectron can be installed, which further improves the performance of the instrument. The dual-stage reflectron consist of two reflectrons with two different electric field, an intense electric field that strongly decelerates the ions and a weaker field. Furthermore, curved field reflectrons have been designed, which act like ion mirror with non-linear electric field distributions and are primarily used in post source decay fragmentation.

Nowadays, research TOF instruments include both reflectrons and delayed extraction, which leads to higher mass resolution and the compensation of the initial kinetic energy distribution of ion with the same m/z ratio.

2.7.5 Detector

The detector is responsible for converting the incident ions into an electric signal, which is proportional to their amount. The most commonly used detector for MS is the secondary electron multiplier (SEM).

The secondary electron multiplier can be a discrete or a continuous dynode (Figures 12 - 13). When one electron hits the surface of the detector, the dynode generates up to ten free secondary electrons, which are accelerated by electric fields. These electrons hit the surface of the dynode in succession and a cascade effect is set into operation. At the end an amplification of the signal to 10^5 and 10^6 can be achieved.

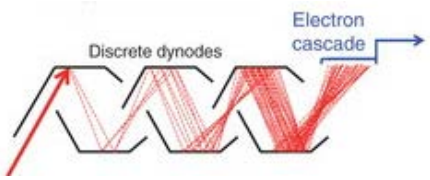


Figure 12: Discrete dynodes of a secondary electron multiplier [23].

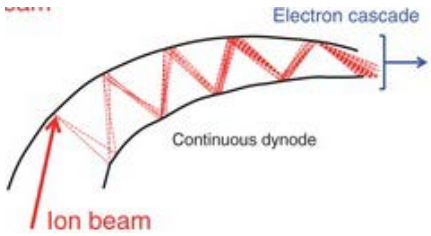


Figure 13: Continuous dynode of a secondary electron multiplier [23].

The microchannel plate (MCP) detector, also widely used in the field of mass spectrometry, is based on the same concept as the SEM (Figure 14). However, due to low noise amplification of low ionic current flows, it exceeds the SEM setup. MCPs are made of glass and contain several thousand very small, parallelly arranged channels to amplify the signal. An amplification up to 10^8 can be achieved with microchannel plate detectors, especially if multiple MCPs are used to detect the signal.

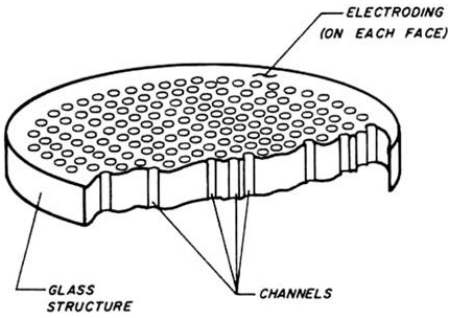


Figure 14: Microchannel plate[24].

2.7.6 Tandem mass spectrometry

Tandem mass spectrometry, also called MS/MS or MS^2, is based on the fragmentation of a single parent ion into fragment ions, which are then detected. Additionally, multiple successive mass spectrometric fragmentations can be executed (MS^2, MS^3, …, MS^n).

At first the parent ion has to be selected in the mass analyzer, forwarded to a collision cell and subsequently fragmented. The fragment ions are then detected, and a mass spectrum of the fragmentation pattern is constructed. There are two possibilities to select the parent ion: in-space or in-time. For the in-space selection at least two mass analyzers are needed, while for the in-time selection only one mass analyzer is needed.

Furthermore, there are three common methods to generate fragments:

- Collision-induced dissociation (CID): The parent ion is fragmented due to the energetic collision with a neutral gas like nitrogen or argon in the collision cell.

- In-source decay (ISD): It is generated by the desorption/ionization process itself. The fragmentation of the parent ion happens before the ion is accelerated into the mass analyzer.
- Post-source decay (PSD): The fragmentation of the parent ion occurs in the field-free region of the flight tube. The fragment ions possess the same velocity as the parent ion, however, the total kinetic energy is split between the fragments in accordance to their masses. This results in poor focusing of the fragment ions and thus poor mass resolution. A curved field analyzer as well as a so-called LIFT-cell, which lifts up the potential of the fragment ions inside a special section of the flight tube, however, improve the fragment mass resolution.

The selection of the single mass of an ion is done with an appropriate mass window. The remaining ions are deflected either with a pair of electrodes or a set of thin wires biased alternatively positively and negatively charged. This so-called ion gate is positioned in front of the collision cell. In addition, the laser power is increased twofold for TOF/TOF, which lead to a decrease in resolution due to wide kinetic energy distribution. The resolving power is at least one order lower than the best result for the parent ion. [19, 20]

The generation of a mass spectrum of fragments can be applied for the structural analysis of molecules, which is commonly done to obtain information about the sequence of a protein of interest in the proteomics field. The method used during this work was MALDI TOF/TOF with PSD generated fragment ions.

2.7.7 Protein sequencing

Tandem mass spectrometry is useful in the determination of the structural analysis of peptides. The peptides show a specific pattern during the fragmentation by MS^2 along the protein backbone. One of the two peptide fragments generated stabilizes the charge and is named according to the nomenclature of Roepstorff and Fohlman (figure 15). If the charge is on the N-terminus fragment the letters *a*, *b*, *c* are used, whereas on the C-terminus fragment the letters *x*, *y*, *z* are applied.

During PSD MS^2 mostly *b* and *y* ions are created, which simplifies the identification process of the database search. However, in low-energy CID MS^2 *a*, *b* and *y* fragments are favored and the presence of immonium ions in the low-mass region is observed. In high-energy CID MS cleaving of the amino acid side chains can occur, which generates *v*, *w* and *d* fragment ions [20, 25].

The peptides, which are generated with tryptic digestion, are often not enough to identify the protein via PMF search due to the high possibility of amino acid combinations that fit the represented m/z values. Therefore, MS^2 is applied for further protein information.

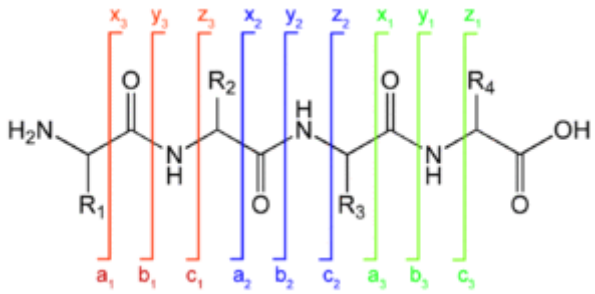


Figure 15: Nomenclature of peptide fragmentation with MS^n [25].

2.8 Peptide Mass Fingerprinting (PMF)

The concept of peptide mass mapping for protein identification by mass spectrometry and subsequent database searching was first introduced in 1993 [26, 27]. A sample with a single or few proteins can be digested by a sequence specific protease to peptides, which are searched in a database and the best match is offered. The strategy for the identification of a protein can be seen in figure 2.

At first, electrophoretic protein separation with polyacrylamide gels is performed. The separation can be done with 1D gels, which separate the proteins according to their size, or with 2D gels, which separate the protein by the isoelectric point and size. The gel band or spot is then excised, *in-gel* digestion with Trypsin/Lys C is performed, and the peptides are extracted and analyzed with MS and MS^2.

Trypsin is a serine protease that cuts the amino acid sequence of a protein after arginine or lysine, while Lys C cuts the protein sequence only after lysine. In this way peptides, which are easily protonated, are generated and can be measured in the positive ion mode due to their stability.

The digested peptides create a unique pattern for each protein and can thus be compared to in-silico generated digestions of a protein database and identified.

The obtained MS data is compared to a blank and a contaminants list to identify peaks derived from the background of the gel itself and its handling and contaminations [28]. The remaining peaks can afterwards be used for the database search. The identification of proteins is achievable if only one protein is present in the sample. Peptides from more than one protein in the mass spectra can lead to misinterpretation or failure of the protein identification.

MALDI MS is the preferred method for PMF due to the direct correlation between the mass of a peptide and the m/z ratios of single charged ions, the high mass accuracy, the tolerance towards contaminants, its simplicity in handling and the high speed and sensitivity of the measurement. [19, 20]

2.8.1 Mascot search

Mascot is the database search engine of Matrix Science (Boston, MA, USA) which was used for this work. The MS and MS^2 data are compared to in-silico digested proteins of the database with the help of optimized search algorithm and the obtained results are demonstrated with a scoring scheme. The scores are appointed according to the best fit of the protease cleavage specificity, the measured molecular mass and the fragment ion pattern [19].

Additionally, different parameters like fixed and variable modifications, used enzyme, missed cleavages, peptide molecular mass tolerance and peptide fragment mass tolerance can be adjusted for the search in the database. Identified proteins are given with a probability of 95 % (figure 16). The probability of a protein being the accurate one to the searched data within the green field is only 5 %, while a protein, which exceeds the green field, already has a probability of 95 %.

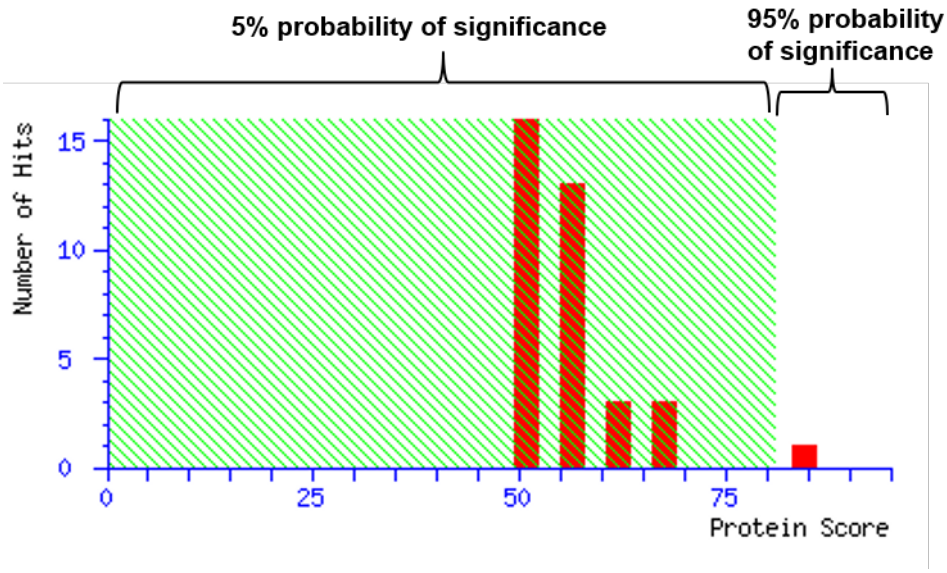


Figure 16: Example for a mascot search with one significant hit and four not significant hits.

3. Materials and Methods

The collection of young apple leaves and the extraction, isolation and purification of the proteins via FPLC as well as the continuous monitoring of the processes was performed by the working group of Prof. Halbwirth employing well-established protocols and was not part of my thesis. The isolation and extraction were performed by a diploma student, Lukas Eidenberger BSc., of Prof. Halbwirth's working group. Nevertheless, some basic aspects concerning these purification steps are given.

The further analysis of the isolated and purified protein(s) was/were accomplished through SDS-PAGE and peptide mass fingerprinting (PMF) in the Institute of Chemical Technologies and Analytics.

The chemicals, buffers and instrumentations used during the isolation and analysis processes by the two working groups are given below.

3.1 Chemicals

3.1.1 Workgroup Phytochemistry and Plant Biochemistry

Table 3: Chemical list used in the workgroup Phytochemistry and Plant Biochemistry.

Chemical	Quality	Company	Lot number
$(NH_4)_2SO_4$	for biochemical purposes	Merck, Darmstadt, DE	1.01211.1000
Citric acid	≥ 99.5 %	Merck, Darmstadt, DE	-
Ethylacetate	pro analysis	Merck, Darmstadt, DE	512 K301723
G6P	Sigma grade	Sigma-Aldrich, St.Louis, MO, USA	093K3789
G6PD	rec. expressed 958 units/mg Protein	Sigma-Aldrich, St.Louis, MO, USA	SLPOL2802V
Glycine	for molecular biology	Sigma-Aldrich, St.Louis, MO, USA	-
HCL conc.	pro analysis 37 %	Merck, Darmstadt, DE	K33464517 427
HEPES	≥ 99.5 % buffer grade	Sigma-Aldrich, St.Louis, MO, USA	458277272
KPi	crystallized Extra pure	Sigma-Aldrich, St.Louis, MO, USA	A217171017
MeOH	for analysis	VWR, Radnor, PA, USA	I680409 318
MES	≥ 99 % for biochemistry	Carl Roth., Karlsruhe, DE	287260817
MgCl2	pro analysis	Merck, Darmstadt, DE	1.05833
Mono S HR 5/5	-	Pharmacia, Uppsala, SE	-

Na acetate	≥ 99 % for molecular biology	Sigma-Aldrich, St.Louis, MO, USA	-
Na-ascorbate	laboratory use only	Sigma-Aldrich, St.Louis, MO, USA	49HO403
NaOH	pro analysis	Sigma-Aldrich, St.Louis, MO, USA	B478298414
p-coumaroyl CoA	≥ 90 %	TransMIT, Gießen, DE	-
PEG 4000	-	Sigma-Aldrich, St.Louis, MO, USA	0001429971
Quartz sand	-	Merck, Darmstadt, DE	TA2035836608
Sodium citrate	for analysis	Merck, Darmstadt, DE	A0158348024
SP Sepharose Fast Flow	-	Pharmacia, Uppsala, SE	-
Triton X-114	proteomics grade	VWR, Radnor, PA, USA	18B2856927

3.1.2 Workgroup Mass spectrometric Bio and Polymer Analysis

Table 4: Chemical list used in the workgroup Mass spectrometric Bio and Polymer Analysis.

Chemical	**Quality**	**Company**	**Lot number**
ACN	for analysis	Merck, Darmstadt, DE	K46689503 603
$AgNO_3$	for analysis	Merck, Darmstadt, DE	K43387312
CHCA	≥ 99.8 % (TLC), powder	Sigma-Aldrich, St.Louis, MO, USA	MKBF4800V
DTT	≥ 99.5 %	Sigma-Aldrich, St.Louis, MO, USA	BCBS8122V
EtOH	absolute, for analysis	Merck, Darmstadt, DE	K49722983 747
FA	98-100 % pro analysis	Merck, Darmstadt, DE	K37957664 744
Formaldehyd solution	36.5-38 % in H_2O	Sigma-Aldrich, St.Louis, MO, USA	SZBF1830V
HAc	≥ 99.8 %	Fluka, St. Gallen, CH	SZBG2450
IAA	≥ 99 %	Sigma-Aldrich, St.Louis, MO, USA	SLBT9667
MeOH	hypergrade LC-MS	Merck, Darmstadt, DE	I0908835 739
MOPS (20X)	research use only	ThermoFisher Scientific, Waltham, MA, USA	1771642
Na_2CO_3	≥ 99.8 %	Sigma-Aldrich, St.Louis, MO, USA	SZBD186AV
$Na_2HPO_4*2H_2O$	pro analysis	Merck, Darmstadt, DE	K37429480 724
$Na_2S_2O_3*5H_2O$	≥ 99.5 % pro analysis	Riedel-de Haën, Seelze, DE	11280

NH_4HCO_3	≥ 99.5 %	Sigma-Aldrich, St.Louis, MO, USA	BCBW3685
Novex Sharp pre-stained protein standard (marker)	research use only	ThermoFisher Scientific, Waltham, MA, USA	1644602
NuPAGE LDS Sample Buffer (4X)	research use only	ThermoFisher Scientific, Waltham, MA, USA	934122
Pierce 660 nm Protein Assay reagent	-	ThermoFisher Scientific, Waltham, MA, USA	OL194419
Potassium hexacyanoferrate (III)	99 %	Sigma-Aldrich, St.Louis, MO, USA	MKBH6304V
TFA	HPLC grade	Sigma-Aldrich, St.Louis, MO, USA	STBD8958V
Trypsin/LysC Mix	mass spec grade	Promega, Fitchburg, WI, USA	0000278420
PepMix 4	laboratory use only	LaserBioLabs (Sophia-Antipolis, FR)	PEP4-14001

3.2 Buffers and Working Solutions

3.2.1 Workgroup Phytochemistry and Plant Biochemistry

Table 5: Buffer list used in the workgroup Phytochemistry and Plant Biochemistry.

Buffer	**Composition**
Disruption buffer I Ionic strength: 0.341 M	125 mM KPi, 11 % Triton X-114 + 0.5 % Na-ascorbate
Disruption buffer II Ionic strength: 0.341 M	125 mM KPi, pH 7.5 + 0.5 % Na-ascorbate
HEPES buffer I Ionic strength: 0.015 M	40 mM HEPES, pH 7.4 + 0.2 % Na-ascorbate
HEPES buffer II Ionic strength: 1 M	40 mM HEPES, 1 M NaCl, pH 7.4 + 0.2 % Na-ascorbate
MES buffer I Ionic strength: 0.007 M	40 mM MES, pH 5.5 + 0.2 % Na-ascorbate
MES buffer II Ionic strength: 1 M	40 mM MES, 1 M NaCl, pH 5.5 + 0.2 % Na-ascorbate

3.2.2 Workgroup Mass spectrometric Bio and Polymer Analysis

Table 6: Buffer list used in the workgroup Mass spectrometric Bio and Polymer Analysis.

Buffer	Composition
Alkylation solution	54 mM IAA in 100 mM NH_4HCO_3 10 mg IAA/mL 100 mM NH_4HCO_3 Stored in the dark
Conservation Solution	1 % HAC, 99 % UHQ water
Development Solution	3 g Na_2CO_3, 150 mL UHQ water, 80 µL Formaldehyde solution (added directly after incubation step) Prepared on the same day as staining procedure
Digestion buffer	95 % 50 mM NH_4HCO_3, 5 % ACN
Fixing Solution	50 % EtOH, 5 % HAc, 45 % UHQ water
Incubation Solution	0.05 g $AgNO_3$, 50 mL UHQ water Prepared on the same day as staining procedure; stored at 4 °C
Matrix solution	3 mg CHCA/mL ACN/UHQ water/TFA (60:40:0.1)
MOPS buffer (1X)	40 mL MOPS buffer (20X), 760 mL UHQ water
Na-phosphate buffer	20 mM, pH 7.3
Reducing solution	10 mM DTT in 100 mM NH_4HCO_3 1.54 mg DTT/mL 100 mM NH_4HCO_3 Stored in the dark
Sensitivation Solution	0.01 g $Na_2S_2O_3$*$5H_2O$, 50 mL UHQ water Prepared on the same day as staining procedure
Stopping Solution	5 % HAc, 95 % UHQ water
Trypsin/LysC Mix solution	10.0 ng trypsin/µL
Washing Solution I	50 % EtOH, 50 % UHQ water
Washing Solution II	UHQ water

3.3 Instrumentation

3.3.1 Workgroup Phytochemistry and Plant Biochemistry

- Allegra 21R centrifuge by Beckman Coulter, Brea, CA, USA
- Analytical Balance Research R 160 P by Sartorius, Göttingen, DE
- Avanti J-26 XP high-speed centrifuge by Beckman Coulter, Brea, CA, USA
- Balance PJ 3000 by Mettler, Columbus, OH, USA
- FPLC ÄKTApurifier by GE, Boston, MA, USA
- Fractionator Frac 950 by GE, Boston, MA, USA
- Household blender AR100 by Moulinex, Ecully, FR
- HPLC UltiMate 3000 by ThermoFisher Scientific, Waltham, MA, USA
- Magnetic stirrer Ikamag RET by IKA, Staufen im Breisgau, DE
- Magnetic stirrer MR Hei Standard by Heidolph, Schwabach, DE

- pH meter InoLab Labor pH-Meter Level 2 by WTW, London, UK
- Sample pump ÄKTAprime by GE, Boston, MA, USA
- Sigma 1-14 table centrifuge by Sigma, Kawasaki, JP
- Thermomixer Comfort by Eppendorf, Hamburg, DE
- Vacuum pump Laboport by KNF Neuberger, Trenton, NJ, USA
- Vortex Genie by Scientific Industries, Bohemia, NY, USA
- Vivaspin 20 by Sartorius, Göttingen, DE

3.3.2 Workgroup Mass spectrometric Bio and Polymer Analysis

- Analog Vortex Mixer by VWR, Radnor, PA, USA
- Analytical balance Extend by Sartorius, Göttingen, DE
- Cooling Centrifuge Sigma 3-30K by Sigma Centrifuges, Osterode, DE
- Galaxy Mini Centrifuge Mini Star silverline by VWR, Radnor, PA, USA
- Gel chamber by Invitrogen, Waltham, MA, USA
- Invitrogen Power Supply by Invitrogen, Waltham, MA, USA
- Millipore Simplicity UV by Merck, Darmstadt, DE
- Thermomixer Comfort by Eppendorf, Eppendorf, Hamburg, DE
- Unicryo MC2L -60 °C by UniEquip, Munich, DE
- Univapo 100 H by UniEquip, Munich, DE

For all analysis work Ultra-High Quality water (UHQ) from the Millipore Simplicity UV was used. The resistivity parameter was 18.2 mΩcm at 25 °C.

3.4 Disposable Materials, work group Mass spectrometric Bio and Polymer Analysis

- Centrifuge Tubes Freestanding, 50 mL by VWR, Radnor, PA, USA
- Centrifuge Tubes with Screw Caps, 15 mL by VWR, Radnor, PA, USA
- epT.I.P.S. Reloads 0.1-10 µL Eppendorf Tips by Eppendorf, Hamburg, DE
- epT.I.P.S. Reloads 2-200 µL Eppendorf Tips by Eppendorf, Hamburg, DE
- epT.I.P.S. Reloads 50-1000 µL Eppendorf Tips by Eppendorf, Hamburg, DE
- MWCO 10 kDa centrifugal filters, low protein binding by VWR, Radnor, PA, USA
- NuPAGE 4-12 % Bis-Tris Protein Gels by ThermoFisher Scientific, Waltham, MA, USA
- Protein Chips and Electrode Cleaner for the Agilent 2100 Bioanalyzer System by Agilent Technologies, Santa Clara, CA, USA
- Protein LoBind Tubes 0.5 mL, PCR clean by Eppendorf, Hamburg, DE
- Safe-Lock Tubes 0.5 mL by Eppendorf, Hamburg, DE
- Safe-Lock Tubes 1.5 mL by Eppendorf, Hamburg, DE
- UV-Cuvette micro, disposable cuvettes, 12.5 × 12.5 × 45 mm by Brand GmbH + Co. KG, Wertheim, DE
- ZipTip Pipette Tips C18 for Sample Preparation by Merck, Darmstadt, DE

3.5 Isolation and extraction of proteins

3.5.1 Collection of plant material

Young leaves from *M. domestica* cv. (cultivar) Golden Delicious were gathered at the experimental orchard of the University of Natural Resources and Life Sciences in Jedlersdorf (Vienna, Austria). The focus of the collecting of sample material was on young terminal leaves, with a length of 2 - 4 cm, light green in color and no signs of pathogenic alterations, infection, and insect or spider infestations. Leaves that showed those signs of alterations were not harvested.

The collection took place during May and June 2017 due to the fact that it was the growing season of the trees and therefore a higher quantity of young leaves is available, but shortly before their blooming period. To allow regeneration of the trees as well as new growth of young leaves the material was only collected once a week.

Immediately after the harvest, the samples were frozen with liquid nitrogen and stored at -80 °C.

3.5.2 Leaf and Cell Disruption

All equipment, like spatulas, containers, falcons and beakers, were cooled with liquid nitrogen before application on the leaves and leaf powder.

For the isolation and extraction of the proteins, the leaf material had to be homogenized in a first step. Therefore, 600 – 700 g of leaves were shredded with the help of an electric household blender. To prevent protein denaturation due to high temperatures inside the mixing chamber, which contains the blade, the chamber as well as the cap of the blender were carefully pre-cooled with liquid nitrogen, before every mixing step. After the blender was cooled down, leaves were put into the chamber and mixed for 15 - 20 s, while the blender was also shaken, to improve the efficiency of the disintegration process. The time of 15 - 20 s was chosen to avoid heat generation of the blender with longer mixing periods. Subsequently, the resulting, homogenized leaf powder was stored at -80 °C.

For cell disruption leaf powder was mechanically grinded with quartz sand under liquid nitrogen. In order to do so, leaf powder was first mixed with ice cold Disruption buffer I in a porcelain mortar, which was cooled with liquid nitrogen. The amount of buffer was chosen in a way, that despite of the higher viscosity of the leaf powder/buffer mixture grinding was still possible. In addition, an abundant amount of quartz sand was added to the mixture and it was ground in the mortar for 5 min. The exact time for grinding in the mortar and its efficiency depends on the leaf powder/Disruption buffer I/quartz sand ratio, which differs for each cycle, and was determined visually by color and viscosity change of the mixture. The obtained slurry was stored on ice until the next extraction step.

3.5.3 Aqueous Two - Phase Extraction with Triton X - 114

The following steps for the purification of the protein of interest were performed on ice: The obtained slurry was stirred thoroughly, filled into 250 mL centrifugation vessels and centrifuged at 3.0×10^4 g, 4 °C for at least 30 min. This resulted in a separation of the slurry into three layers (figure 17):

1. Bottom layer: Solids of the slurry (quartz sand and cell debris)

2. Higher density liquid layer: Liquid aqueous phase containing Triton X – 114 (chlorophyll-containing)

3. Lower density liquid layer: Liquid aqueous phase containing little amount of Triton X – 114 (soluble proteins)

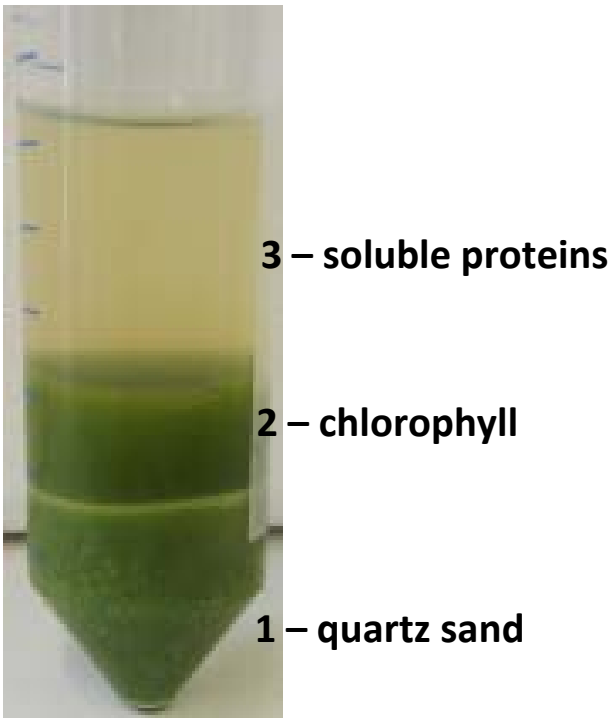


Figure 17: Aqueous two-phase extraction with Triton X-114 of the leave slurry. The separated layers are displayed: 1) bottom layer, 2) higher density liquid layer, 3) lower density liquid layer. The picture was provided by BSc. Lukas Eidenberger.

The liquid phases (combination of layers 2 and 3) were decanted to separate them from the bottom layer and hence all solid particles from the cell disruption process. Additionally, the decanted liquid layers were stirred again and centrifuged at 3.0×10^4 g, 4 °C for 30 min to achieve a complete separation of all solid material. The lower and higher density liquid layers were then separated from each other and stored on ice or at 4 °C.

Hereafter, the higher density liquid layer was extracted with ice cold Disruption buffer II, mixed until homogeneity and centrifuged at 3.0×10^4 g, 4 °C for at least 30 min. The lower density liquid layer was decanted and added to the previously separated and stored liquid. The higher density liquid layer was extracted for a second time with Disruption buffer II. The resulting lower density liquid layer was again pooled with the original one and the higher density layer was then discarded.

For a better comprehension of the procedure an extraction schema is shown in figure 18.

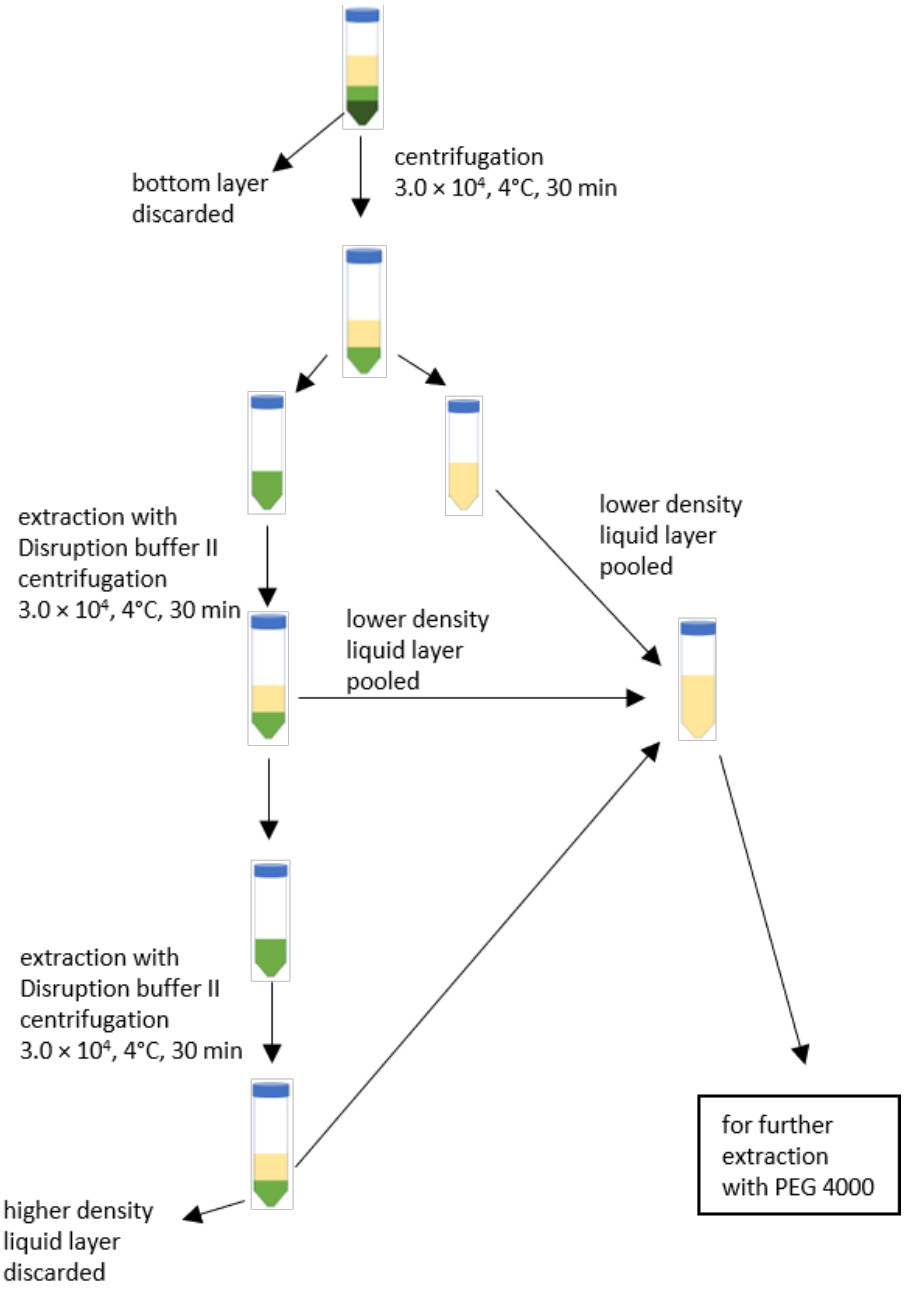


Figure 18: Aqueous Two-Phase Extraction schema with Triton X-114 of the protein slurry.

The pooled lower density liquid phase precipitated with solid ammonium sulfate until a final concentration of 30 % (w/v) ammonium sulfate was reached. Subsequently, the resulting suspension was centrifuged at 3.0 × 10^4 g, 4 °C for at least 20 min to separate the different layers. The liquid phase was decanted, and solid ammonium sulfate was added to it to achieve a concentration of 80 % (w/v). The solution was again centrifuged at 3.0 × 10^4 g, 4 °C for at least 20 min. The obtained pellet from 30 % (w/v) concentration was stored at -20 °C for backup, whereas the pellet from 80 % (w/v) concentration was used for further extraction of proteins.

3.5.4 Aqueous Two - Phase Extraction with PEG 4000

The protein pellet of the 80 % (w/v) ammonium sulfate concentration was solubilized in 70 mL of fresh, ice cold Disruption buffer II. The suspension was stirred until no more protein solubilized in the solution (typically 20 min). Subsequently, 10 % (w/v) PEG 4000 was added and dissolved by stirring with a magnetic stirrer. Subsequently, ammonium sulfate was added until a 30 % (w/v) concentration was achieved. The order of additive addition is crucial as else PEG 4000 and ammonium sulfate do not dissolve completely and PEG 4000 is carried over into the next extraction step. The dissolved protein solution was centrifuged at 5.4×10^3 g, 4 °C for at least 20 min, until the bulk liquid is clear.

This resulted in four different layers (Figure 19):

1. Bottom precipitate of 30 % ammonium sulfate
2. Bulk liquid
3. Second precipitate (floating)
4. Low density liquid phase (mostly PEG 4000)

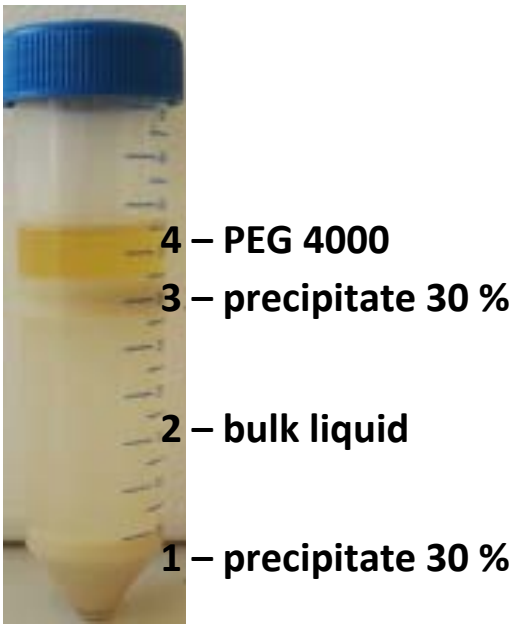


Figure 19: Aqueous two-phase extraction with PEG 4000. The four different layers are clearly distinguishable: 1) bottom precipitate, 2) bulk liquid, 3) second precipitate, 4) low density liquid phase. The picture was provided by BSc. Lukas Eidenberger.

The top liquid layer was pipetted off and discarded. The floating pellet was perforated on both sides and held back carefully with a spatula, whereby the bulk liquid (layer 2) could be decanted off. The pellets were combined and suspended and solubilized in Disruption buffer II.

The bulk solution had an ammonium sulfate concentration of roughly 30 % (w/v), which meant that proteins that precipitate between 30 - 80 % (w/v) were still in solution. Hence, ammonium sulfate was added until a concentration of 80 % (w/v) was achieved and the remaining soluble proteins precipitated. The solution was centrifuged at 5.4×10^3 g, 4 °C for 20 min. The different layers occurred anew, and the resulting pellets were obtained and treated as above. The resulting protein suspension was added to the previously obtained protein suspension from the first round of extraction. This

protein suspension was used for further PEG 4000 – ammonium sulfate extraction steps. Altogether 3-5 cycles of PEG 4000 - ammonium sulfate extraction were performed until the topmost liquid layer's color faded from a strong to a faint yellow. For a better comprehension of the procedure an extraction schema is shown in figure 20.

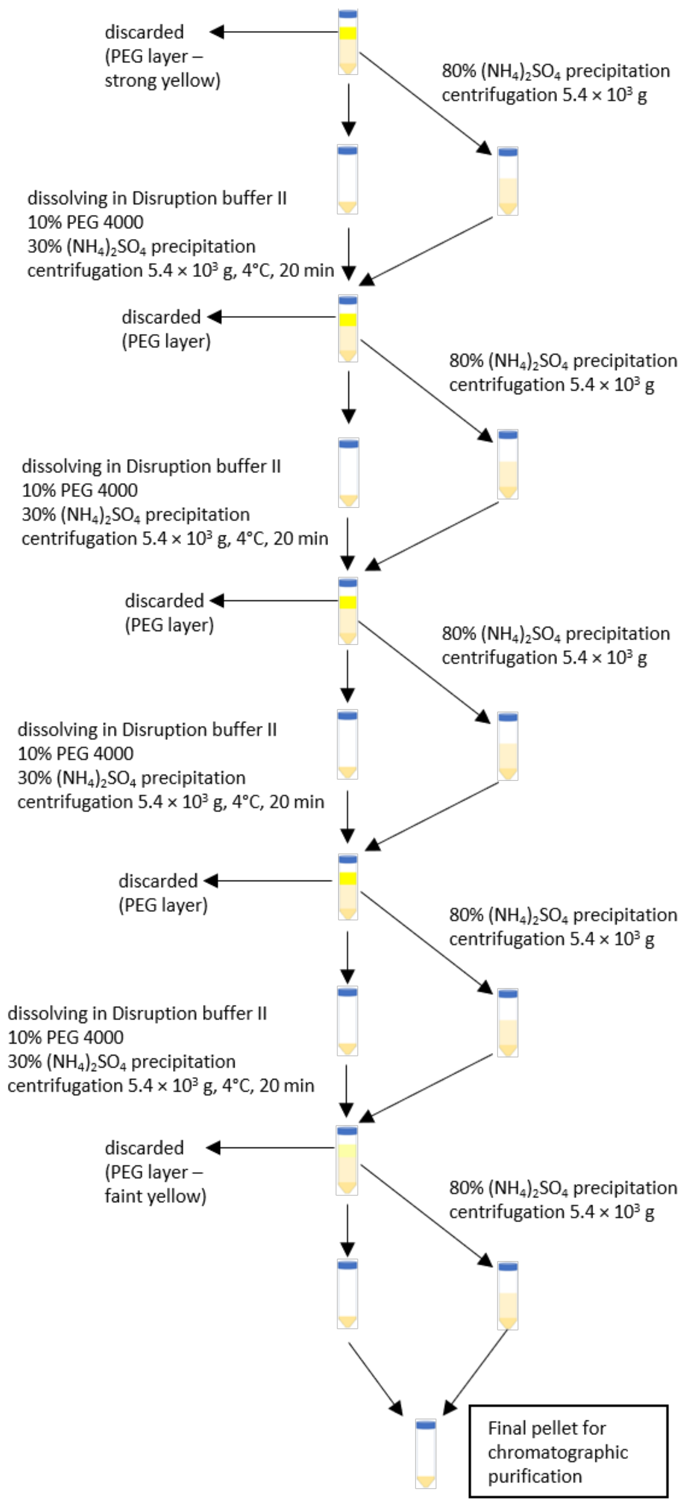


Figure 20: Aqueous Two-Phase Extraction schema with PEG 4000 of the protein solution of the Triton X-114 extraction.

If the pellet had a lower density than the bulk liquid and was therefore floating on top of it, the falcon was punctured at the bottom with a heated needle and the bulk liquid was sucked out with a syringe. The hole in the falcon was then sealed with parafilm for the resuspension of the pellet itself. Subsequently the suspension was transferred into a new falcon; this was the final protein solution.

3.5.5 Buffer Exchange to Low Salt Buffer

The high ionic strength (0.341 M) of the Disruption buffer II is not suitable for application on an ion exchange chromatography column; therefore, the proteins had to be transferred into a buffer system of lower ionic strength.

Thus, the final protein solution was precipitated with an ammonium sulfate concentration of 80 % and the suspension was centrifuged at 5.4 × 10^3 g, 4 °C for 20 min. If the separated liquid was not clear, additional centrifugation for a longer time period was carried out. The obtained pellet was resuspended in Disruption buffer II and solubilized while stirring. Then the suspension was centrifuged at 5.4 × 10^3 g, 4 °C for 30 min and the resulting solution was decanted and the soluble proteins concentrated in a Vivaspin 20 centrifugal concentrator. The concentrator was filled up to the 20 mL mark and centrifuged at 5.0 × 10^3 g, 4 °C for 20 min. This procedure was performed until all the solution was centrifuged and concentrated up to 1 mL of volume, which showed a dark brown color. The concentrated protein solution was pipetted into 20 - 50mL of low salt buffer. The membrane of the concentrator was washed with a few mL of low salt buffer and pooled with the previously obtained solution. Afterwards, the low salt protein solution was centrifuged at 5.4 × 10^3 g, 4 °C for 30 min to separate the remaining solid particles from the liquid phase prior the chromatography steps. The final centrifugation step is necessary due to possible protein precipitation occurring during the exchange of the buffer.

3.5.6 FPLC - Ion Exchange Chromatography

The purification of the protein of interest was done via Fast Protein Liquid Chromatography (FPLC) in three steps. Therefore, two different matrices for cation exchange chromatography were used.

For the first step a SP Sepharose Fast Flow (GE Healthcare, Chicago, IL, USA) packed column was employed and for the further two steps a Mono S HR 5/5 packed column was used. Column characteristics are shown in table 7.

Table 7: Column characteristics of used column packing material.

Property	**SP Sepharose Fast Flow**	**Mono S HR 5/5**
Chromatography mode	Strong cation exchange chromatography	Strong cation exchange chromatography
Ligand	Sulphopropyl	Methyl sulfonate
Ionic capacity	0.18 – 0.25 mmol H^+/mL medium	0.12 – 0.15 mmol H^+/mL medium
Matrix	Cross-linked agarose, 6 %, spherical	Polystyrene/Divinyl benzene
Average particle size (d_{50v})	90 µm	10 µm
Flow velocity (at 1 bar, 25 °C)	400 – 700 cm/h	150 – 900 cm/h

Working temperature	4 – 40 °C	4 – 40 °C
pH stability (CIP)	3 – 14	1 - 14
pH stability (working range)	4 - 13	2 - 12
Chemical stability	Aqueous buffers, 1 M NaOH, 8 M urea, 6 M guanidine hydrochloride, 70 % EtOH	Aqueous buffer, 2 M NaOH, 8 M urea, 6 M guanidine hydrochloride, 100 % EtOH
Storage	20 % EtOH, 0.2 M sodium acetate	20 % EtOH, 0.2 M sodium acetate
Not resistant to	Oxidizing agents, cationic detergents, prolonged exposure to pH < 4	Oxidizing agents, cationic detergents

Prior to application, the column was cleaned with 1M NaOH to remove remains from previous purifications. Then, an equilibration step with a low salt buffer followed. (The same buffer was used for the rebuffering of the protein concentrate.) The signals (Conductivity, UV 280 nm, UV 410 nm, UV 450 nm) were observed for a while and after they stabilized the protein solution applied to the column. The bound proteins were washed with a low salt buffer and eluted with an appropriate elution buffer of higher ionic strength. The elution was done according to a predetermined elution gradient profile. After the purification step, the column was cleaned again with NaOH.

For analyte detection, UV/Vis measurement cells at three specific wavelengths were used. The wavelengths were 280 nm for the detection of aromatic amino acids, 410 nm for heme group containing proteins and 450 nm for cytochrome P450 enzymes. 410 nm and 450 nm were chosen due to the protein of interest most likely being a protein containing a heme group and the importance of cytochrome enzyme class.

During the purification steps the pH, conductivity and pressure of the Äkta purifier system was constantly measured.

The conductivity of the protein solution, rebuffered in the low salt buffer, was measured and if necessary diluted with a 50% low salt buffer to achieve an optimal conductivity of 5 mS/cm. The chromatography itself was performed with a conductivity of the protein solution between 5 and 8 mS/cm.

The eluted fractions were tested via an enzyme assay for the biological activity of putative *p*-coumaroyl-CoA reductase. Due to the large number of eluted fractions, 2-3 fractions were pooled together and tested at once. The fractions, which displayed sufficient biological activity were pooled and carried into the next FPLC purification round. In order to do so, they were concentrated in a Vivaspin filter (200 mL to 3 - 5 mL) before performing the following chromatography. This should prevent a high amount of other proteins interfering with subsequent chromatographic purifications.

Enzyme assay:

At first a master mix (745 µL) for the enzyme assay was prepared according to the compounds and their volumes given in table 8.

Table 8: Compounds for creating 745 µL of master mix for the enzyme assay.

	$NADP^+$ (20 mM)	**G6P (200 mM)**	**$MgCl_2$ (1 M)**	**Sodium citrate (1 M, pH 6.0)**	**H_2O dest.**
Volume [µL]	15	250	25	250	205

For the enzyme assay 14.9 µL of master mix were put into an Eppendorf tube, together with 0.1 µL of G6PD, 5 µL of 2 mM *p*-coumaroyl CoA and 30 µL of the obtained fraction solutions. The compounds were all mixed together and incubated at 30 °C for 1 h. Afterwards, 20 µL of 5 M NaOH were added and vortexed to stop the enzymatic reaction. The test samples were then incubated at 45 °C for 45 min to hydrolyze dihydro-*p*-coumaroyl-CoA and *p*-coumaroyl-CoA. Subsequently, 25 µL of HCl conc. was added, mixed and the products and educts were extracted twice with ethyl acetate, 80 µL and 50 µL respectively. Therefore, ethyl acetate was added to the mixture, thoroughly mixed and centrifuged at 2.2×10^4 g for 3 min. The top phase (ethyl acetate phase) was taken, pooled in a fresh Eppendorf tube and then placed into a vacuum desiccator to evaporate the solvent. The dried proteins were dissolved in 30 µL 50% MeOH and analyzed via HPLC. The characteristics of the used HPLC column are listed in table 9.

For the separation two solutions were employed:

Buffer A: H_2O dest. with 0.1 % FA

Buffer B: ACN with 0.1 % FA

Table 9: Characteristics of the HPLC column used to analyze the enzyme assay.

Characteristic	
Column	Acclaim™ 120 C18 Column
Length [mm]	150
Diameter [mm]	2.1
Particle size [µm]	2.2
Packing material	Spherical silica, 120 Å pore size

To gain information about the exact elution time a standard mix, which consisted of *p*-coumaric acid and dihydro-*p*-coumaric acid, was measured before every enzyme assay measurement cycle.

Three FPLC chromatographic steps were performed:

1. Chromatographic step

- Ammonium sulfate pellet was dissolved in Buffer A
- Conductivity of protein solution was 5.6 mS/cm (1.2 L of solution)
- SP-Sepharose fast flow column
- Column XK 16/20
- Column volume: 24 mL
- Flow rate: 10 mL/min
- Buffer A: 30 mM NaOAc, 30 mM Mes, pH 5.5

- Buffer B: 30 mM NaOAC, 30 mM Mes, 1 M NaCl, pH 5.5
- Gradient: linear until 70% B in 18 CV

2. Chromatographic step

- Biological active fractions pooled, concentrated to 1 mL and diluted to 40 mL with Buffer A
- Mono S HR 5/5 column
- Column volume: 0.982 mL
- Flow rate: 1 mL/min
- Buffer A: 30 mM NaOAc, 30 mM MES, pH 5.5
- Buffer B: 30 mM NaOAc, 30 mM MES, 1 M NaCl, pH 5.5
- Gradient: linear until 35 % B in 90 CV

3. Chromatographic step

- Biological active fractions pooled and diluted to 40 mL with 30 mM NaOAc, 30 mM MES, pH 5.5
- Mono S HR 5/5 column
- Column volume: 0.982 mL
- Flow rate: 1 mL/min
- Buffer A: 40 mM HEPES, pH 7.2
- Buffer B (Elution Buffer): 40 mM HEPES, 500 mM NaCl, pH 7.2
- Gradient: linear until 35 % B in 135 CV
- Elution with pH 5.5

The obtained fractions (1 mL per well) were transferred from the microtiter sample plates into lo-bind protein Eppendorf tubes, each of them washed additionally with 20 µL of Na-phosphate buffer and stored at -80 °C.

3.6 Characterization of proteins

3.6.1 Concentrating and buffer exchange

The biological active fractions (H3 - H11) were thawed and pooled together (end volume of 9 mL) for further analysis. Thus, the sample had to be concentrated and the buffer had to be exchanged. The HEPES buffer, which was used for the final elution of the proteins, with its high ionic strength was exchanged against a Na-phosphate buffer.

The total amount of the pooled biological active fractions turned out to be 9 mL. Half of the pooled sample (4.5 mL) was taken for the concentrating and buffer exchange. Next to the biological active fractions three non-biological active fractions were chosen to be concentrated and rebuffered as a control and comparison.

500 µL of the Pool and B8, C3 and D1 were each pipetted into a 10 kDa MWCO cut off filter (VWR, Radnor, Pennsylvania, United States) and centrifuged at 9.0×10^3 g, 4 °C for 2 min. The process was repeated until all 4.5 mL of the pool were applied to the cut off filter and centrifuged. For the samples

B8, C3 and D1 only 1 mL, which was the total volume after the chromatographic cleanup, was used. Afterwards, 500 µL of the Na-phosphate buffer was pipetted on the filter and centrifuged under the same conditions. A small amount of liquid should remain on the filter after each centrifugation step, respectively. The spin filters were then each placed on a scale and buffer was added until a weight of 0.1 g was reached. Based on weighed volumes the concentration of proteins was roughly increased by a factor of 10 given that no unspecific attachment of proteins to the filter occurred.
The concentrated samples as well as the eluates were stored in 0.5 mL Eppendorf tubes at -80°C.

3.6.2 Protein concentration

The total protein concentration of the original Pool sample and the concentrated Pool sample were determined with the Pierce Assay 660 nm according to the protocol of the manufacturer [29]. The measurements were performed at 660 nm with a path length of 10 mm. Due to the little amount of protein solution a minimal assay was executed with 75 µL of Pierce reagent and 5 µL of sample. After a reaction time of 5 min the samples were measured.

First a calibration was performed using 9 standards, which were diluted from one BSA ampule (2 mg/mL) according to the scheme in table 10. For the reference a Blank with pure UHQ was used.

Table 10: Schema of the preparation of protein standards from one BSA ampule with a stock concentration of 2 mg/ml.

Vial	Volume of Diluent [µL]	Volume of BSA source [µL]	Final BSA concentration [µg/mL]
A	0	300 (of Stock)	2000
B	125	300 (of Stock)	1500
C	325	300 (of Stock)	1000
D	175	175 (of B)	750
E	325	325 (of C)	500
F	325	325 (of E)	250
G	325	325 (of F)	125
H	400	100 (of G)	25
I	400	0	0

Afterwards, triplicates of the original Pool and the concentrated Pool were measured. The results of the measurements are displayed in the Results and Discussion section (see chapter 4.2).

3.6.3 Bioanalyzer

The Agilent Protein 230 Kit and the Protein 230 chips were used for the Agilent 2100 Bioanalyzer. The working solutions were prepared according to the Agilent Protein 230 Kit Quick Start Guide[30]. Sample preparation and analysis were carried out according to standard protocols of the manufacturer.

For the gel-dye mix 25 µL of protein 230 dye concentrate were added to one protein 230 gel matrix aliquot and vortexed and spun down for 15 s. Afterwards, the solution was transferred to a spin filter

and centrifuged at 2.5 × 10^3 g for 15 min. For the destaining solution 650 µL of gel matrix were pipetted into a spin filter and centrifuged at 2500 x g for 15 min. For the denaturing solution 7 µL of 1 M DTT solution was added to 200 µL of sample buffer and vortexed. The obtained solutions were labelled and stored at 4°C.

An aliquot of the ladder, which was already prepared and stored at -20 °C, was thawed for each day of analysis.

The samples B8, C3, D1 and original Pool as well as the desalted versions of these samples were used for the analysis and to evaluate the concentrating step with the 10 kDa cut off spin filters.

Four µL of the protein samples were mixed with 2 µL of denaturing solution in 0.5 mL Eppendorf tubes. The samples as well as the protein 230 ladder aliquot were incubated at 95 °C for 5 min. Afterwards, the samples were cooled down, spun down shortly for 15 s and 84 µL of UHQ water were added in each tube and vortexed. A total volume of the samples and the ladder of 90 µL was obtained.

Subsequently, the loading of the chip was performed, according to the schema in figure 21. Therefore, the chip was placed on the priming station, the plunger was set at 1 mL and the marked gel well was filled with 12 µL of gel-dye mix. The priming station was closed, the plunger pressed until it was held by the clip and held in this position for 60 s. Thereafter, the clip was released and kept like this for 30 s before the priming station was opened and the chip was taken out. All marked gel wells were then filled with 12 µL of gel-dye mix each and 12 µL of destaining solution was also pipetted into the corresponding well. 6 µL of ladder was added into the appropriate well, and the sample wells were each filled with 6 µL of its designated sample. All the solutions were pipetted into the wells avoiding air bubble formation on the well bottom. The chip was placed into the chip shaker and briefly vortexed.

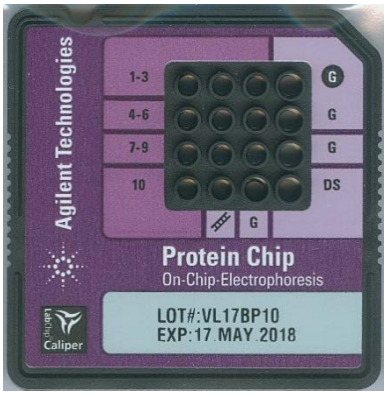


*Figure 21: Bioanalyzer 2100 Agilent Technologies 230 Protein Chip. The wells were filled as following: DS: Destain solution, Ladder symbol: Protein ladder, **G**: Gel-dye mix, G: Gel-dye mix, 1-10: protein samples. At the beginning, **G** is filled with gel-dye mix and pressed through the chip capillaries upon application of the priming station and plunger syringe.*

Before each measurement, a washing chip, filled with 380 µL of UHQ water, was placed into the Bioanalyzer for 1 min to clean its electrodes. Afterwards, the washing chip was taken out and the electrodes were air dried for 1 min.

The 2100 expert software was started and the method P230, which is the corresponding method to the protein 230 chip, was selected. The protein chip was then placed into the Bioanalyzer instrument and the measurement took place. Each sample was measured for 40 s.

3.6.4 SDS - PAGE

To obtain enough material for MALDI analysis, protein samples were prepared 8 times for separation via 1D gel electrophoresis. For the 2 remaining wells a protein standard marker was used. For the 1D gel electrophoresis NuPAGE Bis-Tris Mini Gels were used. The specifications of the gels are displayed in table 11.

Table 11: Specifications of NuPAGE Bis-Tris Mini Gel.

Specification	
Gel Percentage	4 - 12 %
Wells	10
Gel Type	Bis-Tris, cross-linked polymer
Separation Range	15 – 260 kDa (MOPS buffer) 3.5 – 160 kDa (MES buffer)
Gel Size	8 cm × 8 cm
Thickness	1.0 mm
Mode of separation	Molecular weight
Storage	4 – 25 °C

The sample preparation was performed according to the NuPAGE Bis-Tris Mini Gel Electrophoresis Protocol, which is displayed in table 12. The samples were vortexed briefly and incubated at 95 °C for 5 min for protein denaturation. Afterwards, the samples were cooled down to room temperature and briefly centrifuged.

5 µL of the sample equals 0.690 µg of total protein concentration, which was determined with a Pierce Assay (see Chapter 4.2).

Table 12: NuPAGE Bis-Tris Mini Gel Electrophoresis Protocol with the calculated and used amounts of solutions and protein sample for the loading samples.

Components	**Reduced Sample**
Sample	5 µL
NuPAGE LDS Sample Buffer (4X)	2.5 µL
NuPAGE Reducing Agent (10X)	1 µL
Deionized water	1.5 µL
Total volume	10 µL

At first, the gel package was opened, and the gel rinsed with deionized water. The comb was then removed carefully, and the gel wells were washed trice with the running buffer. Additionally, the white tape near the bottom of the gel cassette was also removed and that space was washed trice using the running buffer.

The prepared and washed gel was immediately placed into the gel running chamber, which was subsequently filled up to 3/4 with running buffer.

The whole volume (10 µL) of the prepared samples were loaded onto the gel, as well as two protein standard markers (5 µL each) at both sides of the gel.

After the loading of the gel, the gel chamber was closed, hooked up to the power supply and the gel electrophoresis was started. The starting running conditions for the gel are shown in table 13.

Table 13: Running conditions of the gel at the start.

Voltage [V]	Current [mA]	Power [W]	Time [min]
200	120	25	75

After the electrophoresis was completed, the gel cassette was removed out of the chamber and the gel taken out of the cassette and immediately put into the first solution of the staining procedure (see Chapter 4.4).

3.6.5 Silver Staining (MS compatible)

The staining procedure was immediately started after the gel electrophoresis was completed to fix the proteins in their current position and prevent diffusion of the proteins through the gel matrix structure [16].

The gel was fixed for 20 min in the Fixation Solution and then immediately washed with the Washing Solution I for 10 min. Afterwards, the gel was rinsed three times with Washing Solution II for 20 min each, followed by incubation in the Sensitization Solution for 1 min. Subsequently, the gel was rinsed again twice with Washing Solution II for 1 min each, to get rid of the excess solution. For the silver staining, the gel was then placed into the Incubation Solution, which was stored in the fridge, for 20 min. Afterwards, the gel was rinsed twice with Washing Solution II for 1 min each and incubated in the Development Solution until the bands of the protein standard marker as well as the samples were visible. The reaction was then stopped immediately with Stopping Solution, three times for 5 min each. Afterwards, the acquired gel was scanned and at the end stored in the Conservation Solution at 4 °C [16].

3.6.6 *In-Gel* digestion and ZipTip purification

The gel was washed with UHQ water two times for 10 min each, before the excision of the protein bands of interest. The protein bands were cut out of the gel in roughly 1 mm^3 cubes and transferred into a clean 0.5 mL Eppendorf tube each. In addition, pieces of the gel not containing protein were cut out as blanks and processed in parallel with the protein samples. [31, 32]

The gel particles were washed with a $Na_2S_2O_3$/Potassium hexacyanoferrate (III) (1:1) solution twice for 5 min per change. The solvent volume used in all the steps should roughly equal to the gel volume and valued in this experiment 10 µL. Then, all the liquid was removed and 10 µL ACN was added to the gel particles and incubated for 5 min. The ACN was removed and the milky and sticky gel particles were

rehydrated in 10 µL of 100 mM NH_4HCO_3 for 5 min. Subsequently, 10 µL of ACN was added to achieve a solution ratio of 1:1 and incubated for 15 min. After the incubation the liquid was removed, and the gel particles were dried in a vacuum centrifuge for 30 min.

The dried gel particles were rehydrated in 10 µL freshly prepared reducing solution and incubated for 45 min at 56 °C to reduce the protein. The Eppendorf tubes were chilled to room temperature afterwards, the condensate was briefly centrifuged down and removed from the tubes. It was quickly replaced by 10 µL of freshly prepared alkylation solution and incubated at room temperature for 30 min in the dark. Furthermore, the alkylation solution was removed, and the gel particles were washed with 10 µL of 100 mM NH_4HCO_3 for 5 min. Subsequently, 10 µL of ACN were added (1:1, v/v at the end) and the gel particles were incubated for 15 min. Afterwards, all liquid was removed, and the gel particles dried in a vacuum centrifuge for 30 min.

The dried gel particles were rehydrated in 10 µL of trypsin solution and incubated for 45 min at room temperature. The remaining enzyme solution was removed and replaced by 10 µL of digestion buffer. The gel particles were digested at 37 °C overnight (approx. 16 h).

After the overnight digestion the digested supernatant was recovered in a fresh 0.5 mL Eppendorf tube. At first the gel particles were incubated in 10 µL of 50 mM NH_4HCO_3 for 15 min, followed by the addition of 10 µL of ACN (1:1, v/v) and incubation for 15 min. The supernatant was recovered and pooled with the previous supernatant. Furthermore, 10 µL of 1 % FA were added to the gel particles and incubated for 15 min, followed by the addition of 10 µ of ACN (1:1, v/v) and incubation for 15 min. The supernatant was recovered and pooled with the previously acquired supernatants. The extraction was repeated once more, and all the extracts were pooled together. The samples were dried in a vacuum centrifuge and then dissolved in 10 µL of 0.1 % TFA and briefly sonicated.

For the further analysis via MALDI-MS the samples had to be desalted and concentrated, which was done using Zip-Tip purification. Therefore, the Zip-Tip C18 pipette tips were wet three times with 0.1 % TFA/ACN (1:1) and equilibrated afterwards by wetting the Zip-Tip three times with 0.1 % TFA. The peptides were bound to the column material by pipetting the peptide solution up and down slowly for 8 times. Furthermore, the Zip-Tip was washed three times with 0.1 % TFA and the peptides were eluted with 3 µL of Matrix Solution directly onto the target plate.

The sample preparation for this experiment was the direct-elution method due to its homogenous mixture of matrix and sample peptides. The peptides are therefore bound to a ZipTip C18 pipette tip, directly eluted onto the MALDI target with the appropriate Matrix Solution and air dried at room temperature. A MTP 384 ground steel target from Bruker was used for the measurements.

In addition to the standard digestion process a modified version for more hydrophobic peptides was performed. Therefore, the amount of ACN and the concentration of FA during the extraction of the peptides from the gel particles were modified [33].

After the overnight digestion at 37 °C the supernatant was recovered and placed in a fresh 0.5 mL Eppendorf tube. The gel particles were then incubated in 10 µL of 50 mM NH_4HCO_3 for 15 min, followed by the addition of 10 µL of ACN (1:1, v/v) and incubation for 15 min. The supernatant was recovered and pooled with the previously obtained supernatant. Furthermore, the gel particles were incubated in 10 µL of 5 % FA for 15 min, followed by the addition of ACN (1:1, v/v) and incubation 15 min. The

supernatant was recovered and pooled with the previous supernatants and the extraction was repeated one more time.

The difference during the Zip-Tip purification was that the peptides were eluted with 3 µL of Matrix solution, which consisted of 70 % ACN and 30 % 0.1 % TFA/UHQ. The eluted peptides-matrix mixture was air dried at room temperature on the target.

3.6.7 Protein Identification by MALDI-RTOF – Mass spectrometry

3.6.7.1 Peptide Mass Fingerprint (PMF) Analysis

The mass spectrometric analysis was performed with the UltrafleXtreme (Bruker Daltonics, Bremen, DE). The instruments operates under high vacuum (10^{-6} – 10^{-8} mbar) with a pulsed Nd:YAG laser at 355 nm and a frequency of 1 kHz. The acceleration voltage in the instruments values 20 keV. The mass spectra were recorded in the reflectron time-of-flight mode (RTOF). The instrument setting were 500.0 - 5000.0 Da as a m/z range and a laser power of 50-60 %. 8000 - 10000 single profiles were taken per mass spectra.

As an external calibration substance PepMix 4, which covers a range from 500 - 3500 Da, was used. The stock solution was diluted 1:10 and 1.5 µL were used for each calibration spot. The different standard peptides from PepMix 4 are displayed in table 14, a corresponding mass spectrum in figure 22.

Table 14: Calibration standard peptides of PepMix 4 with m/z values.

Calibration Peptides	**m/z of $[M + H]^+$**
Bradykinin [1-5]	572.7
Angiotensin II	1046.2
Neurotensin	1672.9
ACTH [18-39]	2465.7
Bovine Insulin chain B	3495.9

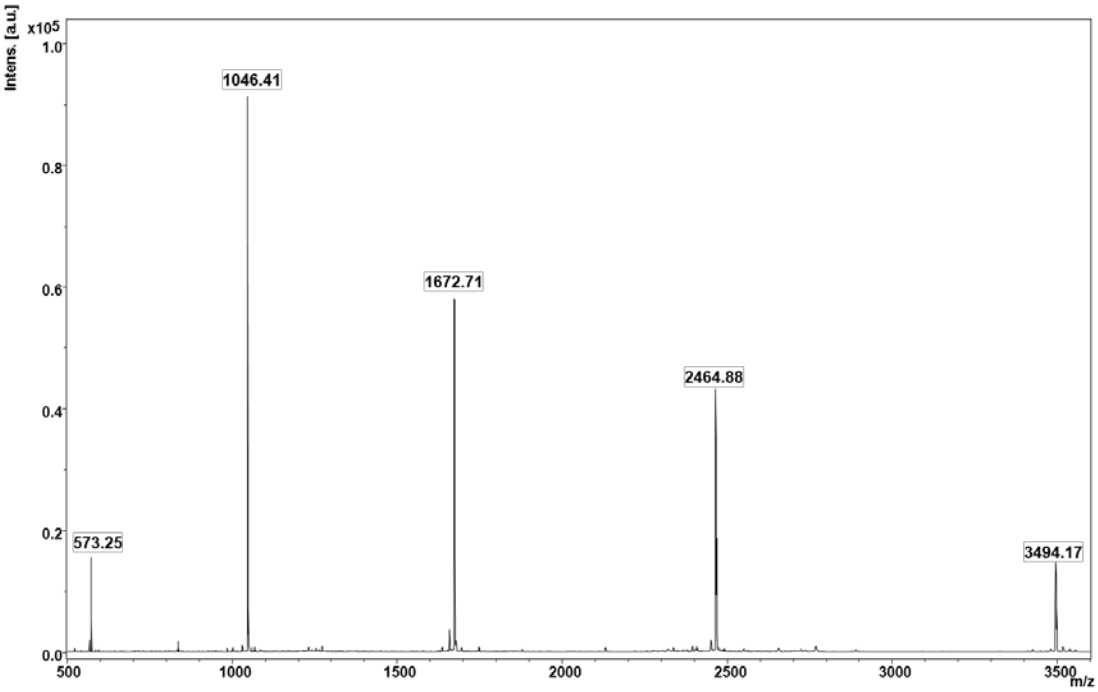


Figure 22: Mass spectrum of PepMix 4, reflectron mode, measured with the same parameters as the samples.

3.6.7.2 Peptide Sequencing

MS^2 analysis was performed with the activation of the LIFT method. The instrument settings were reflectron mode, laser power 80 - 90 % and 9000 - 12000 single profiles per spectra.

An appropriate peak was chosen from the previously acquired mass spectra from the Peptide mass fingerprint analysis via the FlexAnalysis software (Bruker Daltonics, Bremen, DE). The chosen peak was then fragmented with the LIFT method and the fragments were detected.

3.6.7.3 Data Analysis via MASCOT Search

The peptide mass fingerprint spectra were analyzed using FlexAnalysis software from Bruker and searched in MASCOT via the BioTool option in the FlexAnalysis software.

The parameters for the Database searching were:

- Database: NCBI
- Taxonomy: Viridiplantae (Green plants); Rosaceae
- Enzyme: Trypsin
- Missed cleavages: 2
- Fixed modifications: Carbamidomethyl (C)
- Variable modifications: Acetyl (Protein N-Term), Oxidation (M), Deaminated (NQ)
- Peptide mass tolerance: 0.3 Da
- Mass type: $[M + H]^+$
- Last access date: 30.11.2018

For the MS^2 spectra the same parameters were used with the addition of fragment tolerance of 0.5 Da.

The database Viridiplantae, which consists of all the proteins of Green plants, as well as the database Rosaceae was chosen for the relevance they have to the origin of the samples. *M. domestica* belongs to the Taxonomy Rosaceae and is therefore related to organism like pear, peach, rose, plum, almonds, strawberry, blackberry, raspberry and cherry.

The obtained MS and MS^2 data were searched in both Viridiplantae and Rosaceae to compare the results and gain more information on the possible enzyme in the protein database of *M. domestica* itself.

4. Results and Discussion

4.1. Isolation, Extraction and Purification

During the extraction and purification steps, the biological activity of the enzyme (*p*-coumaroyl-CoA reductase) was tested periodically. Chromatograms of the standard mixture (black line) and a test enzymatic reaction (blue line) are compared in Figure 23. The first, smaller peak at 6.25 min elution time is dihydro-*p*-coumaric acid, while the second peak at 6.75 min elution time is *p*-coumaric acid. Additionally, UV/Vis spectra of dihydro-*p*-coumaric acid are displayed in figure 24.

Sample fractions were enzymatically tested and resulting chromatograms compared to a chromatogram of the standard mixture of both standards. Additionally, the UV/Vis spectra of peaks in question were investigated. If there was no peak at the dihydro-*p*-coumaric acid standard elution and no possible peak in the UV/Vis spectrum, the enzymatic assay was defined as negative. If peaks with a similar elution time were recorded, the area ratio between the standard peak and the sample peak was compared and an approximation of the conversion rate and therefore the concentration of the enzyme was determined.

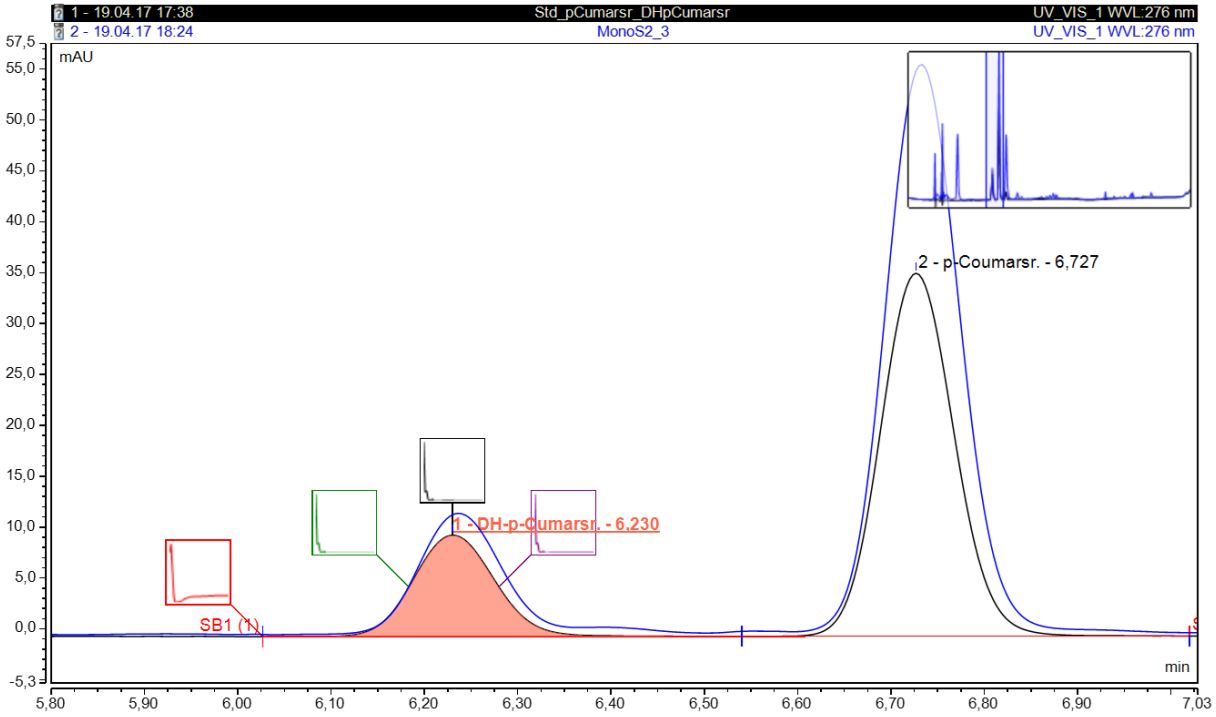


Figure 23: Illustration of the chromatograms for the detection of putative p-coumaroyl-CoA reductase enzymatic activity. Black line: Chromatogram of the standard mixture containing p-coumaric acid and dihydro-p-coumaric acid, Blue line: Exemplary chromatogram of an enzymatic test extract. The chromatogram illustration was provided by the master student Lukas Eidenberger, BSc. From Prof. Halbwirth's workgroup.

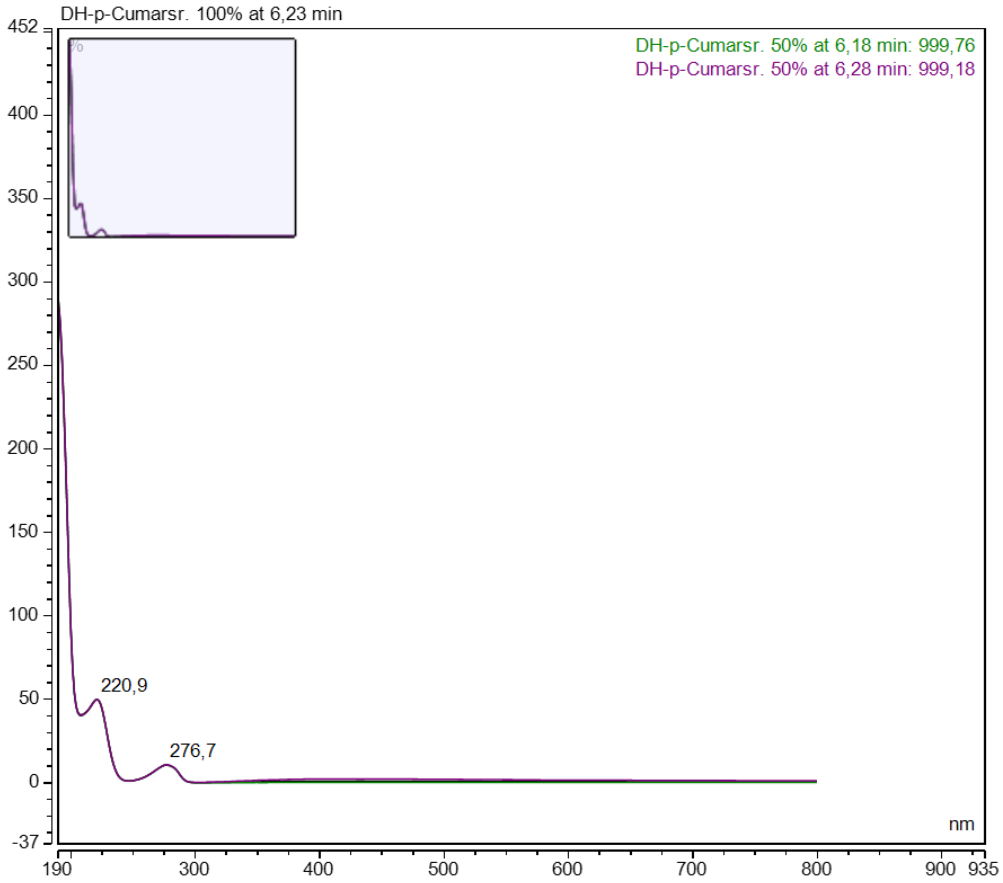


Figure 24: Illustration of the UV/Vis spectra to the corresponding chromatograms. The UV/Vis spectra were provided by the master student Lukas Eidenberger, BSc. From Prof. Halbwirth's workgroup.

The purification of the isolated and extracted proteins was carried out via FPLC. The details of the three steps are listed in the Materials and Methods section and additionally given below. The chromatograms of each step are displayed in figures 25-28. The biological activity of the fractions, which showed a peak at the UV 280 nm measurement, where tested with the enzyme assay and those that tested positive are outlined in black in the chromatograms. Furthermore, the biological active fractions of the first and third chromatographic step were analyzed with the bioanalyzer.

Three chromatography steps were performed:

1. Chromatographic step

- Ammonium sulfate pellet was dissolved in Buffer A
- Conductivity of protein solution was 5.6 mS/cm (1.2 L of solution)
- SP-Sepharose fast flow column
- Column XK 16/20
- Column volume: 24 mL
- Flow rate: 10 mL/min
- Buffer A: 30 mM NaOAc, 30 mM Mes, pH 5.5
- Buffer B: 30 mM NaOAC, 30 mM Mes, 1 M NaCl, pH 5.5
- Gradient: linear until 70% B in 18 CV

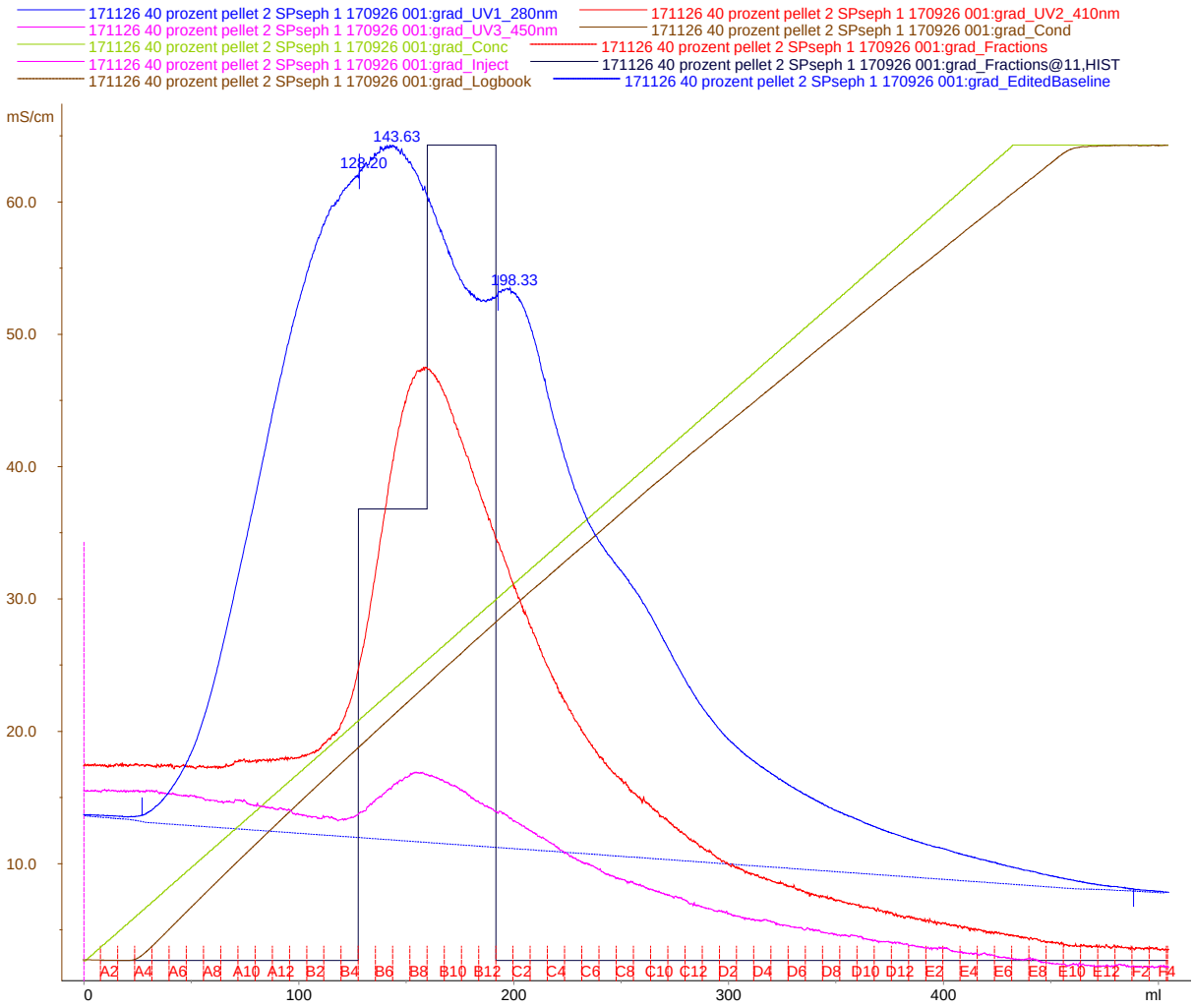


Figure 25: Chromatogram of the first chromatographic step. Blue line: UV 280 nm measurement; Red line: UV 410 nm measurement; Pink line: UV 450 nm measurement; Brown line: Conductivity measurement; Green line: Linear gradient of Buffer B; Black line: Outline of the biologically active fractions (B5-B12), which were then used for the second chromatographic step. The chromatograms were provided by the workgroup of Prof. Halbwirth.

2. Chromatographic step

- Biological active fractions pooled, concentrated to 1 mL and diluted to 40 mL with Buffer A
- Mono S HR 5/5 column
- Column volume: 0.982 mL
- Flow rate: 1 mL/min
- Buffer A: 30 mM NaOAc, 30 mM MES, pH 5.5
- Buffer B: 30 mM NaOAc, 30 mM MES, 1 M NaCl, pH 5.5
- Gradient: linear until 35 % B in 90 CV

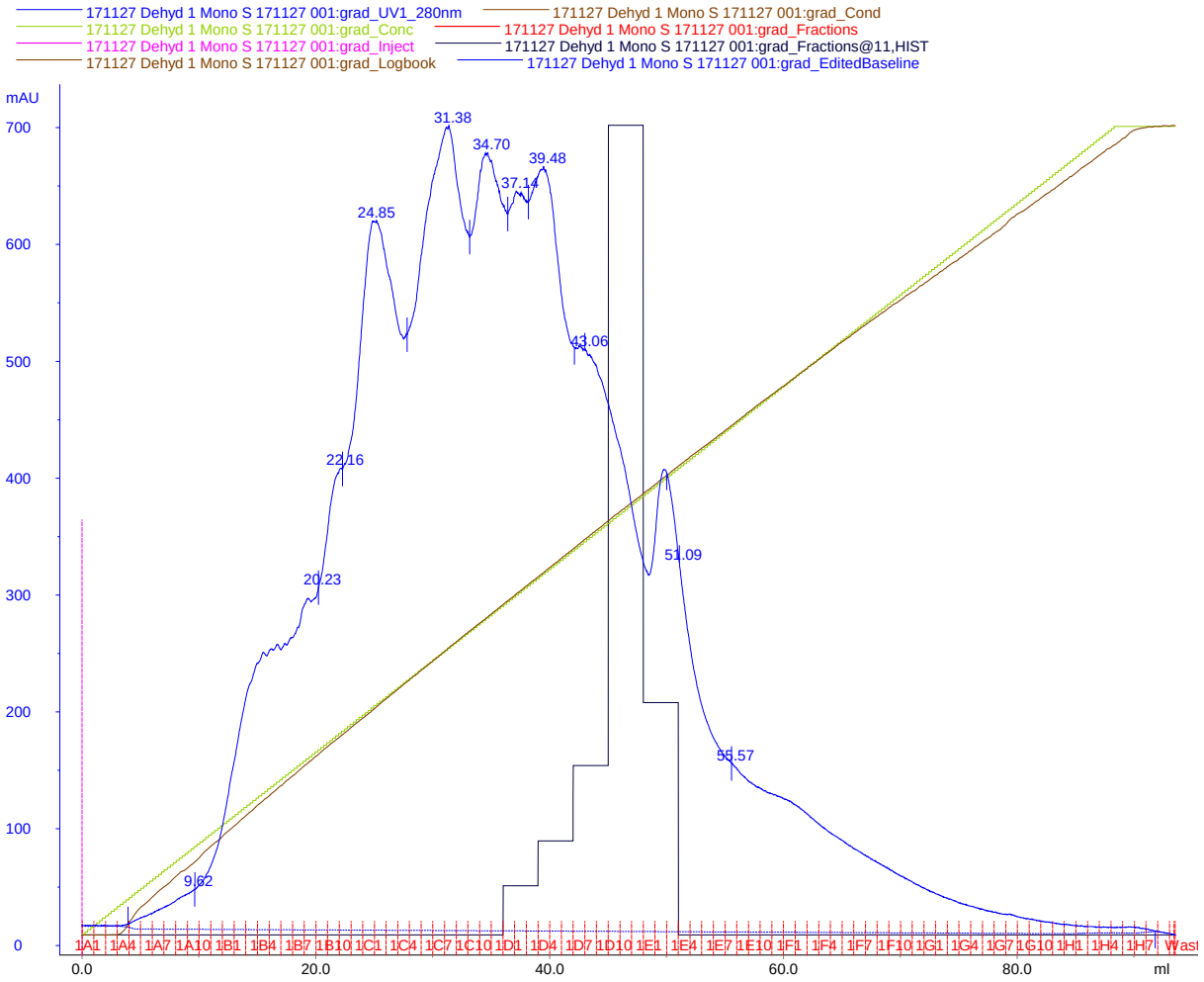


Figure 26: Chromatogram of the second chromatographic step. Blue line: UV 280 nm measurement; Brown line: Conductivity measurement; Green line: Linear gradient of Buffer B; Black line: Outline of the biologically active fractions (D1-E3), which were then used for the third chromatographic step. The chromatograms were provided by the workgroup of Prof. Halbwirth.

3. Chromatographic step

- Biological active fractions pooled and diluted to 40 mL with 30 mM NaOAc, 30 mM MES, pH 5.5
- Mono S HR 5/5 column
- Column volume: 0.982 mL
- Flow rate: 1 mL/min
- Buffer A: 40 mM HEPES, pH 7.2
- Buffer B (Elution Buffer): 40 mM HEPES, 500 mM NaCl, pH 7.2
- Gradient: linear until 35 % B in 135 CV
- Elution with pH 5.5

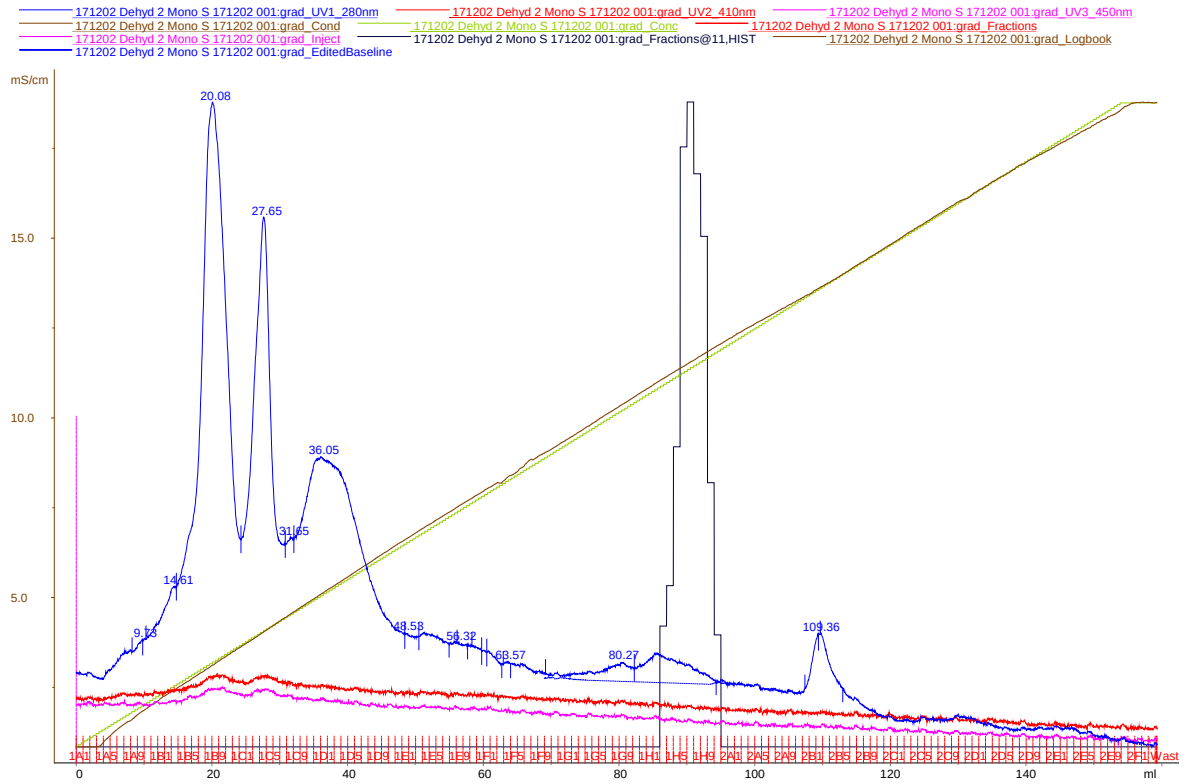


Figure 27: Chromatogram of the third chromatographic step. Blue line: UV 280 nm measurement; Red line: UV 410 nm measurement; Pink line: UV 450 nm measurement; Brown line: Conductivity measurement; Green line: Linear gradient of Buffer B; Black line: Outline of the biologically active fractions (H3-H11). The chromatograms were provided by the workgroup of Prof. Halbwirth.

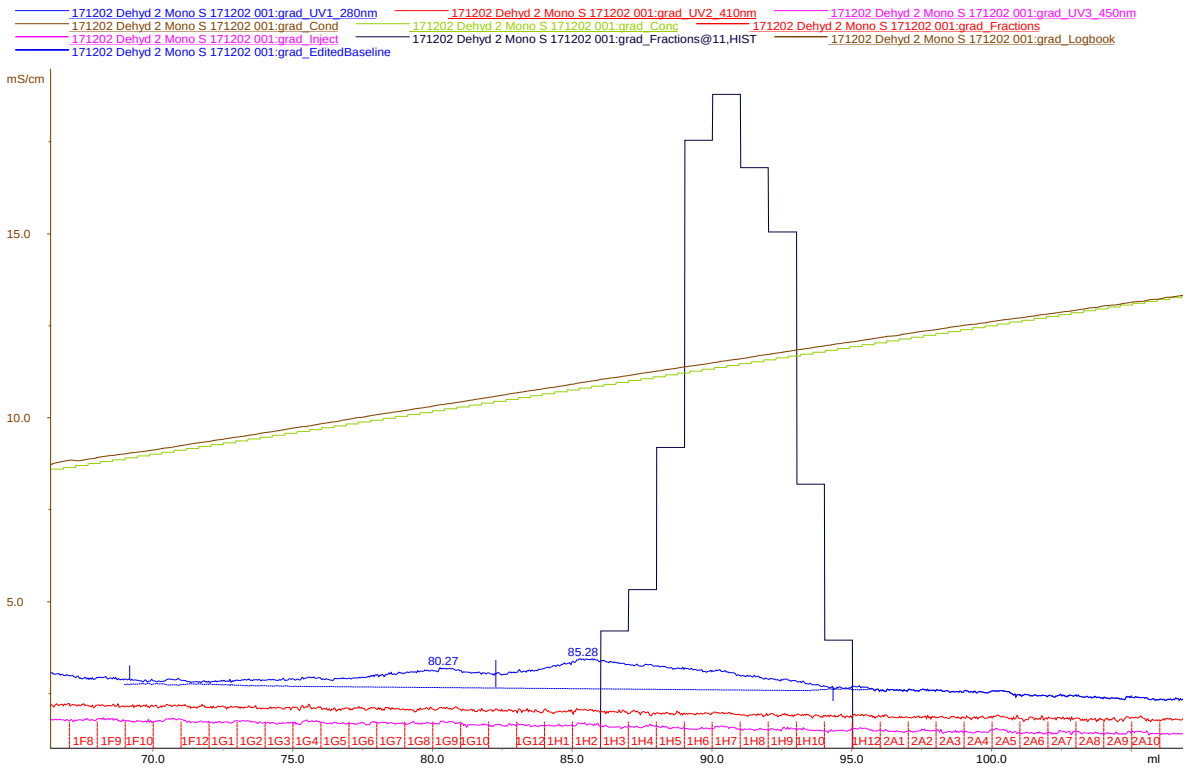


Figure 28: Biologically active fractions of chromatogram of the third chromatographic step (close up, compare to Figure 20). Blue line: UV 280 nm measurement; Red line: UV 410 nm measurement; Pink line: UV 450 nm measurement; Brown line: Conductivity measurement; Green line: Linear gradient of Buffer B; Black line: Outline of the biologically active fractions (H3-H11). The chromatograms were provided by the workgroup of Prof. Halbwirth.

The elution pattern during the three chromatographic purification changes. This is not only due to the fact that different purification setups are employed but also indicates that proteins are removed from the original sample. Mind that only fractions, which display biological activity, are pooled after one corresponding chromatographic step and taken to the next round of purification. The third and last step shows, that there is still a bulk of protein in the sample, which eludes at 20-60 mL. Based on enzymatic activity testing, the protein of interest eludes between 85 – 95 CV. Those biological fractions were taken and further analyzed with the Bioanalyzer and SDS-PAGE instrumentation.

4.2 Protein concentration

The measured standards for the Pierce 660 nm Protein Assay are displayed in table 15 as well as in figure 29. To gain an efficient regression line as well as a high coefficient of determination (≥ 0.99) standard I was excluded from the regression. The coefficient of determination obtained by the standard measurements was $R^2 = 0.9942$.

Table 15: Standard concentration used for the standard curve of the Pierce 660 nm Protein Assay.

Standard	BSA Concentration [µg/mL}	Average Absorbance []
A	2000	1.026
B	1500	0.773
C	1000	0.519
D	750	0.389
E	500	0.243
F	250	0.141
G	125	0.114
H	25	0.097
I	0	0.093

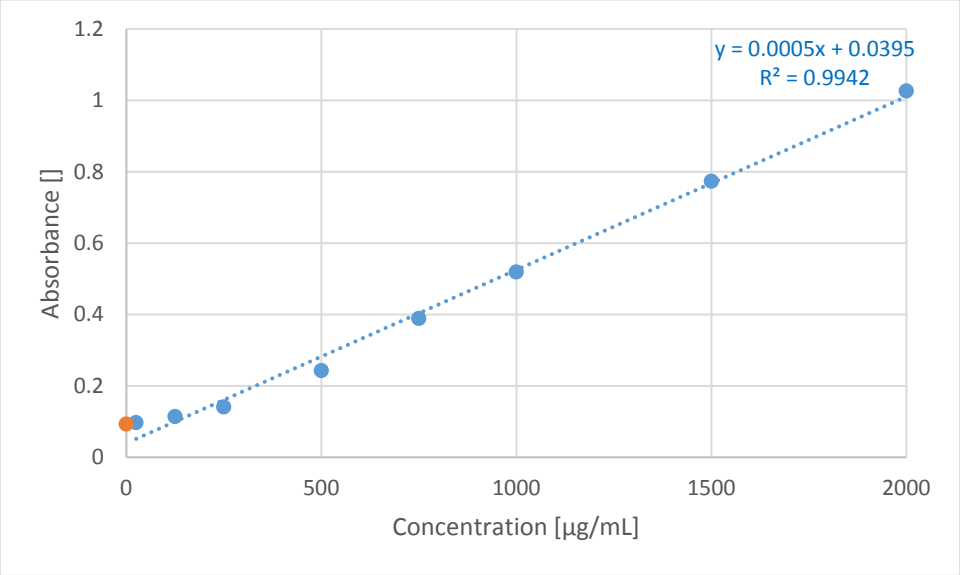


Figure 29: Pierce 660 nm Protein assay of the standards. Standard I was excluded from the standard curve. Standards A-H: blue dots; Standard I: orange dot.

The concentration of the protein sample before and after the buffer exchange was photometrically measured and determined with the equation gained from the standard regression. The protein concentration of the pooled sample before and after the buffer exchange are given in table 16. Based on these numbers a slight increase in protein concentration after buffer exchange was achieved.

Table 16: Protein concentration calculated according to the equation of the regression line of the standard measurements.

Sample	Average Absorbance []	Average Concentration [µg/mL]
Pool (original)	0.099	121.5
Pool (buffer exchange)	0.107	138.0

4.3 Bioanalyzer

The idea behind the Bioanalyzer approach was to monitor the successive chromatographic steps of the FPLC to find out, which protein peaks are enriched, and which ones removed during the purification process. Therefore, the samples with the highest biological activity were analyzed to gain more insight over analyte purification and the reproducibility of the chromatographic system. Unfortunately, the samples of the second chromatography were not obtainable. Thus, only a differentiation between the first and the last chromatographic step can be made.

The samples with the highest biological activity (see chapter X) from the first purification step were taken and analyzed with the Agilent 2100 Bioanalyzer system. The samples (B1-C6) were loaded onto two protein chips and the obtained electropherograms were fitted with peaks in the Origin software by OriginLab (Northampton, MA, USA, version 9.1).

The different electropherograms and their fitted peak information are shown in tables 17-34 and figures 30-47.

Table 17: Fitted peak information of the electropherogram of sample B1.

Peak	x	A	FWHM	Height
1	24.08	8.74	0.24	34.86
2	24.47	7.89	0.47	15.74
3	25.02	5.32	0.47	10.62
4	29.78	125.12	0.71	166.38
5	30.16	204.0	0.70	275.70
6	30.95	36.35	0.71	48.34

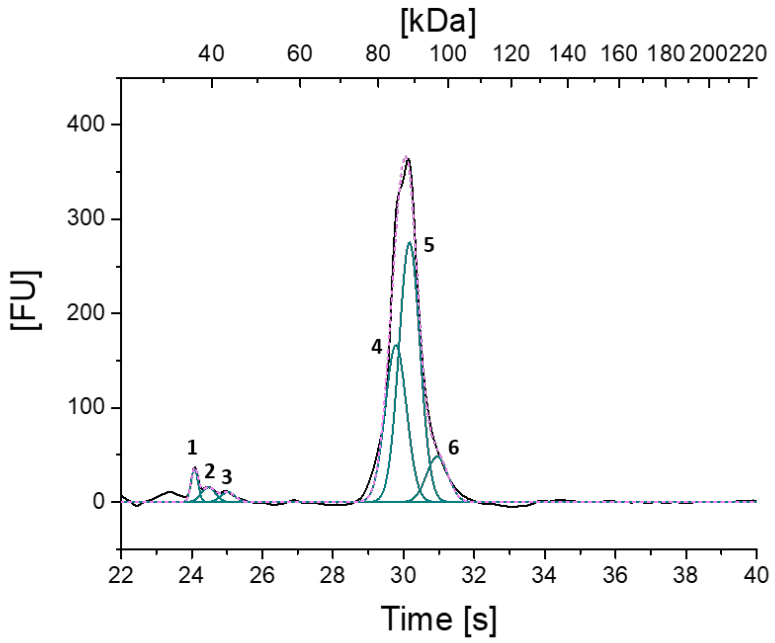


Figure 30: Electropherogram of sample B1 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-6.

Table 18: Fitted peak information of the electropherogram of sample B2.

Peak	x	A	FWHM	Height
1	23.75	4.58	0.56	7.75
2	24.23	12.42	0.65	18.02
3	24.75	15.81	0.88	16.82
4	25.11	5.35	0.65	7.77
5	29.55	97.87	0.71	129.69
6	29.76	76.08	0.31	233.75
7	30.12	281.77	0.55	483.20
8	30.61	82.45	0.88	87.72

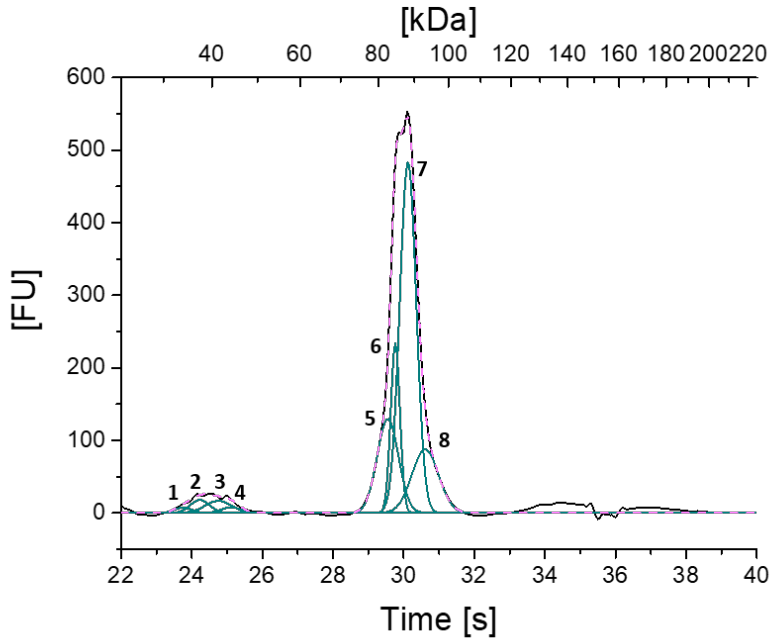


Figure 31: Electropherogram of sample B2 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-8.

Table 19: Fitted peak information of the electropherogram of sample B3.

Peak	x	A	FWHM	Height
1	24.48	10.76	0.57	17.88
2	24.99	7.09	0.36	18.36
3	29.30	80.90	0.50	152.04
4	29.67	228.36	0.38	568.90
5	30.05	309.03	0.48	603.93
6	30.38	134.35	0.88	142.93

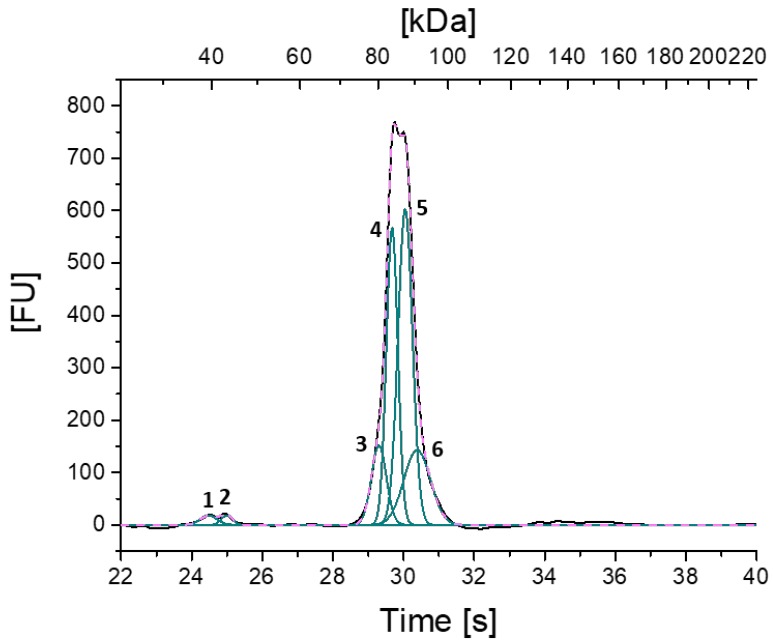


Figure 32: Electropherogram of sample B3 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-6.

Table 20: Fitted peak information of the electropherogram of sample B4.

Peak	x	A	FWHM	Height
1	24.44	14.30	0.48	28.04
2	24.97	13.38	0.35	36.36
3	27.41	3.00	0.59	4.79
4	29.27	117.18	0.49	224.59
5	29.67	380.60	0.43	822.64
6	30.06	281.33	0.44	599.35
7	30.40	127.26	0.82	145.48

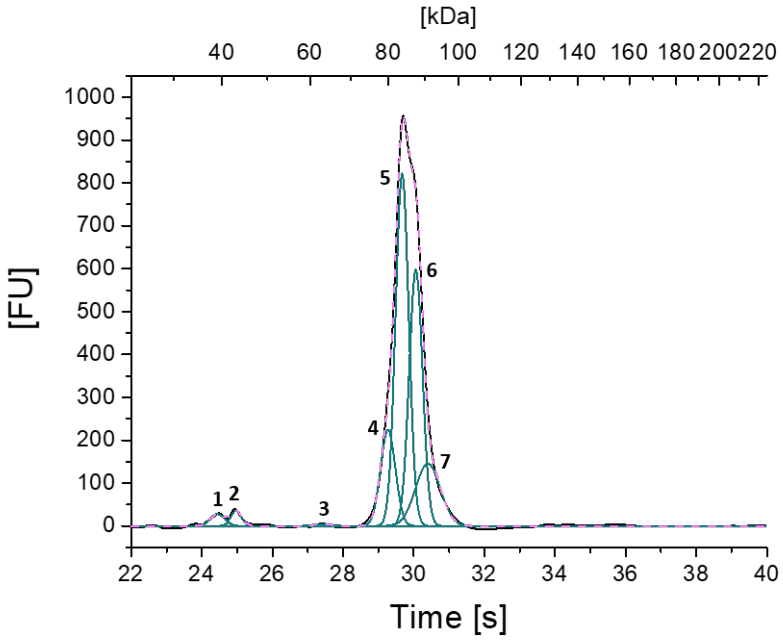


Figure 33: Electropherogram of sample B4 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-7.

Table 21: Fitted peak information of the electropherogram of sample B5.

Peak	x	A	FWHM	Height
1	24.44	11.78	0.47	23.66
2	25.00	10.31	0.36	26.87
3	27.44	3.42	0.29	10.95
4	29.45	258.93	0.63	388.64
5	29.68	118.14	0.32	348.12
6	30.00	235.89	0.50	439.31
7	30.15	155.14	1.06	137.53

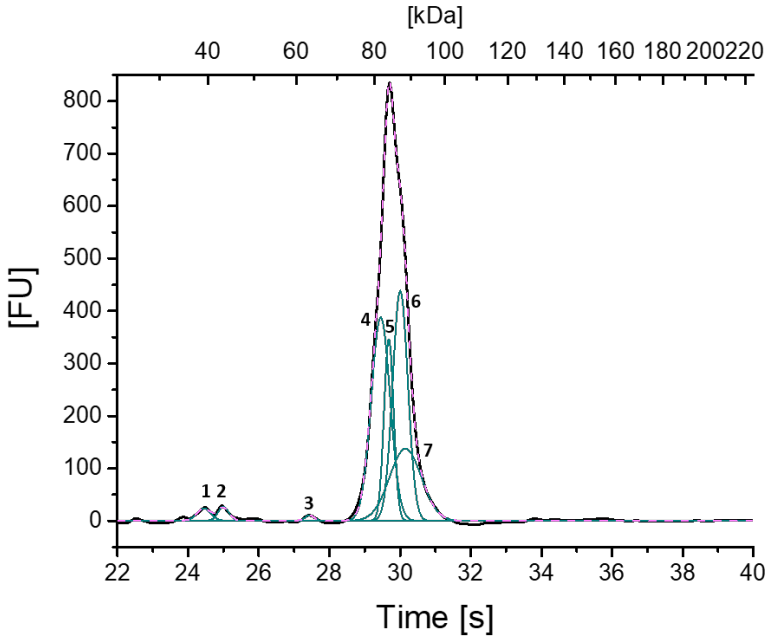


Figure 34: Electropherogram of sample B5 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-7.

Table 22: Fitted peak information of the electropherogram of sample B6.

Peak	x	A	FWHM	Height
1	23.88	3.00	0.35	7.98
2	24.45	9.77	0.33	28.16
3	24.98	10.91	0.37	27.58
4	27.42	5.00	0.33	14.11
5	29.43	225.58	0.64	331.61
6	29.77	201.50	0.47	401.94
7	30.19	108.24	0.47	215.90

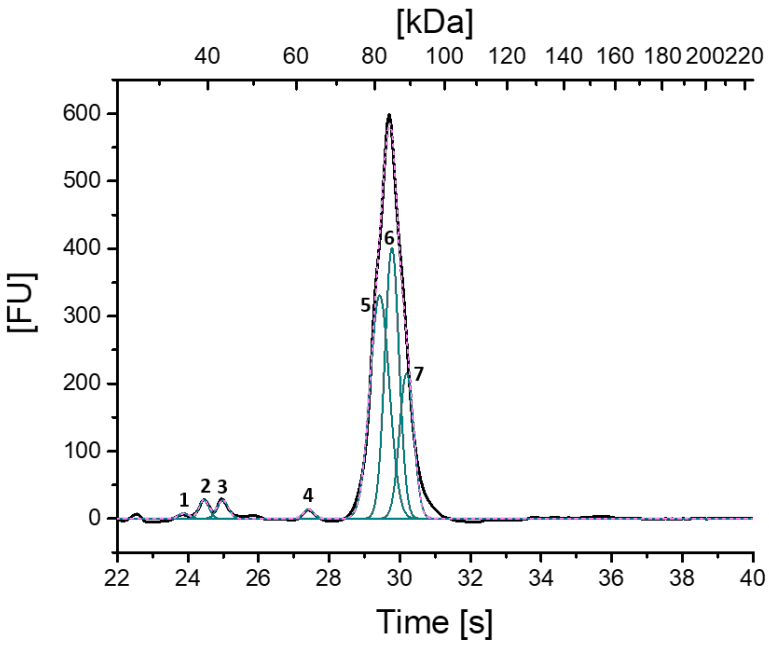


Figure 35: Electropherogram of sample B6 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-7.

Table 23: Fitted peak information of the electropherogram of sample B7.

Peak	x	A	FWHM	Height
1	22.46	3.25	0.23	13.16
2	23.82	1.00	0.35	0.66
3	24.43	6.42	0.34	17.92
4	24.96	6.33	0.34	17.71
5	25.80	1.67	0.47	3.34
6	27.32	2.49	0.29	7.95
7	27.50	1.50	0.29	4.79
8	29.07	41.81	0.54	72.21
9	29.29	37.01	0.28	125.56
10	29.64	204.28	0.56	344.00
11	30.05	114.59	0.76	141.54

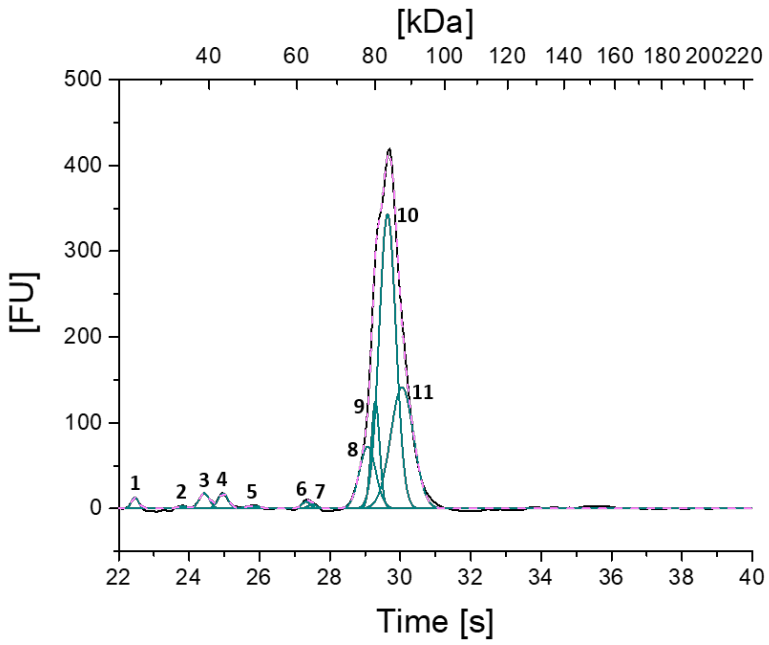


Figure 36: Electropherogram of sample B7 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-11.

Table 24: Fitted peak information of the electropherogram of sample B8.

Peak	x	A	FWHM	Height
1	22.50	7.76	0.41	17.70
2	23.71	0.69	0.19	3.43
3	24.50	4.29	0.36	11.11
4	25.01	3.30	0.31	10.07
5	25.94	1.33	0.31	4.05
6	26.66	1.37	0.39	3.29
7	27.40	2.65	0.32	7.68
8	28.95	14.83	0.39	35.65
9	29.37	57.31	0.38	141.11
10	29.75	109.07	0.53	193.26
11	30.20	45.31	0.68	62.76

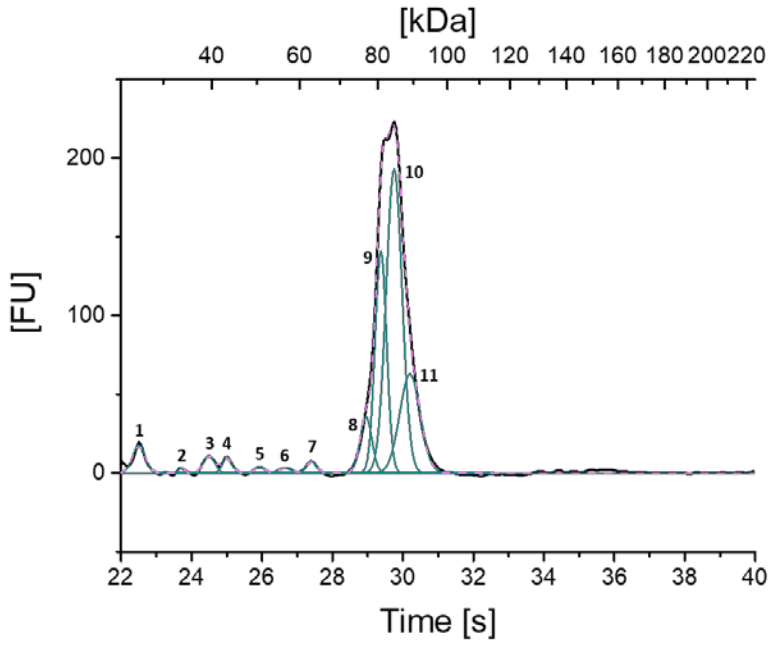


Figure 37: Electropherogram of sample B8 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-11.

Table 25: Fitted peak information of the electropherogram of sample B9.

Peak	x	A	FWHM	Height
1	22.46	10.93	0.47	21.80
2	24.34	1.20	0.24	4.78
3	24.59	2.07	0.26	7.44
4	24.99	1.55	0.27	5.48
5	25.95	1.40	0.35	3.73
6	26.64	3.43	0.51	6.32
7	27.35	1.87	0.35	5.05
8	29.27	31.98	0.77	39.18
9	29.40	9.78	0.26	35.47
10	29.75	68.56	0.66	97.47
11	30.34	13.55	0.60	21.35

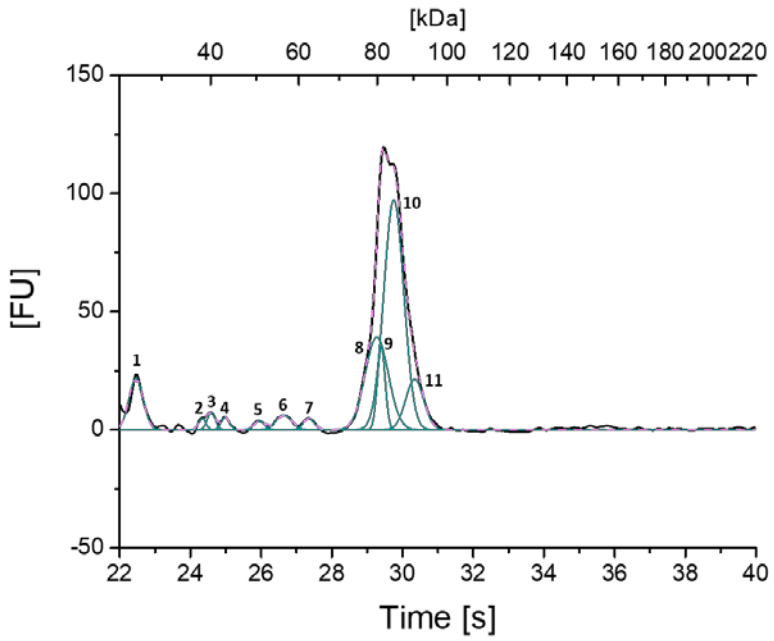


Figure 38: Electropherogram of sample B9 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-11.

Table 26: Fitted peak information of the electropherogram of sample B10.

Peak	x	A	FWHM	Height
1	22.55	13.73	0.47	27.38
2	24.41	1.50	0.24	5.98
3	24.66	1.60	0.21	7.15
4	24.94	1.50	0.38	3.74
5	25.99	1.99	0.33	5.62
6	26.67	8.65	0.69	11.81
7	27.41	2.40	0.37	6.06
8	29.15	18.97	0.63	28.27
9	29.46	24.72	0.37	62.88
10	29.80	19.69	0.41	44.94
11	30.10	33.23	0.81	38.53
12	31.48	4.16	0.82	4.78

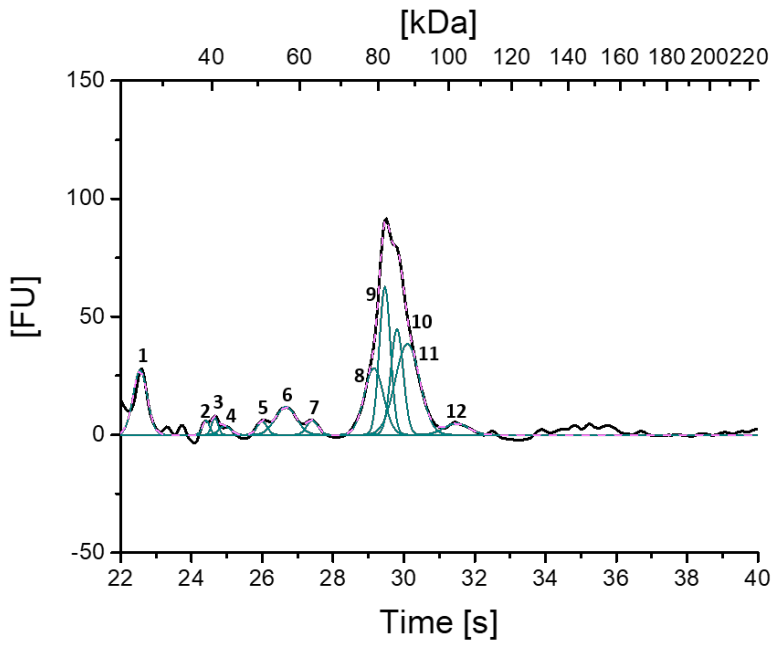


Figure 39: Electropherogram of sample B10 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-12.

Table 27: Fitted peak information of the electropherogram of sample B11.

Peak	x	A	FWHM	Height
1	22.45	13.78	0.47	27.48
2	25.87	2.60	0.39	6.26
3	26.51	11.43	0.70	15.43
4	27.27	1.56	0.36	4.01
5	28.75	3.87	0.52	6.95
6	29.38	31.86	0.63	47.77
7	29.79	5.04	0.35	13.43
8	30.09	15.33	0.68	21.11
9	31.32	3.80	0.76	4.70

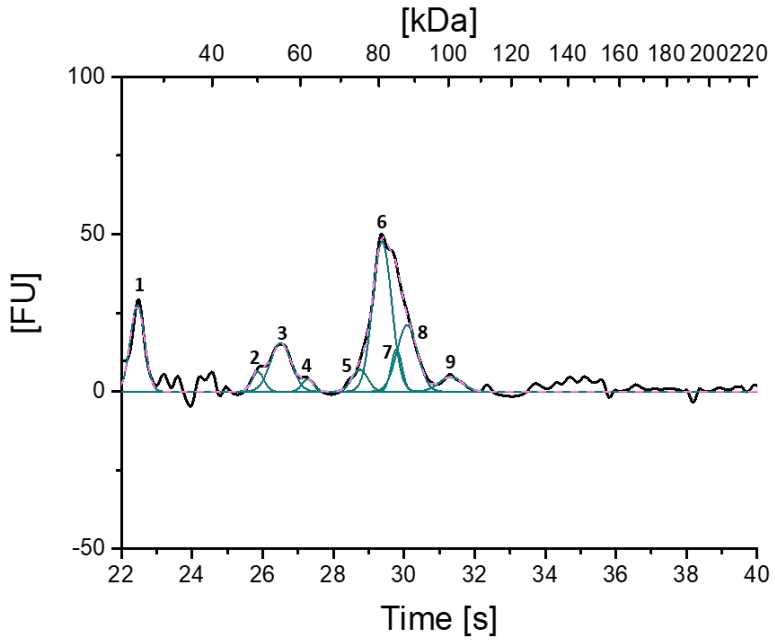


Figure 40: Electropherogram of sample B11 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-9.

Table 28: Fitted peak information of the electropherogram of sample B12.

Peak	x	A	FWHM	Height
1	22.40	15.10	0.47	30.13
2	25.82	3.30	0.43	7.28
3	26.45	11.82	0.70	15.78
4	27.23	1.18	0.39	2.87
5	28.64	1.93	0.58	3.13
6	29.23	6.06	0.56	10.12
7	29.59	19.85	0.89	20.97
8	30.13	6.87	0.86	7.49
9	31.28	3.25	0.60	5.10

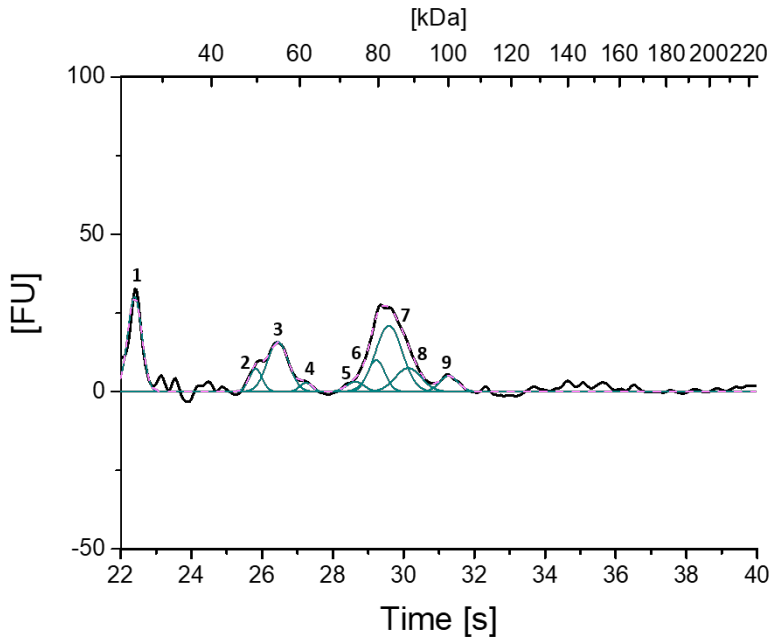


Figure 41: Electropherogram of sample B12 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-9.

Table 29: Fitted peak information of the electropherogram of sample C1.

Peak	x	A	FWHM	Height
1	22.21	24.07	0.47	48.01
2	22.96	4.70	0.33	13.50
3	25.64	6.18	0.55	10.56
4	26.29	14.33	0.78	17.25
5	29.09	6.20	1.12	5.21
6	29.33	13.15	0.81	15.20
7	29.95	8.95	0.79	10.65
8	31.06	3.38	0.59	5.42

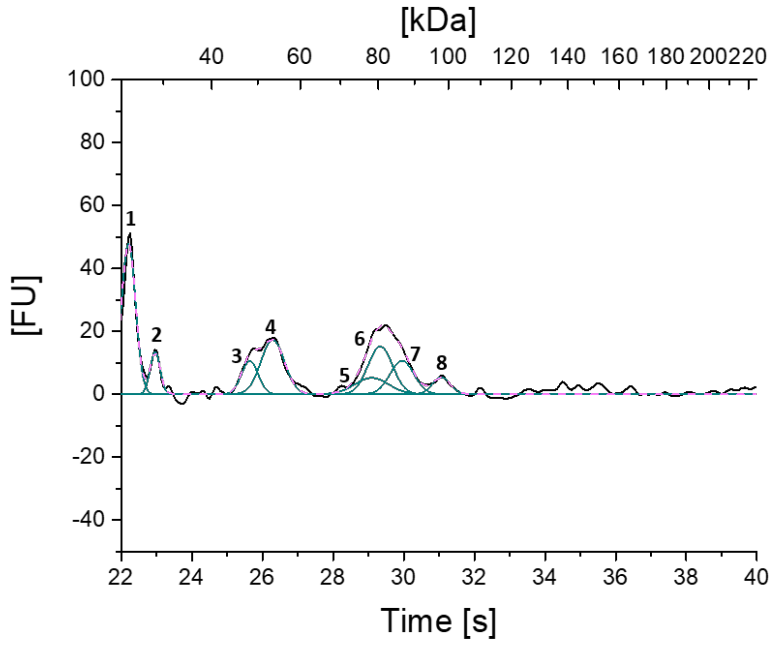


Figure 42: Electropherogram of sample C1 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-8.

Table 30: Fitted peak information of the electropherogram of sample C2.

Peak	x	A	FWHM	Height
1	22.80	6.98	0.41	15.92
2	25.34	4.48	0.65	6.46
3	25.68	12.42	0.77	15.22
4	26.19	3.96	0.48	7.80
5	26.36	7.14	0.96	6.97
6	28.96	7.62	0.94	7.60
7	29.11	1.69	0.61	2.59
8	29.64	13.36	0.94	13.32
9	30.91	2.92	0.56	4.88

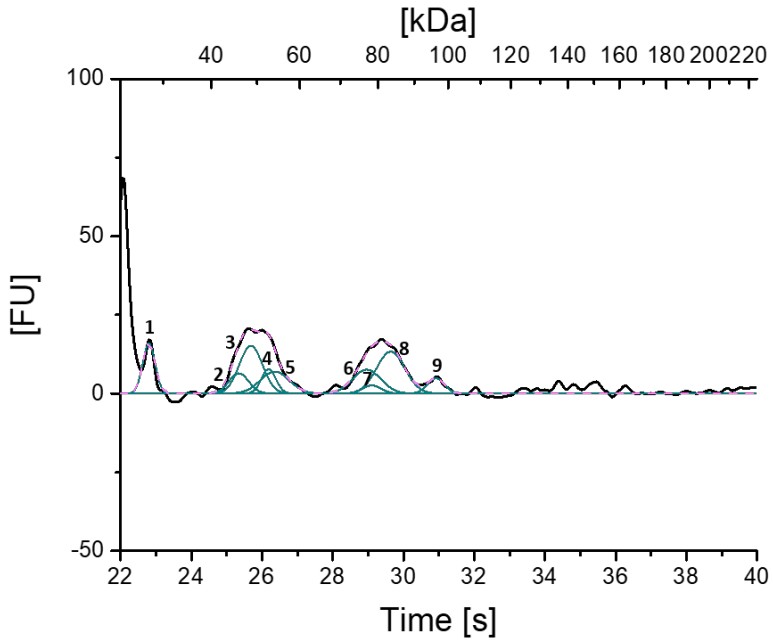


Figure 43: Electropherogram of sample C2 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-9.

Table 31: Fitted peak information of the electropherogram of sample C3.

Peak	x	A	FWHM	Height
1	22.75	7.89	0.41	17.99
2	25.20	4.46	1.07	3.92
3	25.36	10.40	0.59	16.68
4	25.84	10.73	0.76	13.34
5	26.17	12.96	0.98	12.44
6	28.91	8.89	0.97	8.58
7	29.53	7.39	0.86	8.12
8	29.87	4.40	1.05	3.94
9	30.90	1.99	0.47	3.99

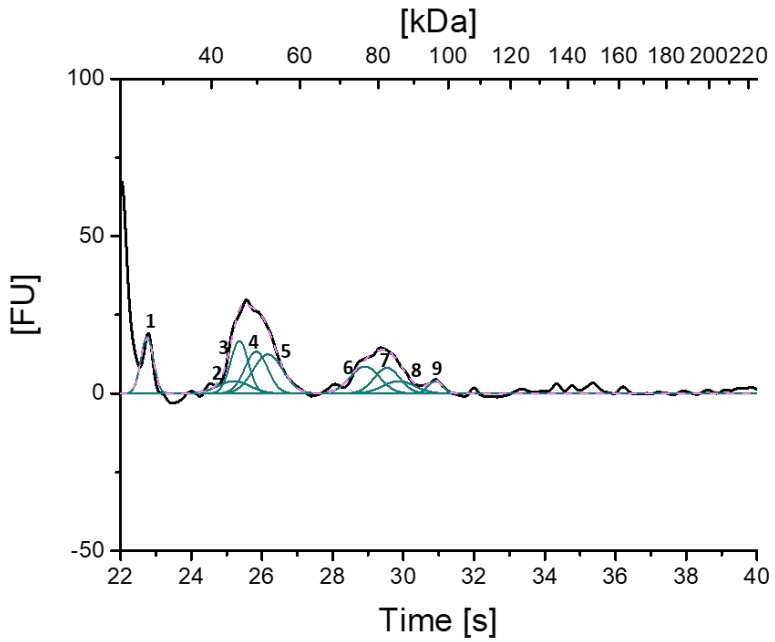


Figure 44: Electropherogram of sample C3 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-9.

Table 32: Fitted peak information of the electropherogram of sample C4.

Peak	x	A	FWHM	Height
1	22.60	4.16	0.35	11.06
2	24.42	1.22	0.35	3.23
3	24.99	7.77	0.41	17.66
4	25.43	16.94	0.55	28.72
5	25.93	10.84	0.60	17.06
6	26.40	3.22	0.62	4.89
7	27.87	1.37	0.47	2.76
8	28.56	3.48	0.44	7.50
9	29.08	3.86	0.59	6.14
10	29.54	6.04	0.86	6.57
11	30.73	1.02	0.47	2.06

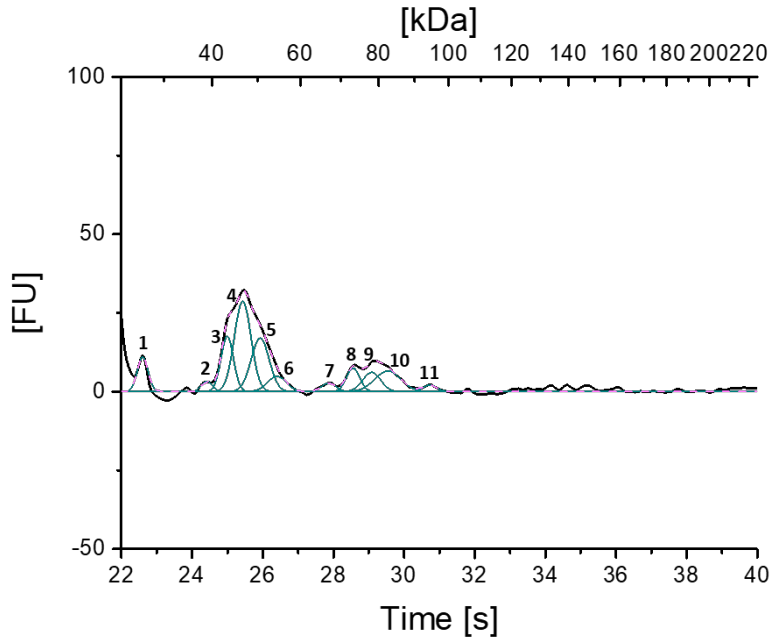


Figure 45: Electropherogram of sample C4 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-11.

Table 33: Fitted peak information of the electropherogram of sample C5.

Peak	x	A	FWHM	Height
1	24.10	1.61	0.39	3.85
2	24.74	9.98	0.46	20.53
3	25.18	12.04	0.46	24.85
4	25.54	22.60	0.94	22.54
5	27.58	1.84	0.48	3.56
6	28.25	3.37	0.46	6.91
7	28.97	7.88	1.06	6.98
8	29.20	0.93	0.82	1.06

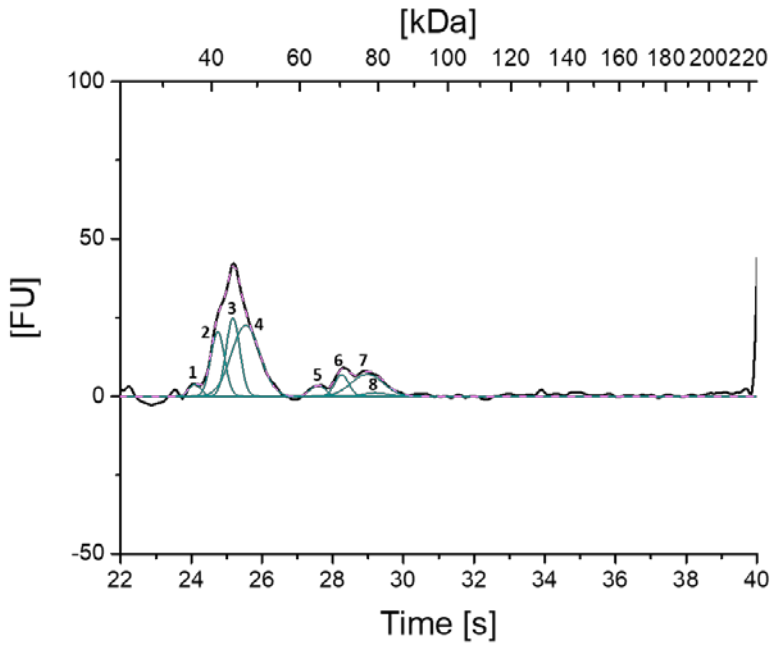


Figure 46: Electropherogram of sample C5 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-8.

Table 34: Fitted peak information of the electropherogram of sample C6.

Peak	x	A	FWHM	Height
1	24.21	1.77	0.41	4.08
2	24.89	12.40	0.45	26.09
3	25.29	15.00	0.43	32.56
4	25.67	26.43	0.94	26.36
5	27.67	2.53	0.47	5.08
6	28.36	4.33	0.44	9.35
7	29.05	7.58	1.01	7.04

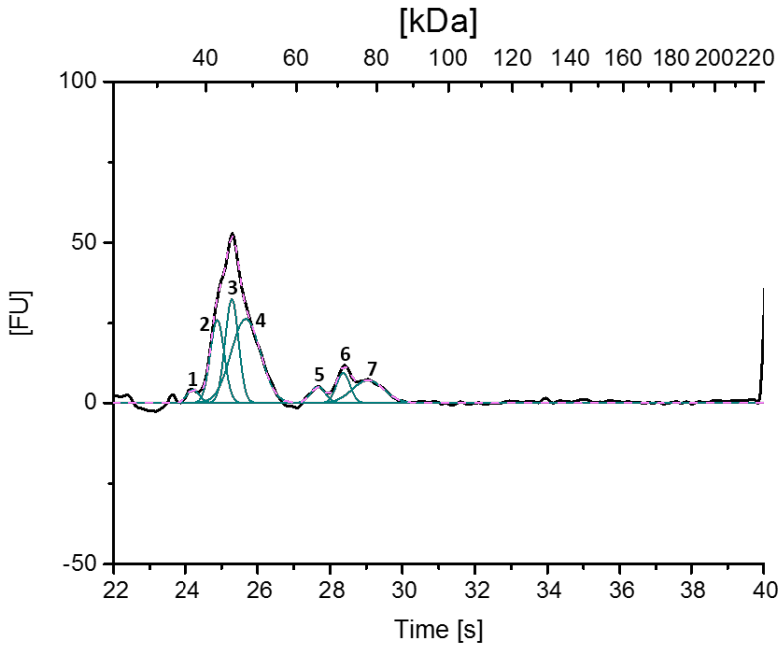


Figure 47: Electropherogram of sample C6 with fitted peaks. The black line is the original electropherogram, the pink dotted line is the fitted electropherogram and the green lines are the single fitted peaks. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. The peaks are numbered from 1-7.

In addition, the electropherograms of all samples were compared in the estimated molecular weight range of the enzyme, which can be seen in figure 48. The samples B6-B8 show a higher peak response in the range of roughly 40 - 45 kDa than the other samples (black, red and light green line). Mind that enzymatic activity was likewise found for these fractions. This could be an indication that the protein concentration of the enzyme of interest is higher in those fractions, than in the other ones. However, the peak height does not give any indication towards the biological activity of the proteins in the different chromatographic fractions. The peak at 80 kDa decreases over the time of the elution during the chromatography.

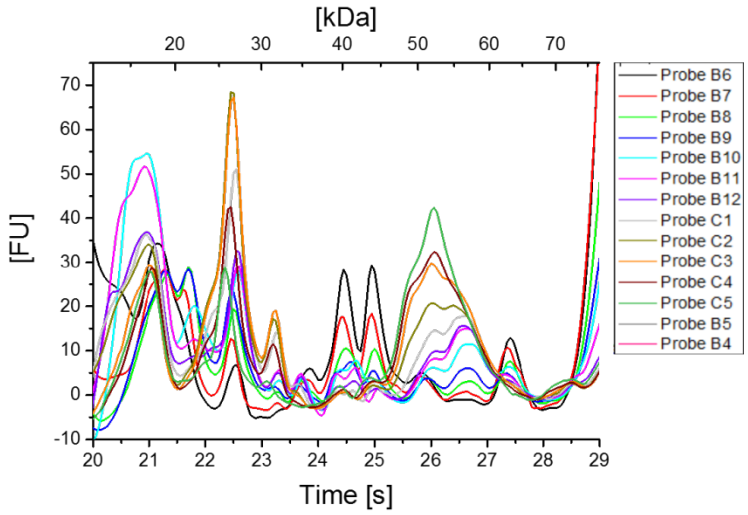


Figure 48: Electropherograms of the samples B4 – C5 put on top of each other. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa]. Probably, the molecular weight range of 40 – 45 kDa is of interest (unpublished data, group of H. Halbwirth). The samples B6 – B8 display a higher peak response than the other samples.

For the analysis of the third and last chromatographic step, three samples without biological activity (B8, C3, D1) were chosen and the pooled sample of the fractions with the highest biological activity (Pool). The samples were measured before and after the buffer exchange to notice if any differences due the change occur. The different samples are displayed in figures 49-56.

The detected fluorescence of the samples was below 10 [FU] and is barely distinguishable from the background, therefore no peaks could be fitted into the obtained electropherograms.

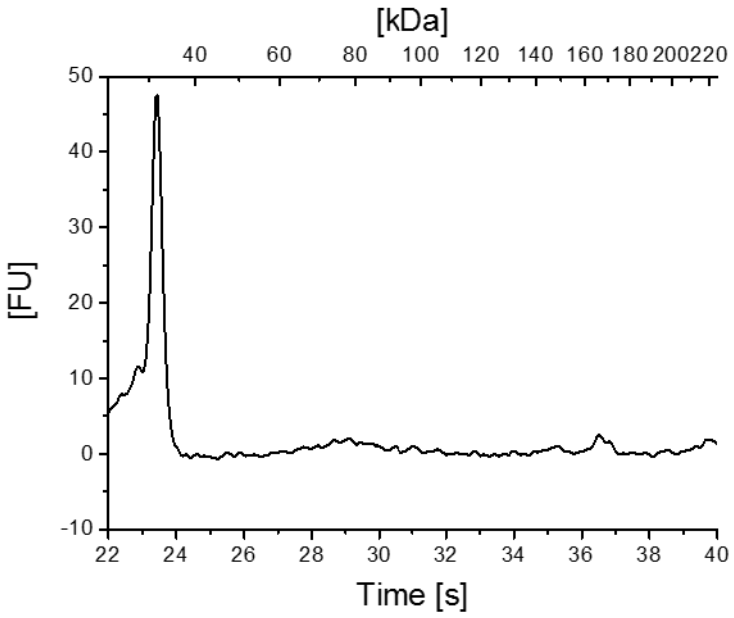


Figure 49: Electropherogram of sample B8 with the chromatographic elution buffer. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa].

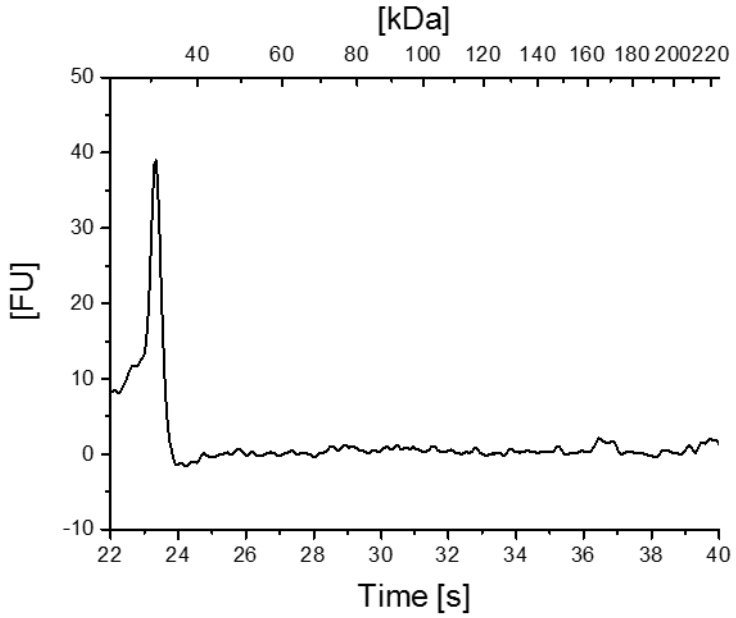


Figure 50: Electropherogram of sample B8 after the buffer exchange. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa].

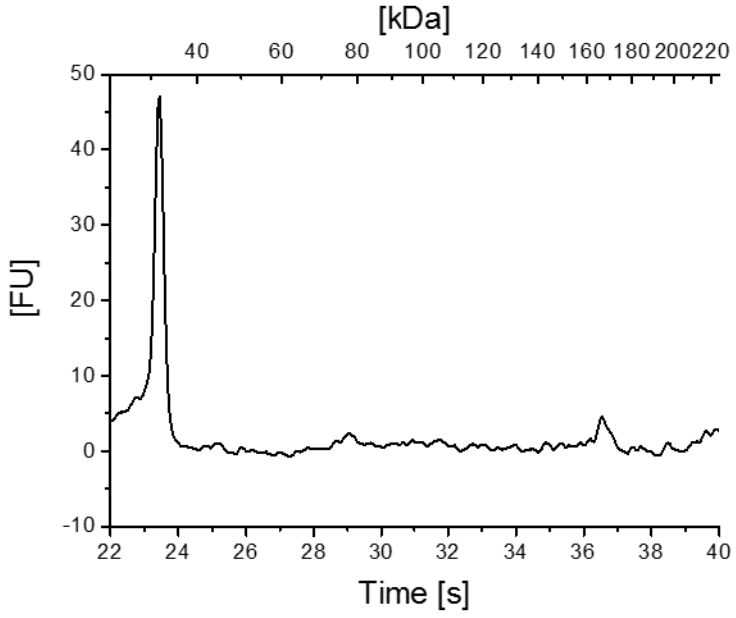


Figure 51: Electropherogram of sample C3 with the chromatographic elution buffer. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa].

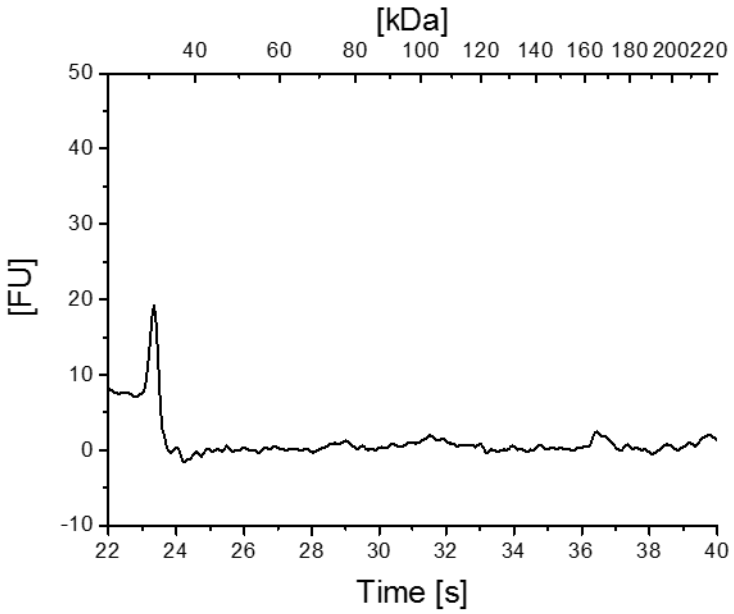


Figure 52: Electropherogram of sample C3 after the buffer exchange. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa].

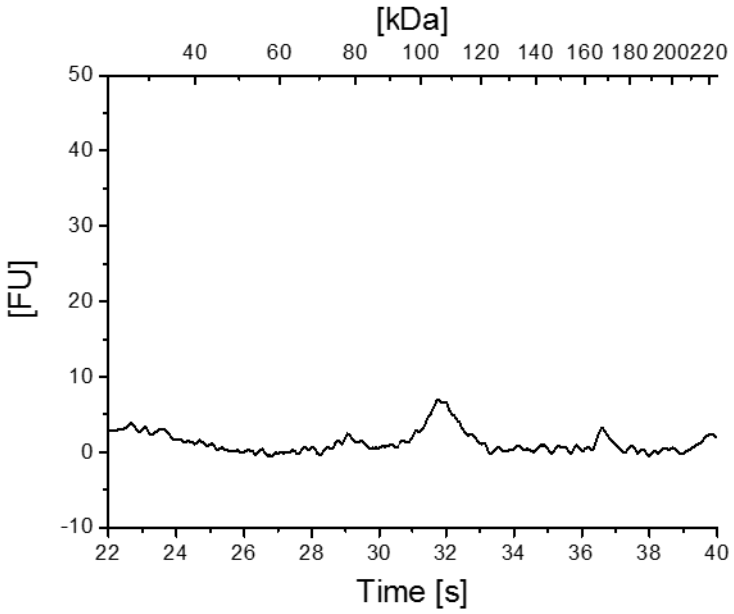


Figure 53: Electropherogram of sample D1 with the chromatographic elution buffer. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa].

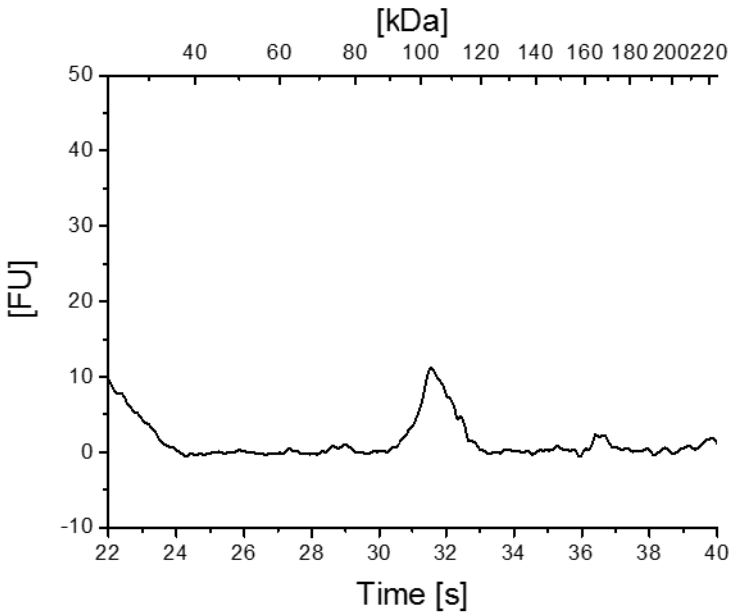


Figure 54: Electropherogram of sample D1 after the buffer exchange. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa].

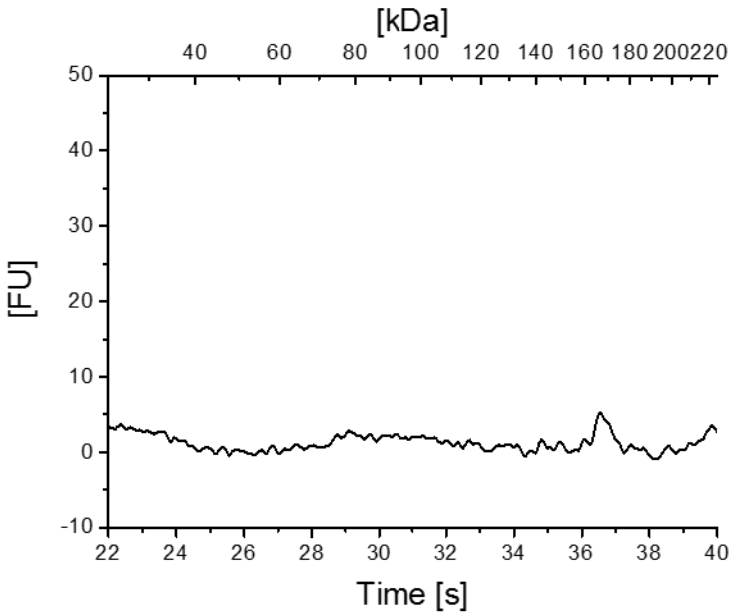


Figure 55: Electropherogram of the Pool with the chromatographic elution buffer. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa].

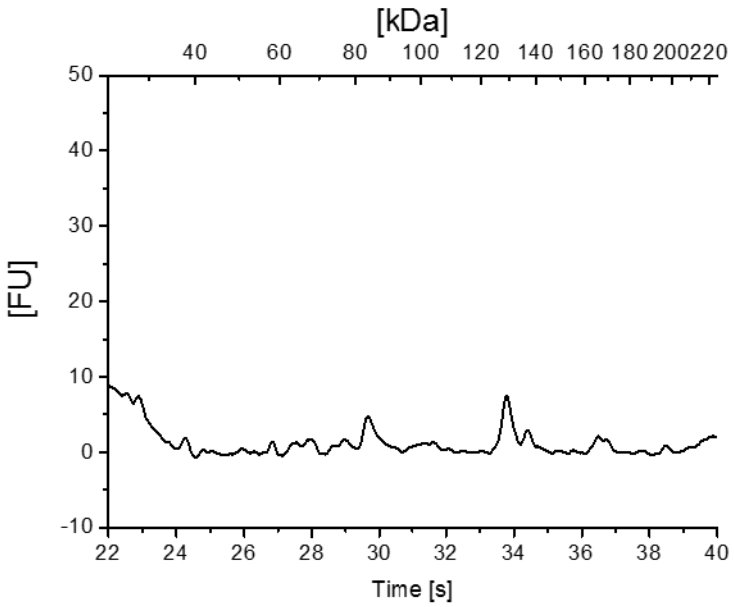


Figure 56: Electropherogram of the Pool after the buffer exchange. The elution time of the peaks [s] was plotted against the fluorescence units [FU] and the molecular weight [kDa].

In comparison to the samples from the first chromatography, it is noticeable that the majority of peaks and therefore proteins, were seemingly eliminated during the process. Hence, the fractions from the third chromatography should have a higher purity than from the other two steps, if they still include any proteins.

Between the samples B8, C3 and D1 and the Pool sample, which shows biological activity (measured directly after the chromatographic step), no differences in the pattern or a higher fluorescence signal is observable. However, the samples before and after the buffer exchange reveal an increase of the quality of the signal to noise ratio. The signal is more pronounced and distinguishable from the background measurement.

4.4 SDS-PAGE

Taking the Pool sample to SDS- PAGE analysis, gels Gel 1 and Gel 2 were silver stained after the 1D gel electrophoresis separation and scanned before the incision of the gel bands and the appropriate blanks. The scanned silver stained gels are displayed in figure 57 and 58.

The total protein amount used for each replicate in Gel 1 is 690 ng. Unfortunately, there was not enough solution of concentrated protein sample (Pool) for a complete second gel, therefore the dilution (1:3 and 1:9) created for the measurement of the Pierce 660 nm Protein Assay were used as well.

It is of note that in contrast to analyses carried out via the Agilent 2100 Bioanalyzer setup, SDS-PAGE analysis with silver staining did relate the presence of proteins over the complete analyzed molecular weight range. This effect might be due to differences in the mode of analyte labelling prior detection.

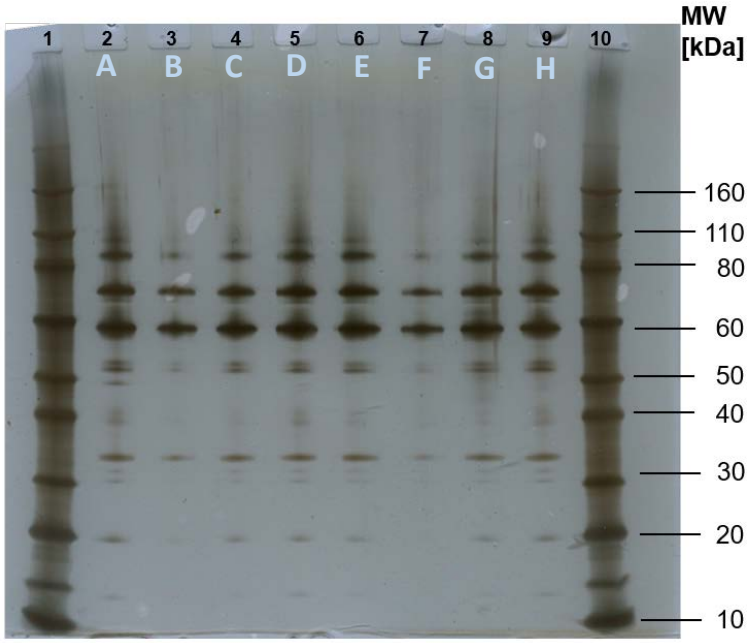


Figure 57: Scan of silver stained Gel 1 with the molecular weight marker. The letters A-H describe the different lanes and are replicates of the protein sample. The total protein concentration of each replicate is 690 ng.

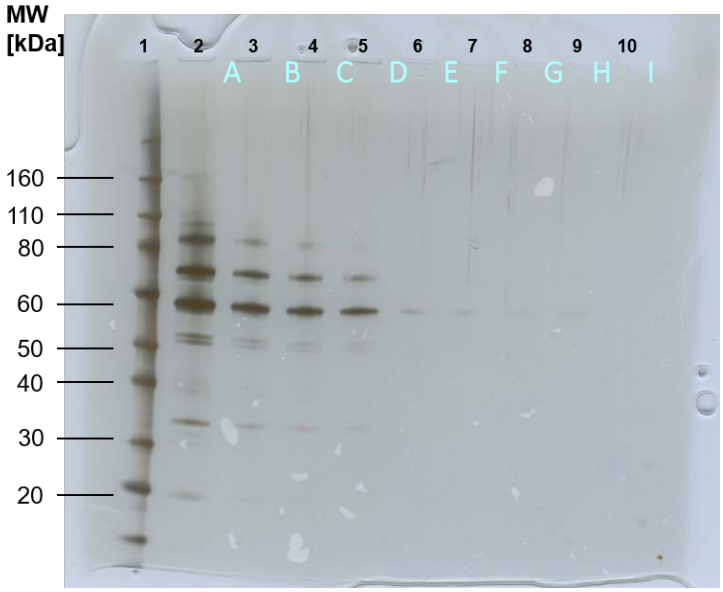


Figure 58: Scan of silver stained Gel 2 with the molecular weight marker. The letters A-H describe the different lanes. A: concentrated protein sample (Pool), B-D: protein sample dilution 1:3, E-H: protein sample dilution 1:9. The total protein concentration of the samples are A: 690 ng, B-D: 230 ng, E-H: 77 ng.

The most intense gel bands appear in a molecular weight range of 60-950 kDa, in addition faint gel bands between 20-60 kDa are visible. For the *in-gel* digestion and identification 8 gel bands were chosen, including the 3 most intense gel bands between 60 and 950 kDa and the gel bands at roughly 53 kDa, 50 kDa, 40 kDa, 35 kDa as well as 20 kDa. The faintly visible gel band at 40 kDa was selected due to an indication that the protein of interest most likely having a molecular weight of 40-41 kDa, according to prior investigations of the working group of Prof. Halbwirth (unpublished data).

The protein distribution in the two gels clearly indicate that it is not a single purified protein, but a mixture of multiple proteins. Bands very close to each other could also indicate isoforms, i.e. proteoforms of the same protein. Therefore, it could be that each gel band could consist of more than one protein, which was considered in the analysis of the Mascot search results.

The gels with the cut-out gel bands and blanks are shown in Figure 59 and 60.

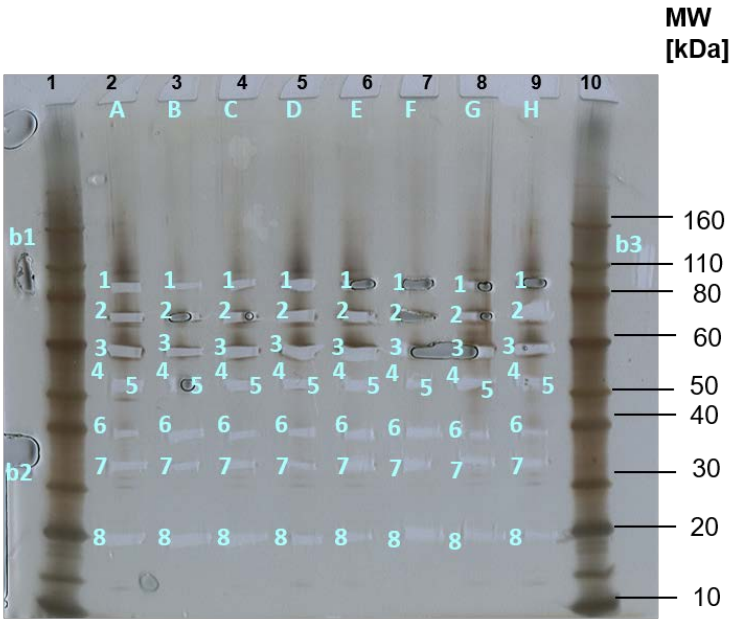


Figure 59: Scan of silver stained Gel 1 with the indications of the cut-out gel bands of each lane. The lanes are described with the letters A-H and the different bands with 1-8. A-H are replicates of the protein sample.

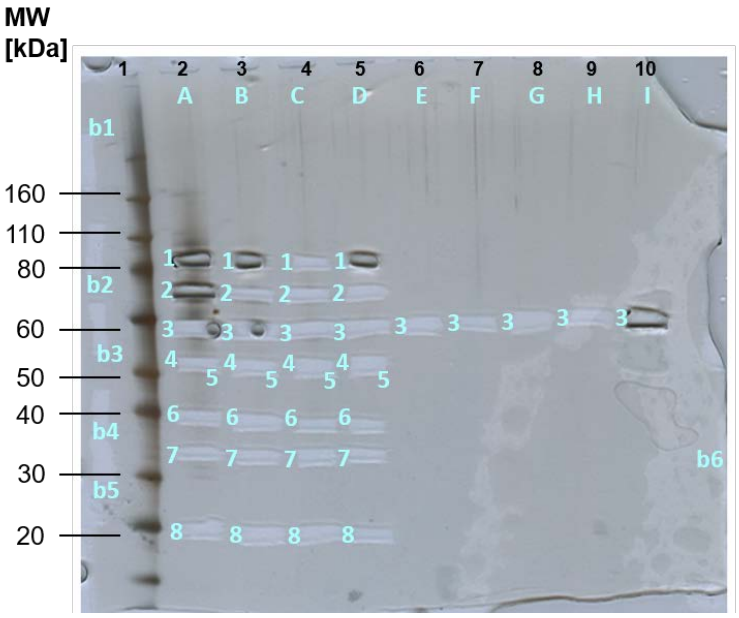


Figure 60: Scan of silver stained Gel 2 with the indications of the cut-out gel bands of each lane. The lanes are described with the letters A-H. A: concentrated protein sample (Pool), B-D: protein sample dilution 1:3, E-H: protein sample dilution 1:9.

The molecular weight values of proteins, obtained through the molecular weight marker on the gel, are estimated as given in table 35.

Table 35: Molecular weights of the cut-out gel bands from Gel 1 and Gel 2.

Gel band	Molecular weight [kDa]
1	95
2	72
3	59
4	53
5	52
6	41
7	34
8	20

All gel bands from the lanes A-H from Gel 1 were digested separately. The gel extracts from the lanes D and E as well as the extracts from the lanes F, G and H were pooled together, dried in the vacuum centrifuge, as described in Chapter X, and analyzed together. This approach was performed to try and raise the peptide amount in the sample spots, which were measured via MALDI RTOF MS.

The gel bands from the lanes A-C from Gel 2 were digested separately and the gel extracts from the lanes B and C were pooled and analyzed together due to the lower total protein amount in those lanes. The lanes D-H were not digested, but rather kept as a backup for additional experiments. The gel bands from lane B and C were extracted using a higher amount of FA and ACN. In addition, the elution of the peptides on the MALDI target were performed with an ACN/ 0.1 %TFA ratio of 70:30. The higher ACN amount was a different approach to ascertain if more peptides from the protein of interest could be extracted and detected. The idea behind this approach was the extraction of more hydrophobic peptides from the gel pieces, in case that the protein of interest itself displays a more hydrophobic behavior [33].

4.5 Protein identification

The protein identification was performed on the extracted gel bands from Gel 1 and Gel 2. The different samples are named according to the lane and the row in the gel (e.g. Gel 1: 1A, 2A, 3A, …; Gel 2: 1A*, 2A*, …). The pooled samples are named in the same fashion (e.g. Gel 1: 1DE, 1FGH,…; Gel 2: 1BC*, 2BC*, …).The MS/MS spectra are named after the sample itself and the m/z value of the fragmented peak (e.g. 1A_1695, 2A_1409, ….)

The obtained mass spectra of the gel bands were compared to the ones gained from the blanks, to filter out the ‘background’ peaks, which derive from the gel itself, and to a contamination list, to eliminate peaks derived from contaminations like keratin, trypsin self-digestion, detergents and BSA.

Afterwards, the chosen peaks of the peptides were searched in the database Viridiplantae (Green plants) of NCBIprot and in the database Rosaceae, which includes *M. domestica* and all the related species like prune, pear, peach, rose, strawberry, raspberry, blackberry, cherry and even almond.

The mass spectra are displayed in the range of m/z 500 – 3500 due to no more appearances of peaks above m/z 3500.

4.5.1 Gel Band 1

In the figures 61-67 the mass spectra of the filtered peptide peaks are shown. The peaks used for the search in Mascot are listed in the Appendix (see Chapter 7.1.1). Additionally, figure 68 displays the MS/MS spectrum, which allows for more information on the single peptides themselves. The peptide sequence and the fragmentation pattern of 1A_1695 is visible in figure 69.

The mass spectra show similar peaks in the region m/z 500 – 1700, but more variance in the higher m/z range.

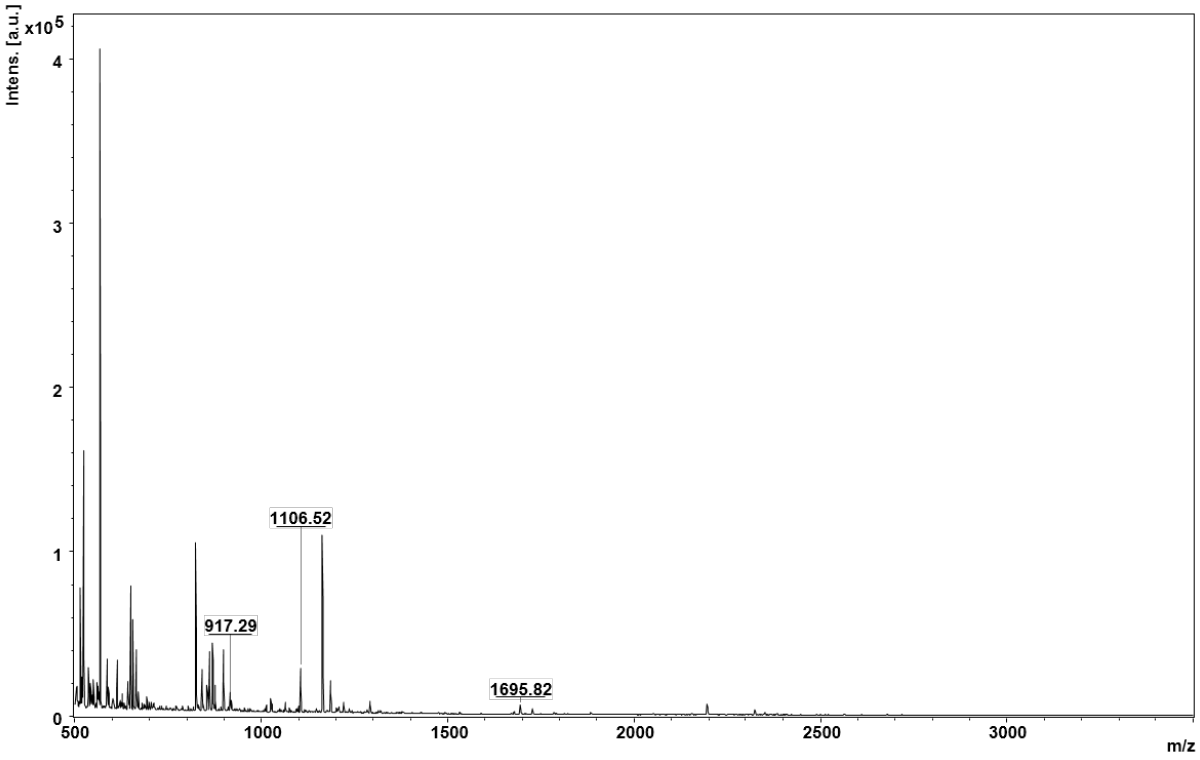


Figure 61: Mass spectrum of sample 1A of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

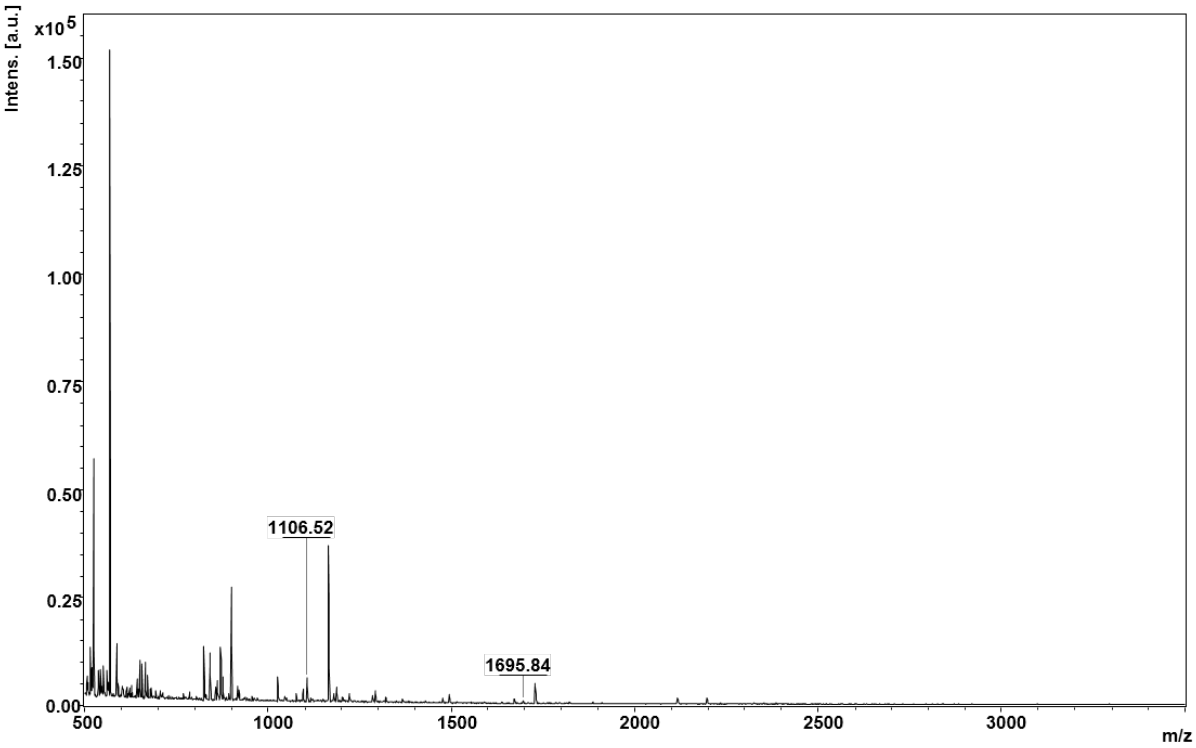


Figure 62: Mass spectrum of sample 1B of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

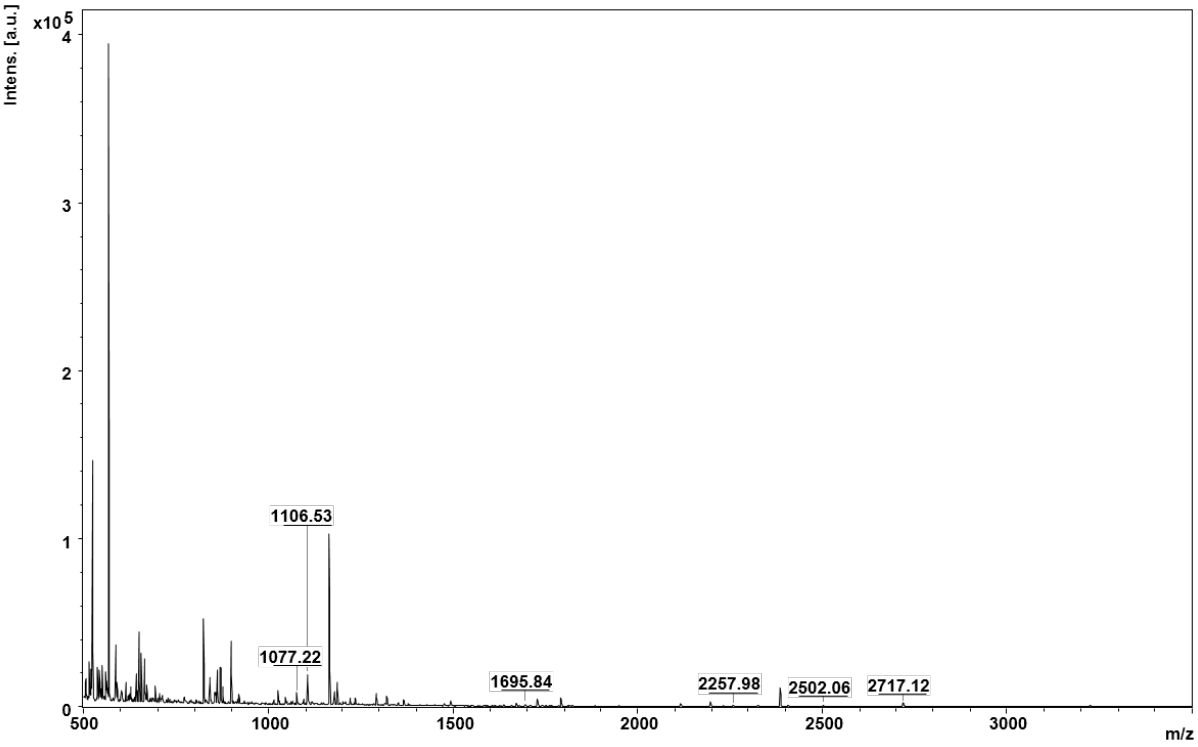


Figure 63: Mass spectrum of sample 1C of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

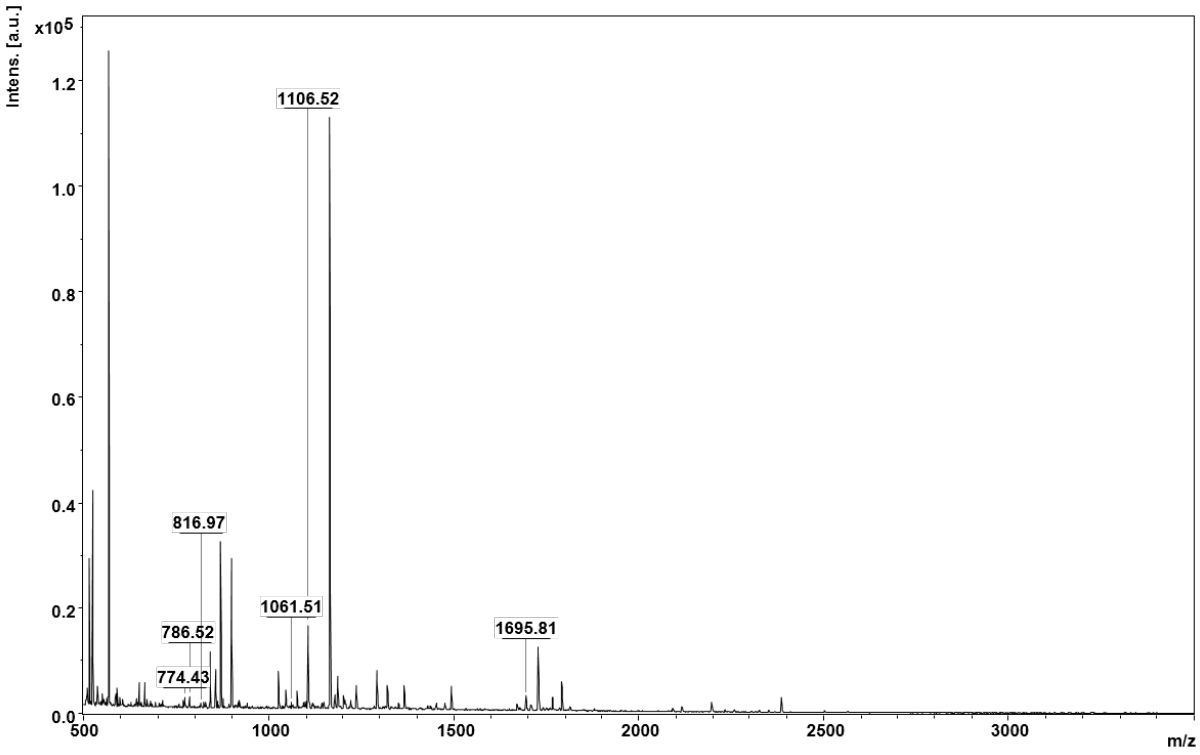


Figure 64: Mass spectrum of sample 1DE of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

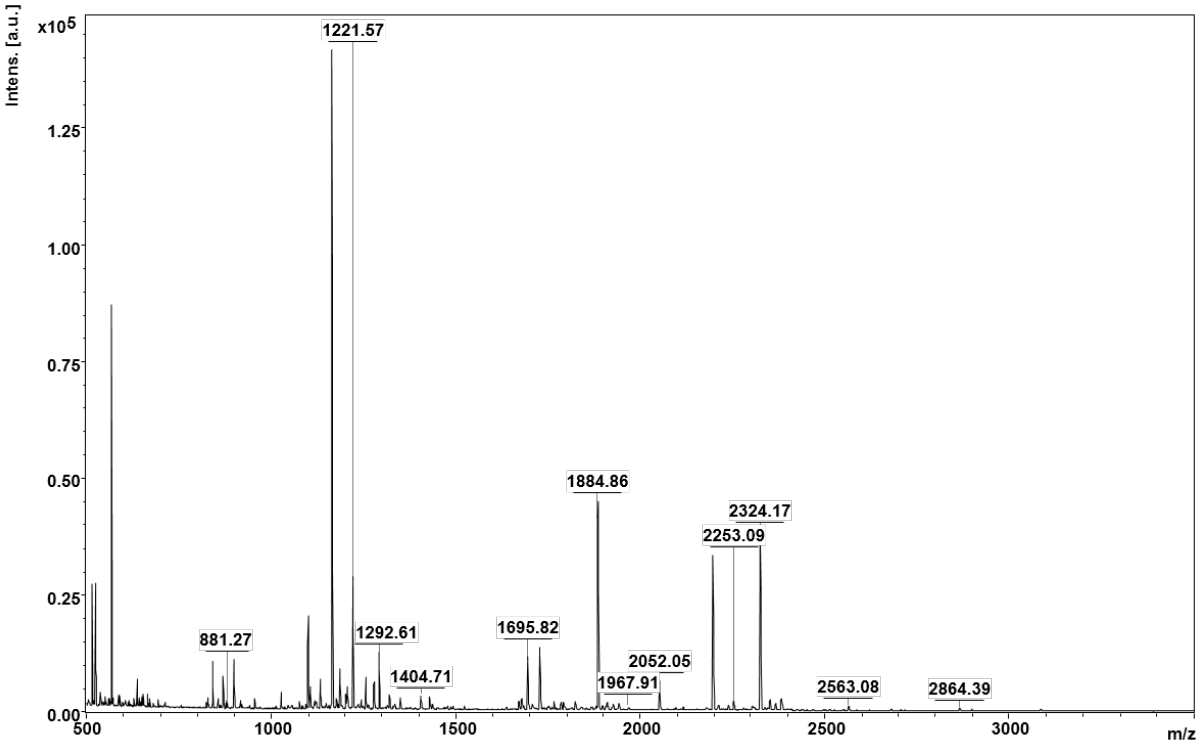


Figure 65: Mass spectrum of sample 1FGH of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

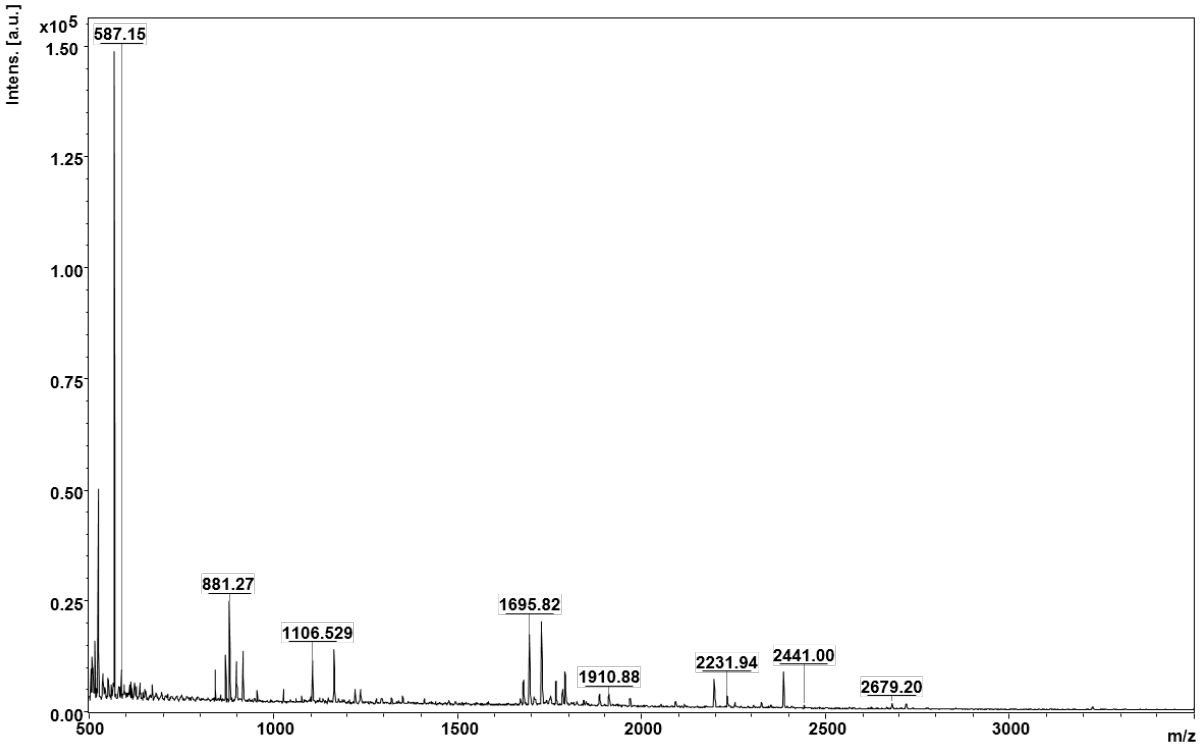


Figure 66: Mass spectrum of sample 1A of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

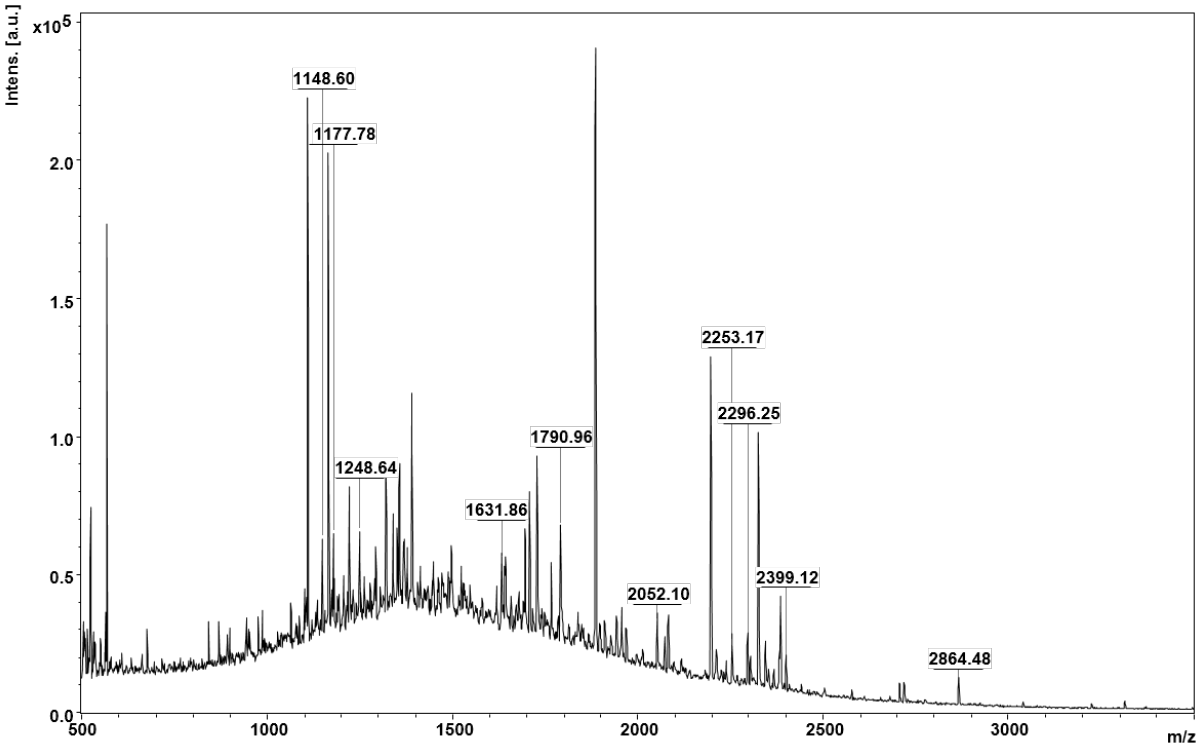


Figure 67: Mass spectrum of sample 1BC of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

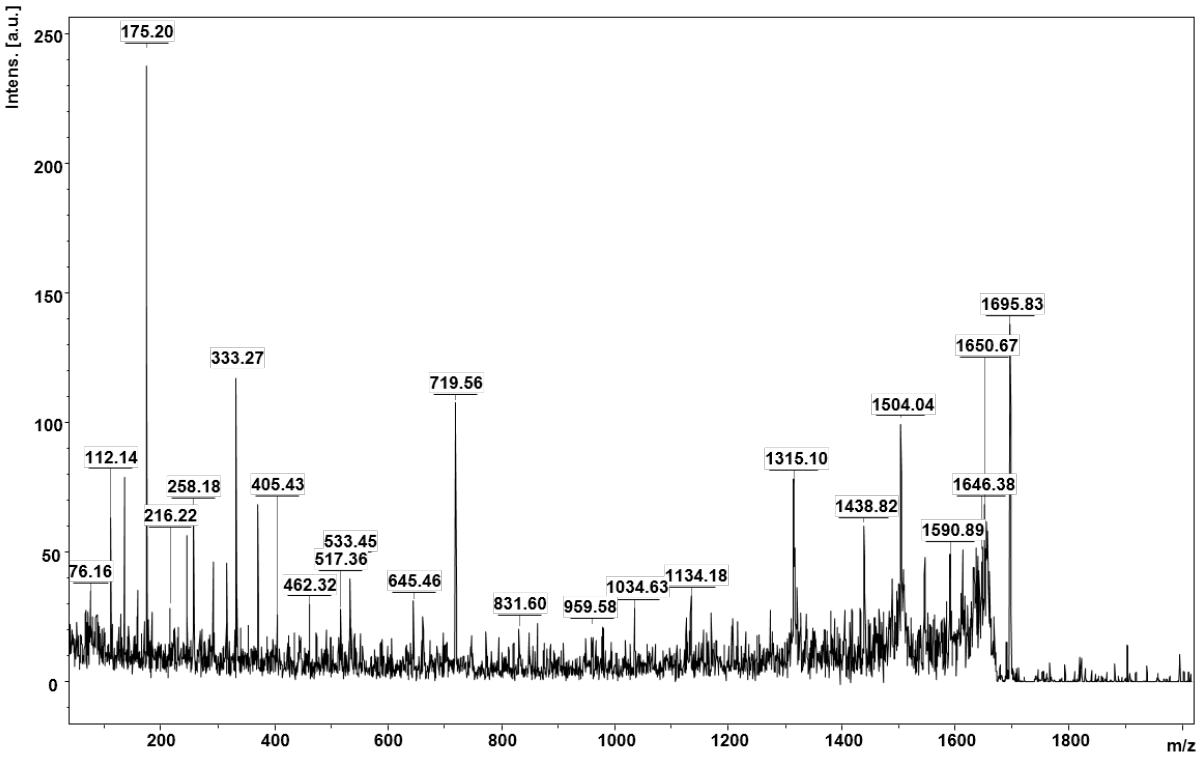


Figure 68: MS/MS spectrum of sample 1A_1695 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

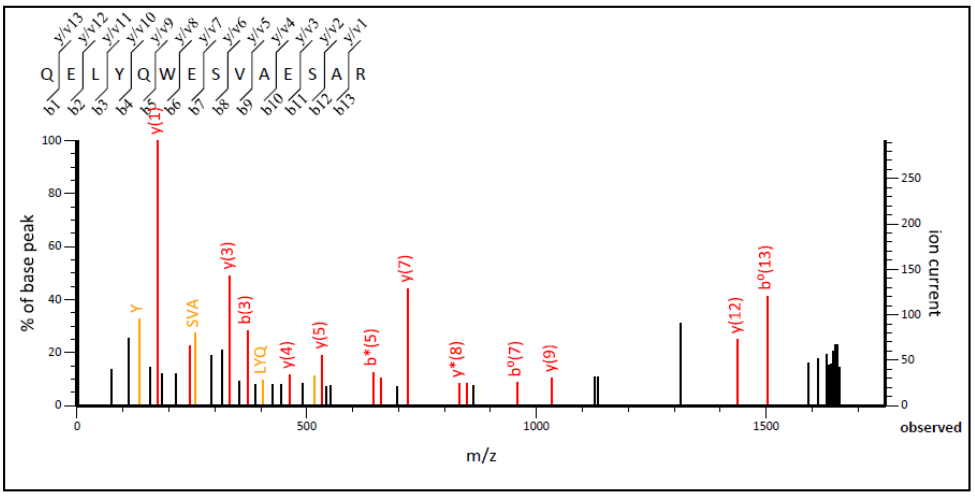


Figure 69: Peptide fragmentation and sequence of sample 1A_1695. The sequence is displayed in the left top corner

The results of the different samples of Gel Band 1 are listed in table 36. Due to the proteins being isolated and extracted from *M. domestica*, only results from this species were considered for the protein identification.

Sample 1A_1695 could be identified as alpha-xylosidase 1-like from *M. domestica* in Viridiplantae as well as in Rosaceae. In addition, the search in Rosaceae yielded the match alpha-xylosidase 1 from *M. domestica*. The peptide was listed as a unique peptide, which makes the result very reliable. Unique

peptides are specific to only a certain protein or protein variation, i.e. proteoform, in different related species.

The theoretical molecular weight of both proteins fit to the rough estimation of the molecular weight (95 kDa) from the gels.

The analyzed Gel band 1 could also consist of more than 1 protein, but due to the MS^2 spectrum of 1A_1695 it can be certainly be concluded that the one protein, or one of the proteins, is alpha-xylosidase 1-like or alpha-xylosidase 1 from *M. domestica*.

Table 36: Mascot search results of Gel Band 1 for all samples. Only significant hits with relevant theoretical molecular weights, compared to the rough estimate of the gel bands, are listed.

Sample	Taxonomy	Name of protein and Species	Matches	Score/ Thres-hold	th. MW [Da]	th. pI	Protein coverage [%]
1A_1695	Rosaceae	alpha-xylosidase 1-like [*M. domestica*] XP_008368212.1	1/1	71/46	103503	5.51	1
1A_1695	Rosaceae	alpha-xylosidase 1 [*M. domestica*] XP_008368212.1	1/1	71/58	103779	5.93	1
1A_1695	Viridiplantae	PREDICTED: alpha-xylosidase 1-like [*M. domestica*] XP_008368212.1	1/1	71/46	103616	5.51	1

4.5.2 Gel Band 2

In the figures 70-76 the mass spectra of the filtered peptide peaks are shown. The peaks used for the search in Mascot are listed in the Appendix (see Chapter 7.1.2). Additionally, figures 77, 79, 81, 83, 85 display the MS/MS spectra, which give more input on the single peptides themselves. The peptide sequence and the fragmentation patterns of 2A_1409, 2A_1511, 2C_1511, 2A*_1409 and 2A*_1511 are visible in figures 78, 80, 82, 84, 86.

The mass spectra show similar peak patterns along the displayed range, except for a few peaks > 2500 m/z, which will give reliable results for the identification.

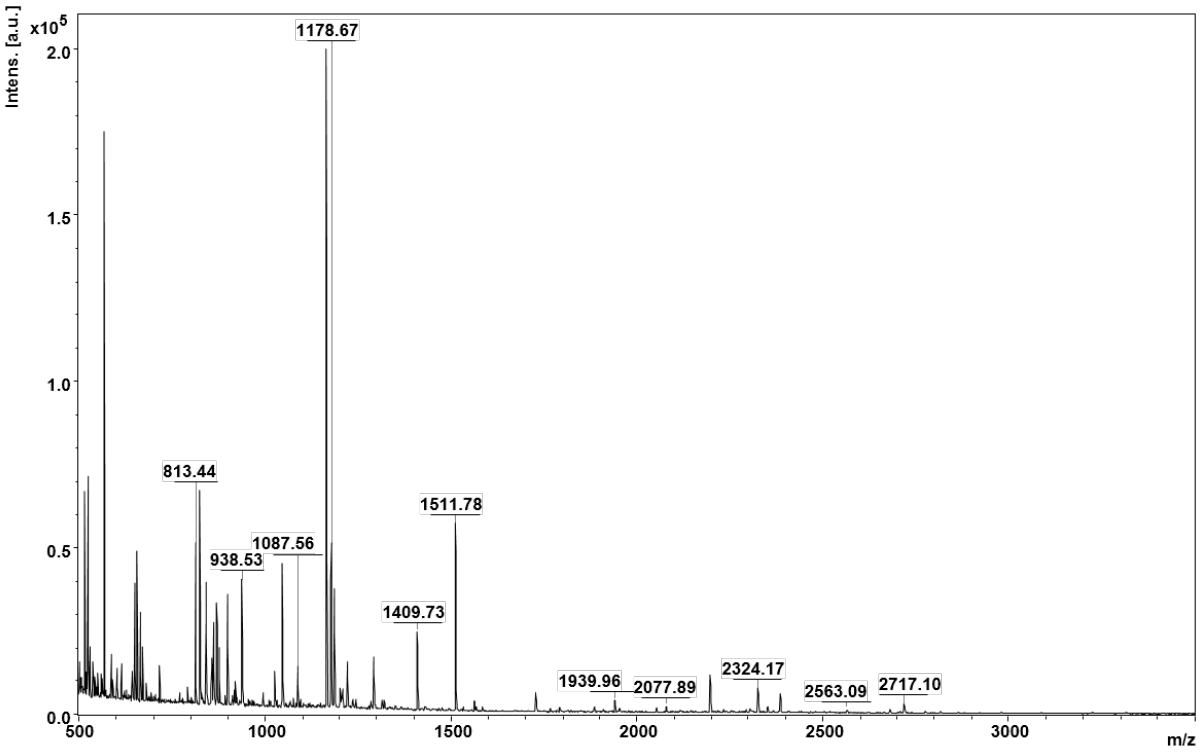


Figure 70: Mass spectrum of sample 2A of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

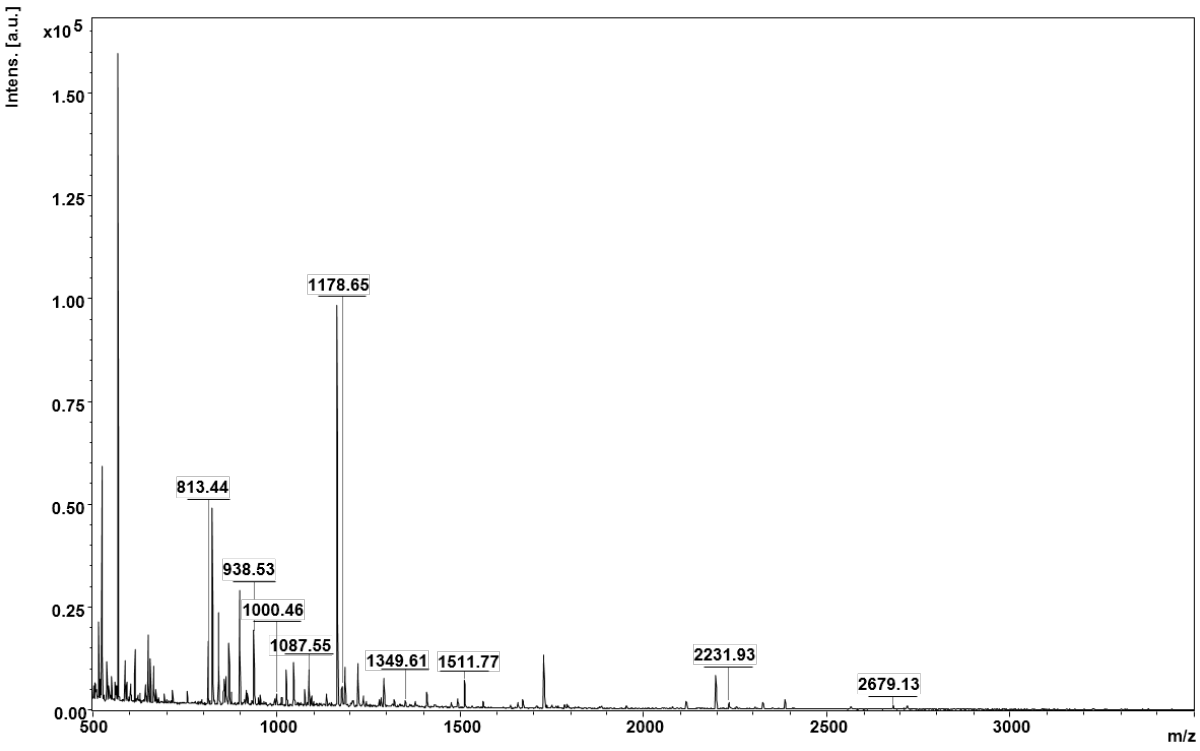


Figure 71: Mass spectrum of sample 2B of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

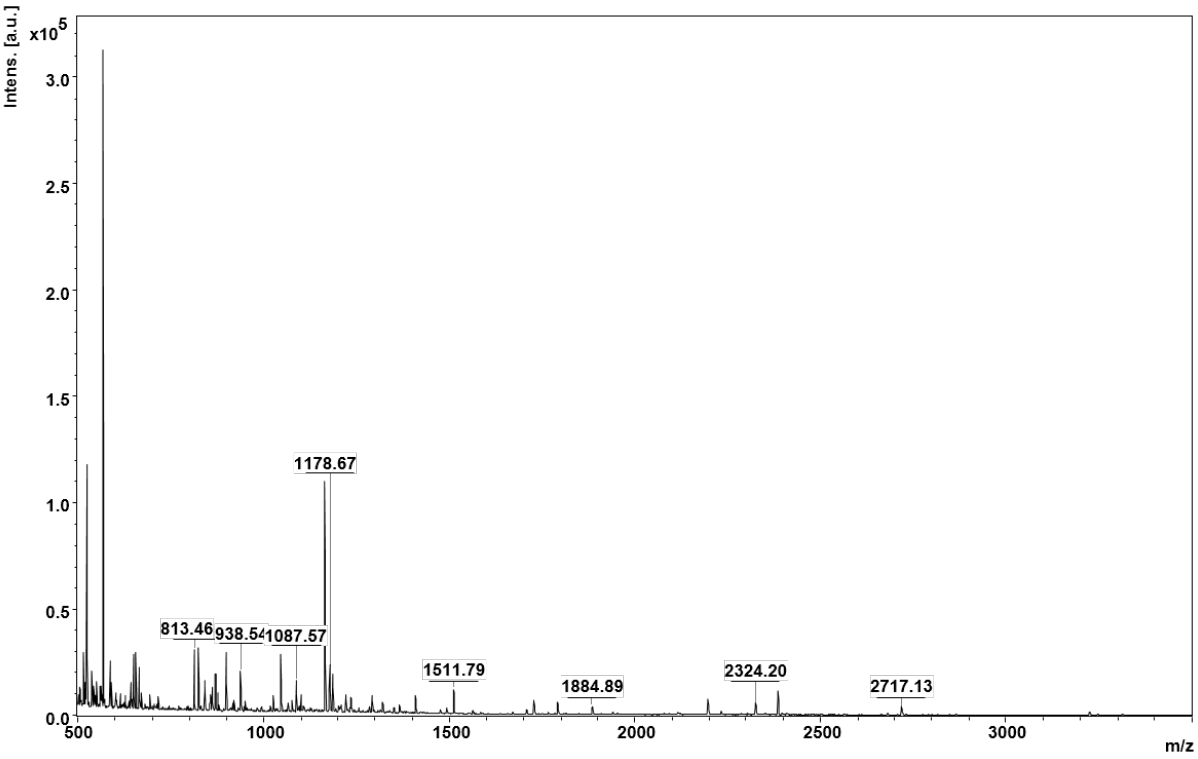


Figure 72: Mass spectrum of sample 2C of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

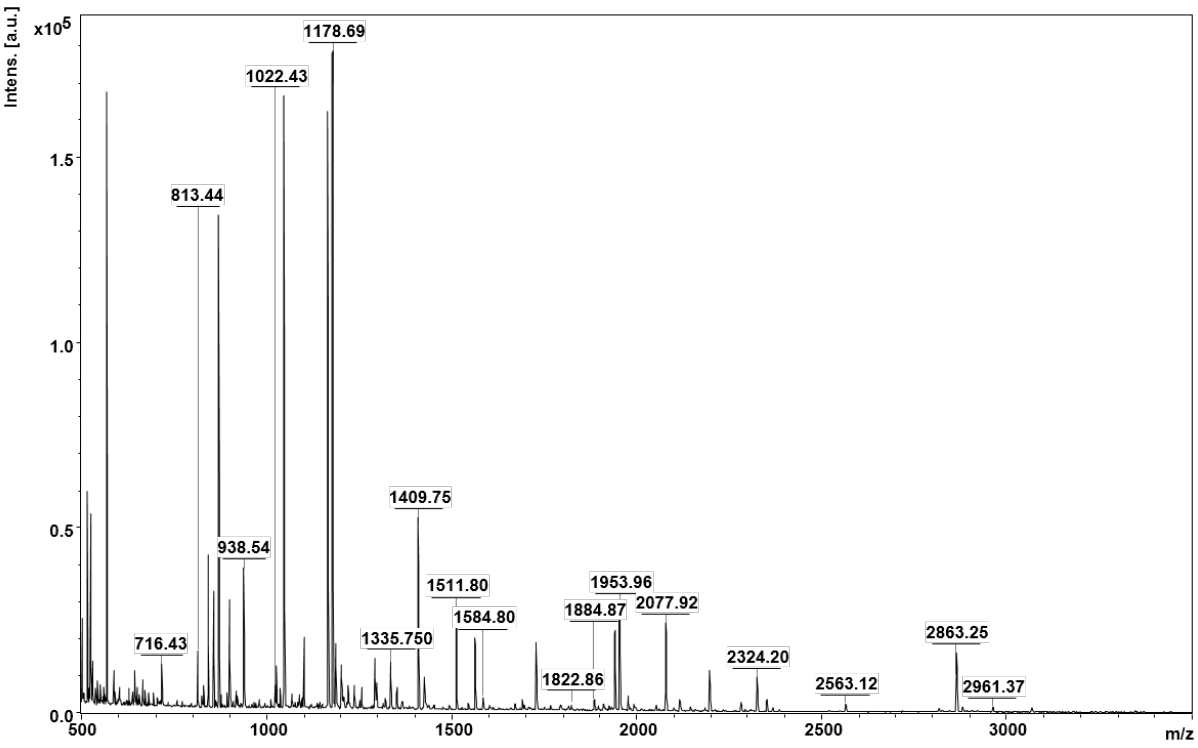


Figure 73: Mass spectrum of sample 2DE of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

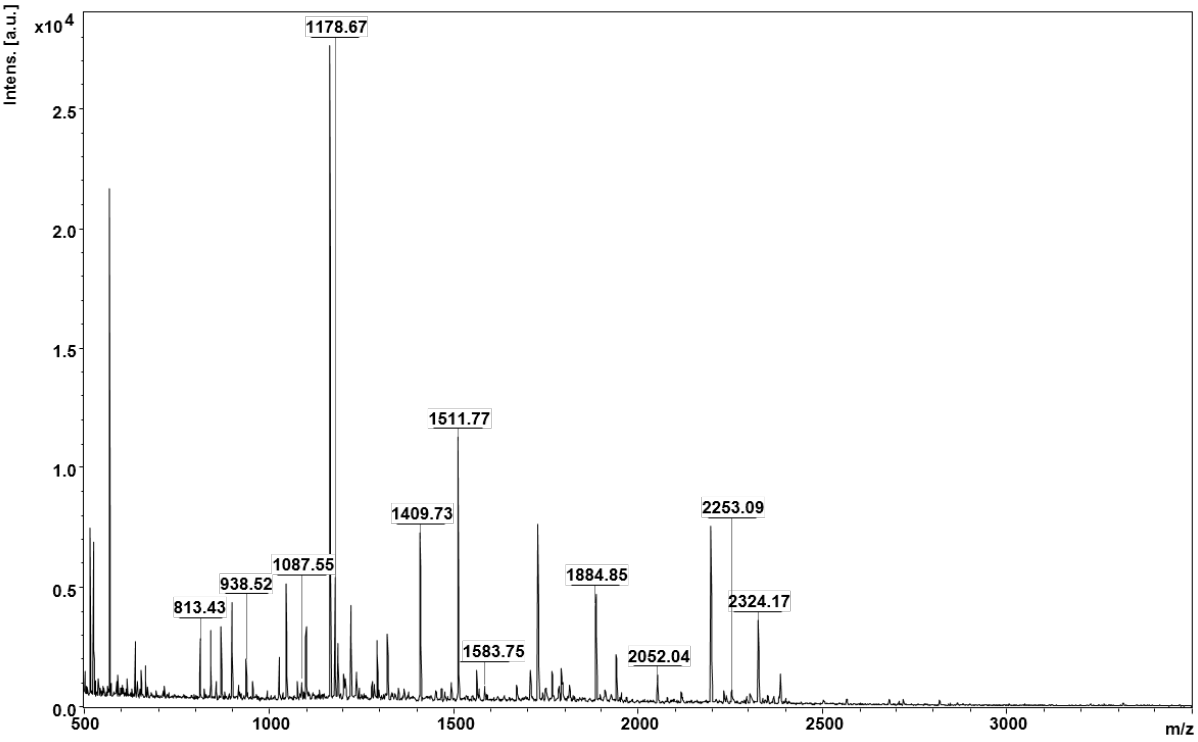


Figure 74: Mass spectrum of sample 2FGH of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

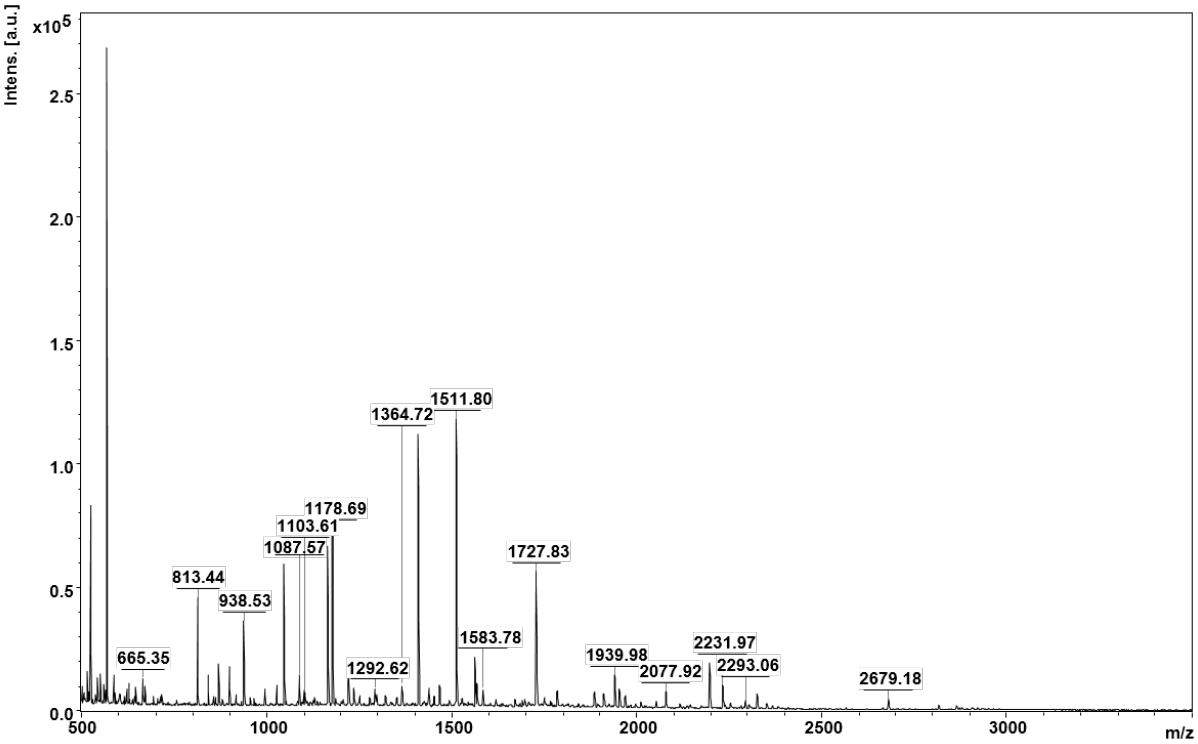


Figure 75: Mass spectrum of sample 2A of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

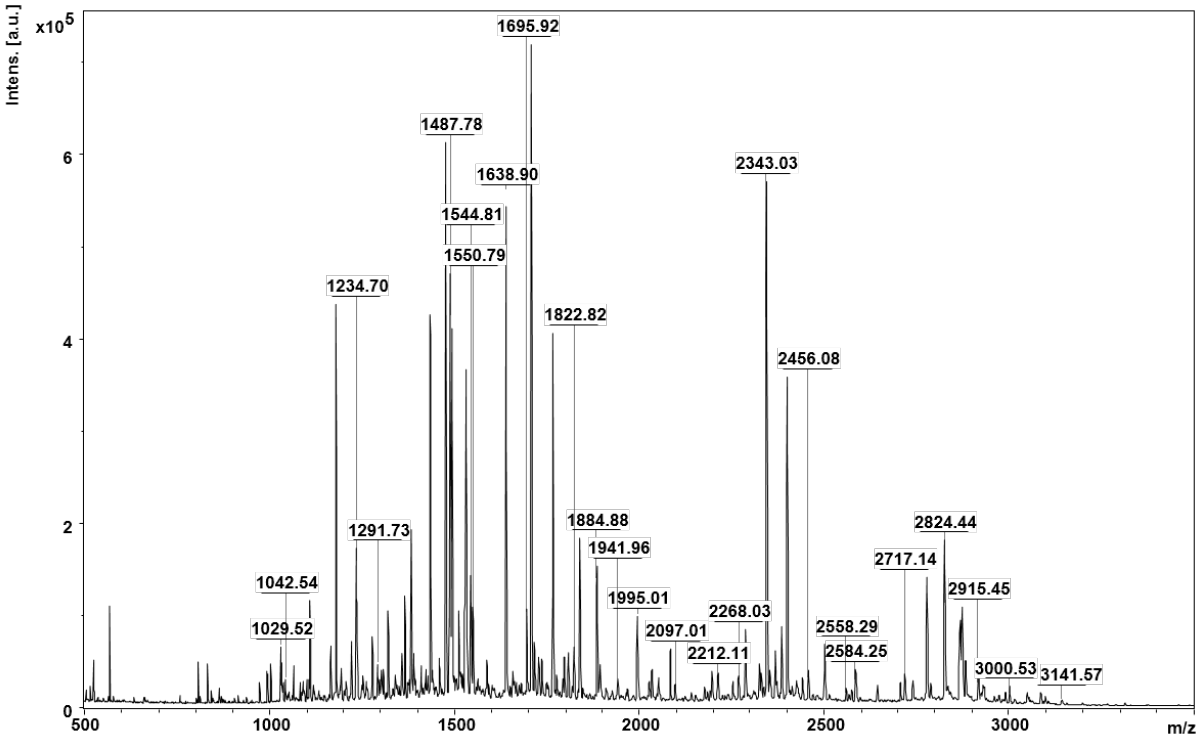


Figure 76: Mass spectrum of sample 2BC of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

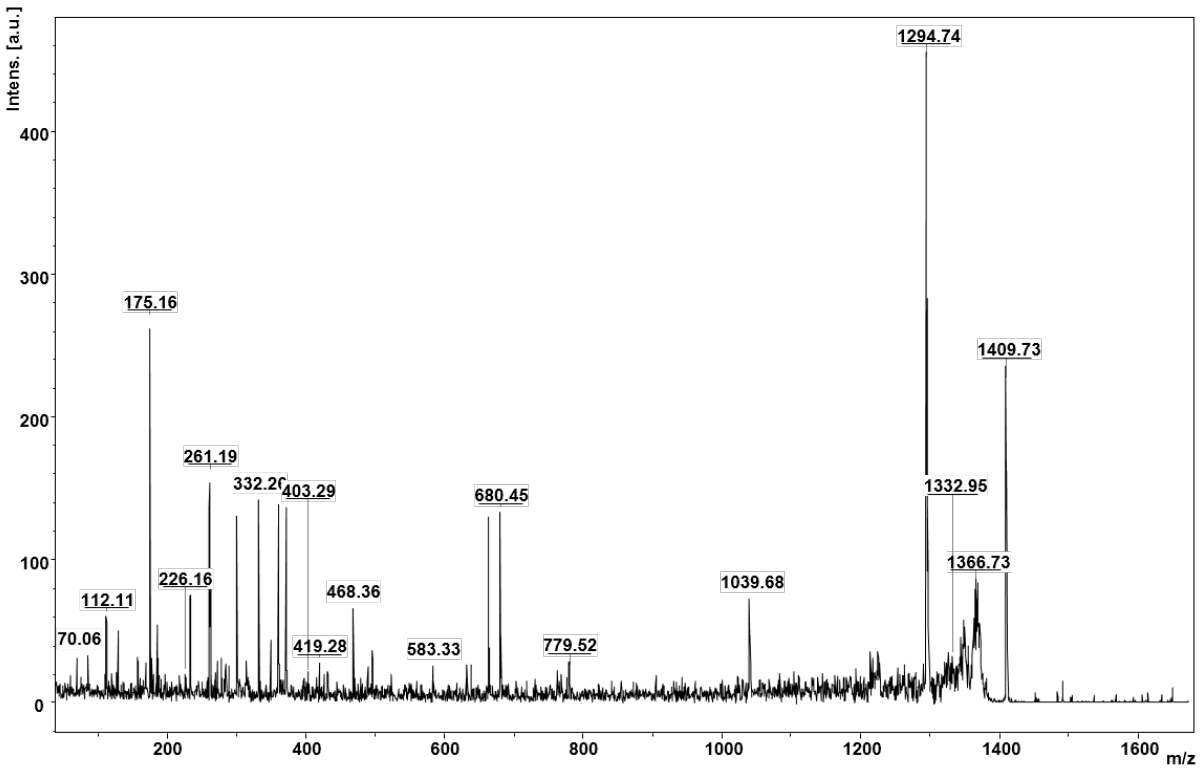


Figure 77: MS/MS spectrum of sample 2A_1409 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

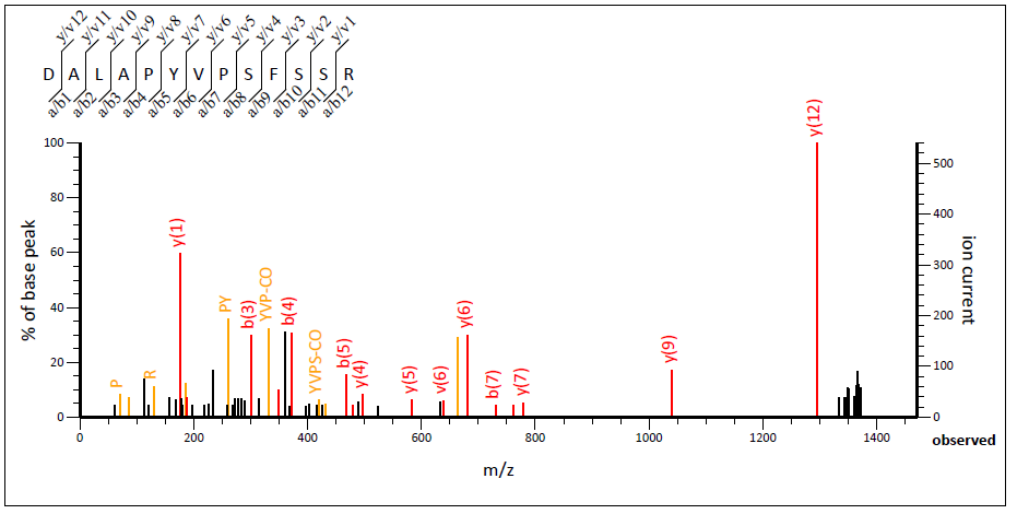


Figure 78: Peptide fragmentation and sequence of sample 2A_1409. The sequence is displayed in the left top corner

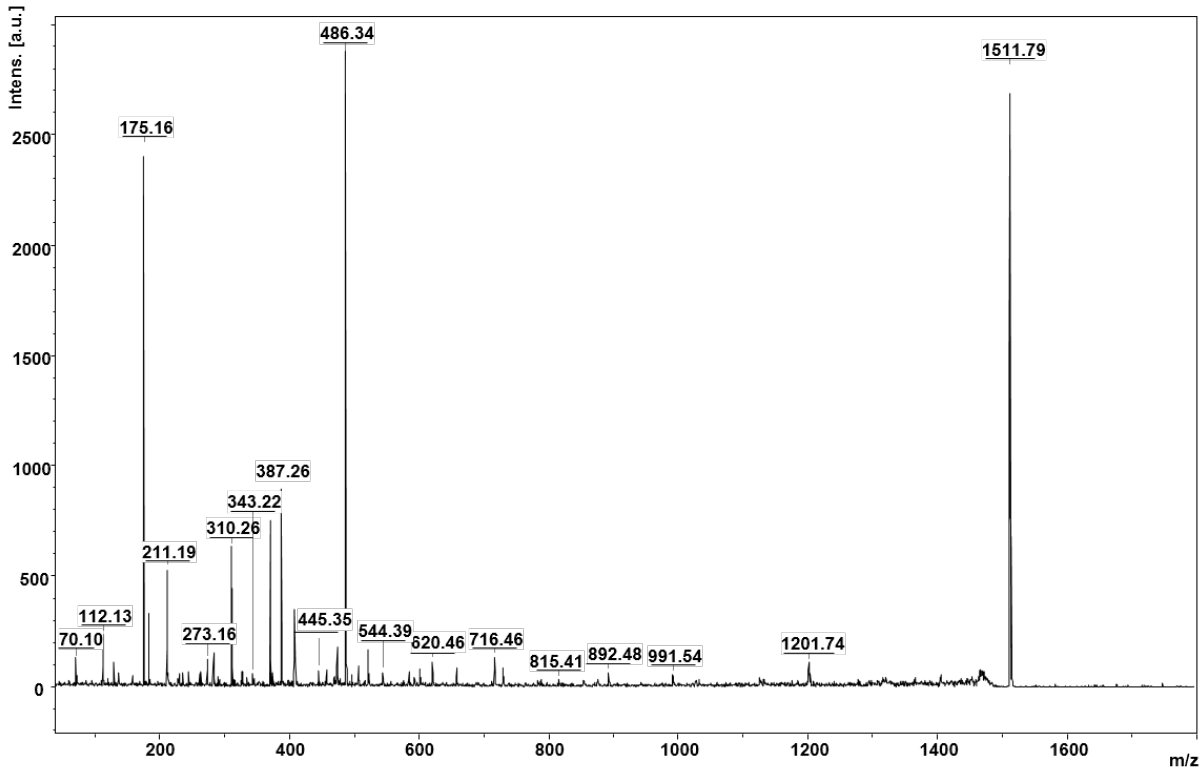


Figure 79: MS/MS spectrum of sample 2A_1511 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

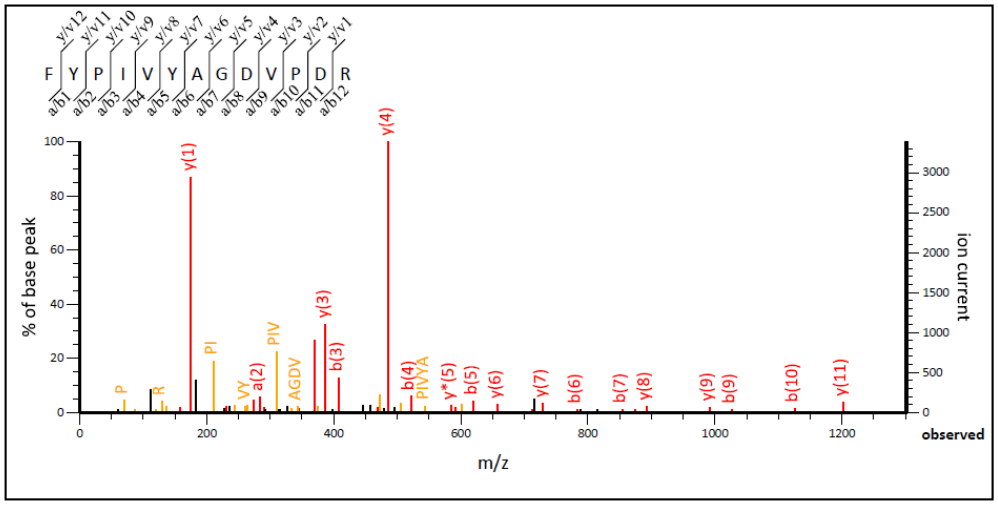


Figure 80: Peptide fragmentation and sequence of sample 2A_1511. The sequence is displayed in the left top corner.

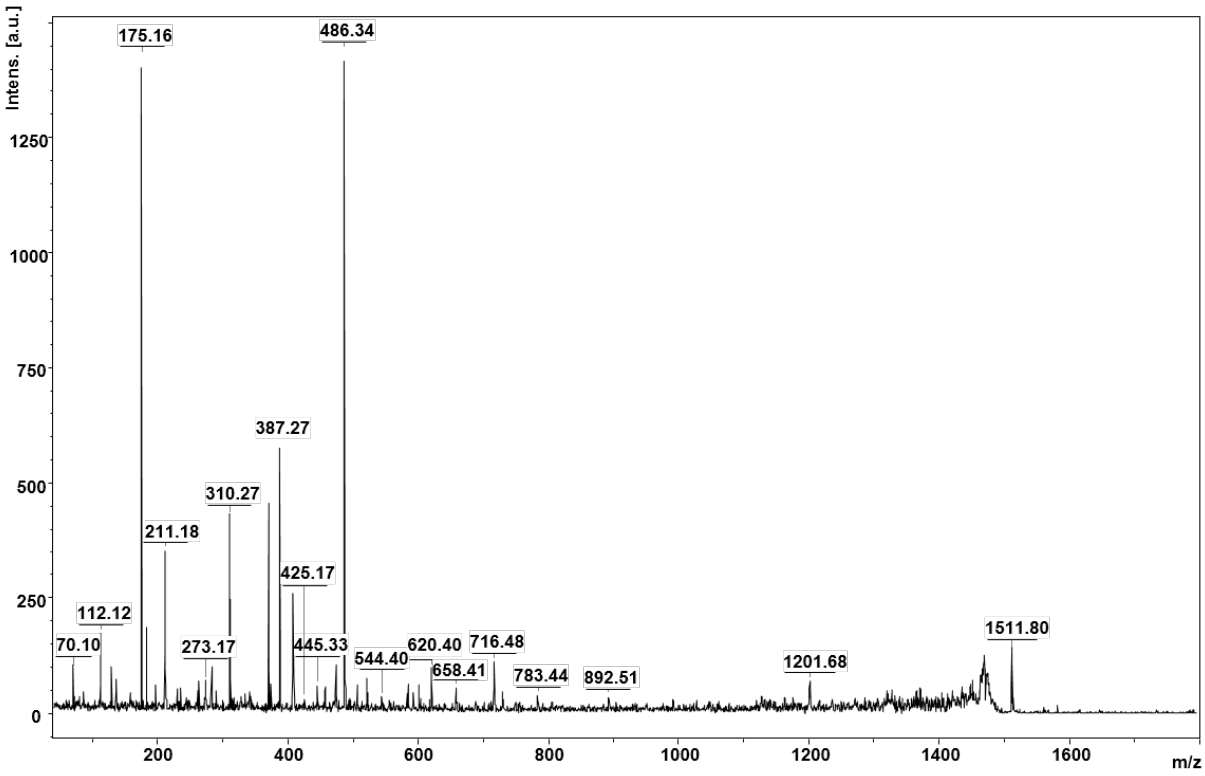


Figure 81: MS/MS spectrum of sample 2C_1511 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

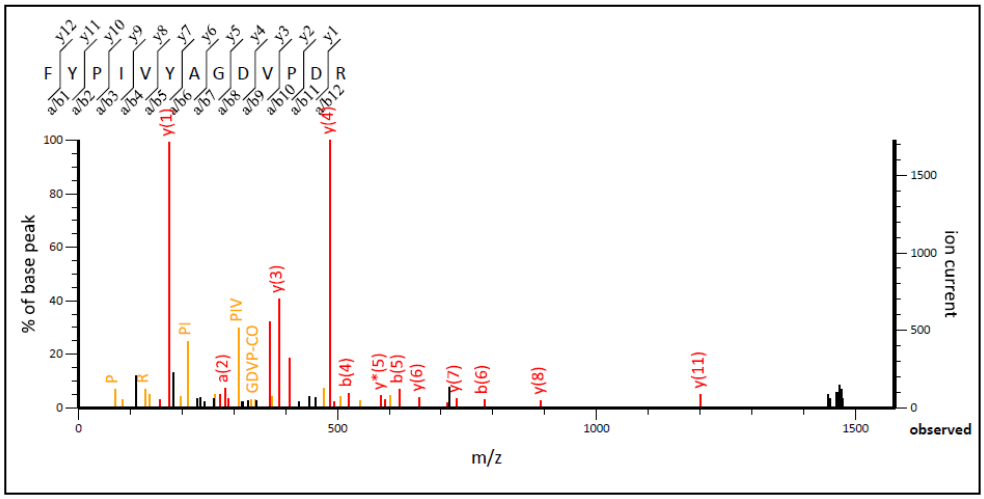


Figure 82: Peptide fragmentation and sequence of sample 2C_1511. The sequence is displayed in the left top corner.

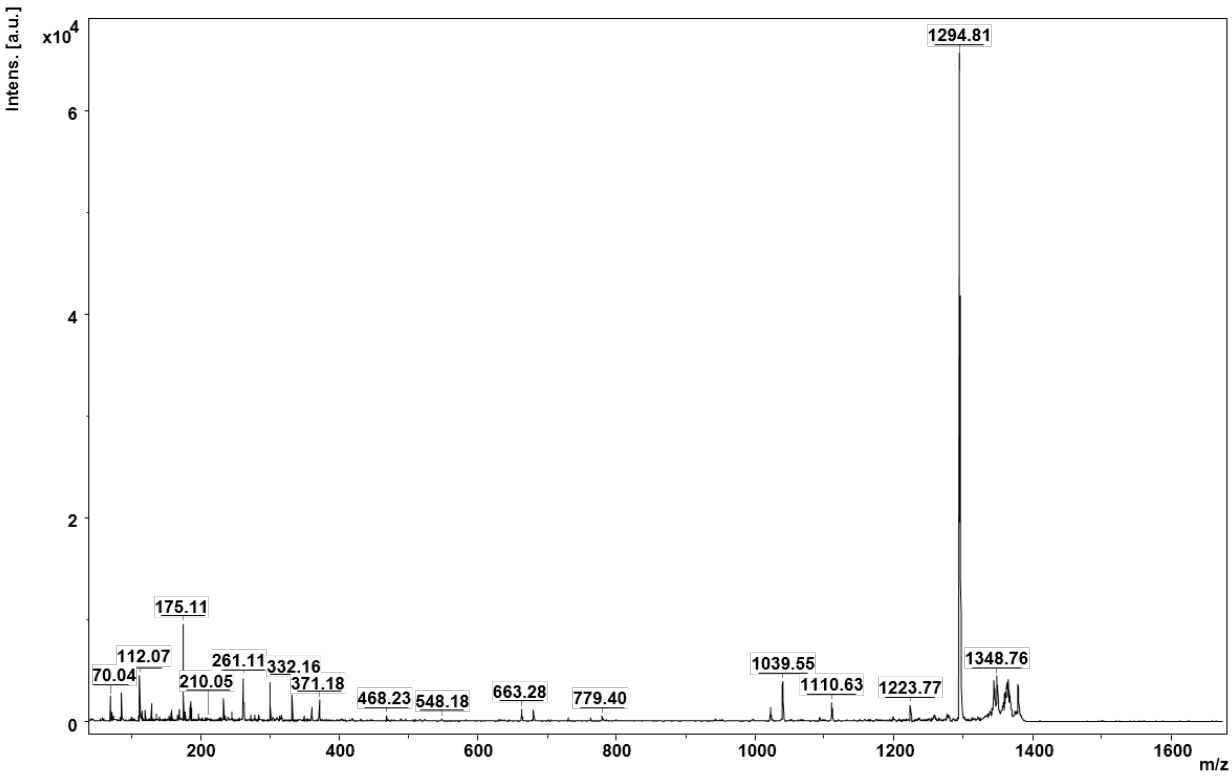


*Figure 83: MS/MS spectrum of sample 2A*_1409 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

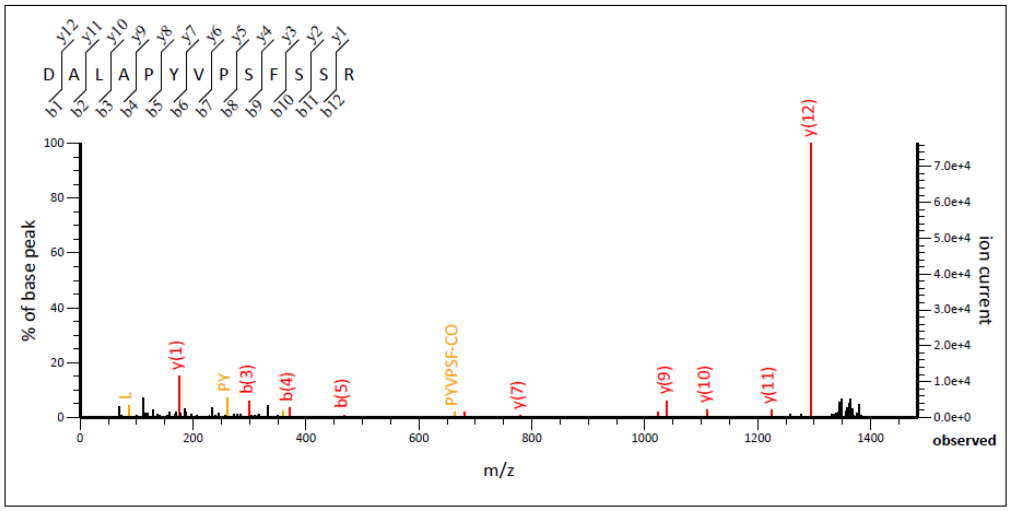


*Figure 84: Peptide fragmentation and sequence of sample 2A*_1409. The sequence is displayed in the left top corner.*

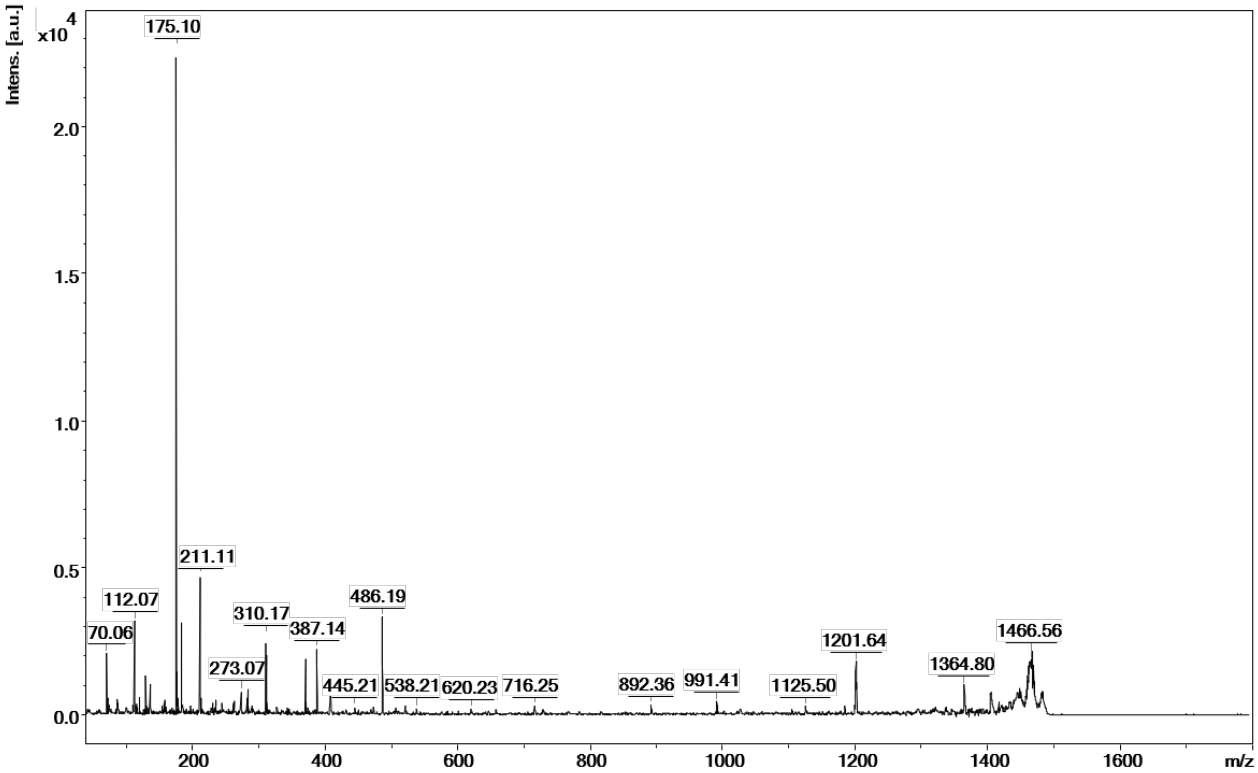


*Figure 85: MS/MS spectrum of sample 2A*_1511 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

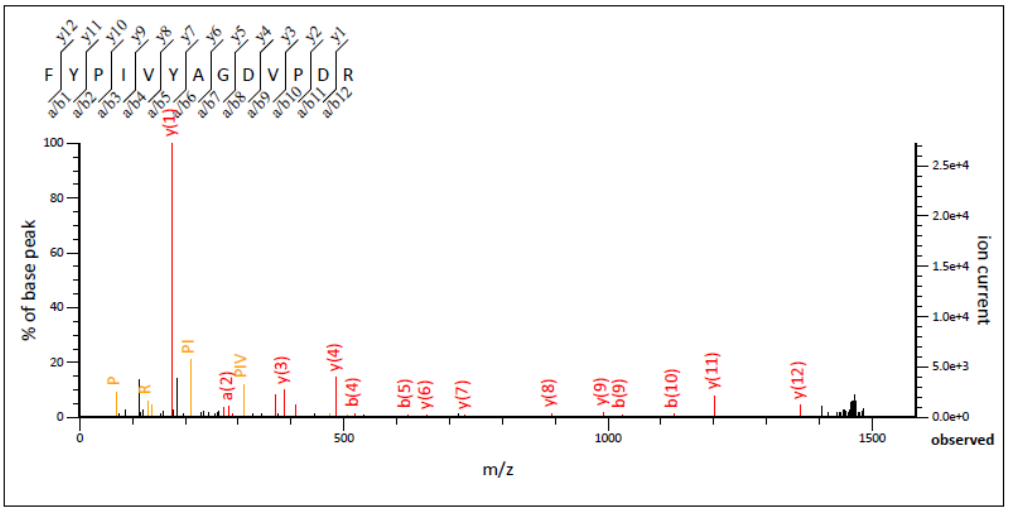


*Figure 86: Peptide fragmentation and sequence of sample 2A*_1511. The sequence is displayed in the left top corner.*

The Mascot search results are listed in table 37. Only mass spectra with a relevant theoretical weight, similar to the estimation from the gel band (72 kDa) are shown. Due to the proteins being isolated and extracted from *M. domestica*, only results from this species were considered for the protein identification.

The MS^2 spectra of 2A_1409, 2A_1511, 2B_1511, 2C_1511, 2A*_1409 and 2A*_1511 all identified the protein to be cucumisin-like isoform X1, cucumisin-like isoform X2 or cucumisin-like (partial), which are classified as unique peptides and therefore make the result very reliable.

The analysed Gel band 2 more than likely consists of multiple proteins, but cucumisin-like isoform X1 or X2 or partial from *M. domestica* can be considered as a correct result with 95 % probability.

Table 37: Mascot search results of Gel Band 2 for all samples. Only significant hits with relevant theoretical molecular weights, compared to the rough estimate of the gel bands, are listed.

Sample	Taxonomy	Name of protein and Species	Matches	Score/ Thres-hold	th. MW [Da]	th. pI	Protein coverage [%]
2A_1409	Rosaceae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	67/46	79768	6.76	1
2A_1409	Rosaceae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	67/46	76338	6.74	1
2A_1409	Rosaceae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	67/46	64987	8.39	2

2A_1511	Rosaceae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	85/46	79768	6.76	1
2A_1511	Rosaceae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	85/46	76338	6.74	1
2A_1511	Rosaceae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	85/46	64987	8.39	2
2B_1511	Rosaceae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	46/46	79768	6.76	1
2B_1511	Rosaceae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	46/46	76338	6.74	1
2B_1511	Rosaceae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	46/46	64987	8.39	2
2C_1511	Rosaceae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	46/46	79768	6.76	1
2C_1511	Rosaceae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	46/46	76338	6.74	1
2C_1511	Rosaceae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	46/46	64987	8.39	2
2A*_1409	Rosaceae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	58/46	79768	6.76	1
2A*_1409	Rosaceae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	58/46	76338	6.74	1
2A*_1409	Rosaceae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	58/46	64987	8.39	2
2A*_1511	Rosaceae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	80/46	79768	6.76	1

2A*_1511	Rosaceae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	80/46	76338	6.74	1
2A*_1511	Rosaceae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	80/46	64987	8.39	2
2A_1409	Viridiplantae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	67/58	79768	6.76	1
2A_1409	Viridiplantae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	67/58	76338	6.74	1
2A_1409	Viridiplantae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	67/58	64987	8.39	2
2A_1511	Viridiplantae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	85/58	79768	6.76	1
2A_1511	Viridiplantae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	85/58	76338	6.74	1
2A_1511	Viridiplantae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	85/58	64987	8.39	2
2C_1511	Viridiplantae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	75/58	79768	6.76	1
2C_1511	Viridiplantae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	75/58	79768	6.76	1
2C_1511	Viridiplantae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	75/58	79768	6.76	1
2A*_1409	Viridiplantae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	58/57	79768	6.76	1

2A*_1409	Viridiplantae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	58/57	76338	6.74	1
2A*_1409	Viridiplantae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	58/57	64987	8.39	2
2A*_1511	Viridiplantae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	80/58	79768	6.76	1
2A*_1511	Viridiplantae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	80/58	76338	6.74	1
2A*_1511	Viridiplantae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	80/58	64987	8.39	2

4.5.3 Gel Band 3

In the figures 87-93 the mass spectra of the filtered peptide peaks are shown. The peaks used for the search in Mascot are listed in the Appendix (see Chapter 7.1.3). Additionally, figures 94, 96, 98, 100, 102 and 104 display the MS/MS spectra, which give more input on the single peptides themselves.

The peptide sequence and the fragmentation patterns of 3A_1192, 3A_1511, 3A_1623, 3B_1192, 3C_1192 and 3C_1673 are visible in figures 95, 97, 99, 101, 103 and 105.

The mass spectra show similar peak behavior along the displayed range, which will give reliable results for the identification.

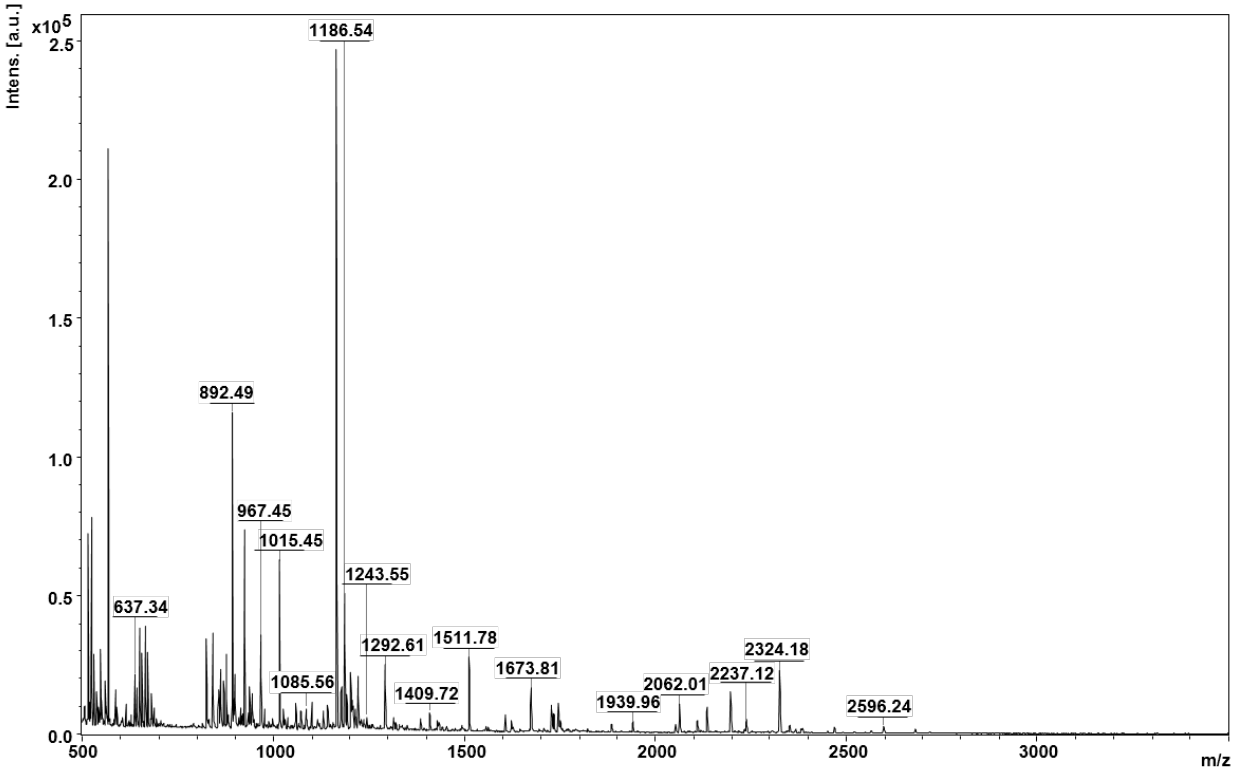


Figure 87: Mass spectrum of sample 3A of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

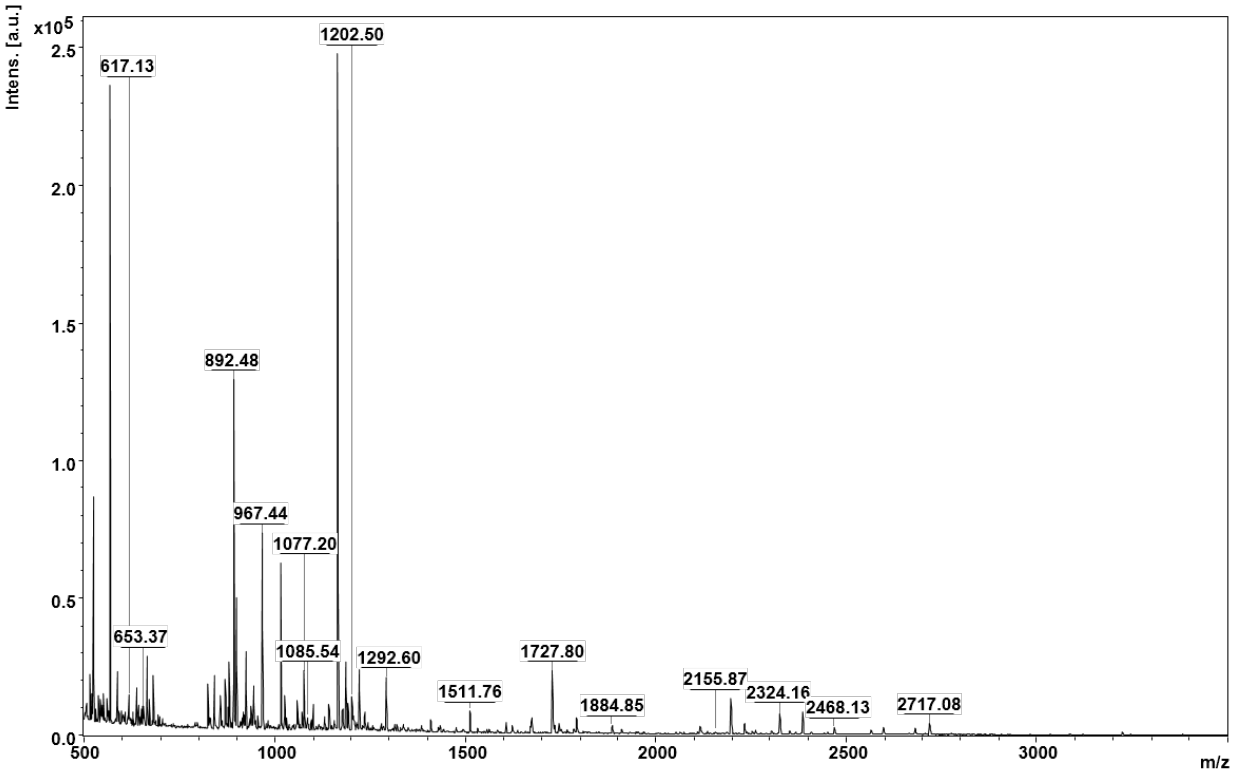


Figure 88: Mass spectrum of sample 3B of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

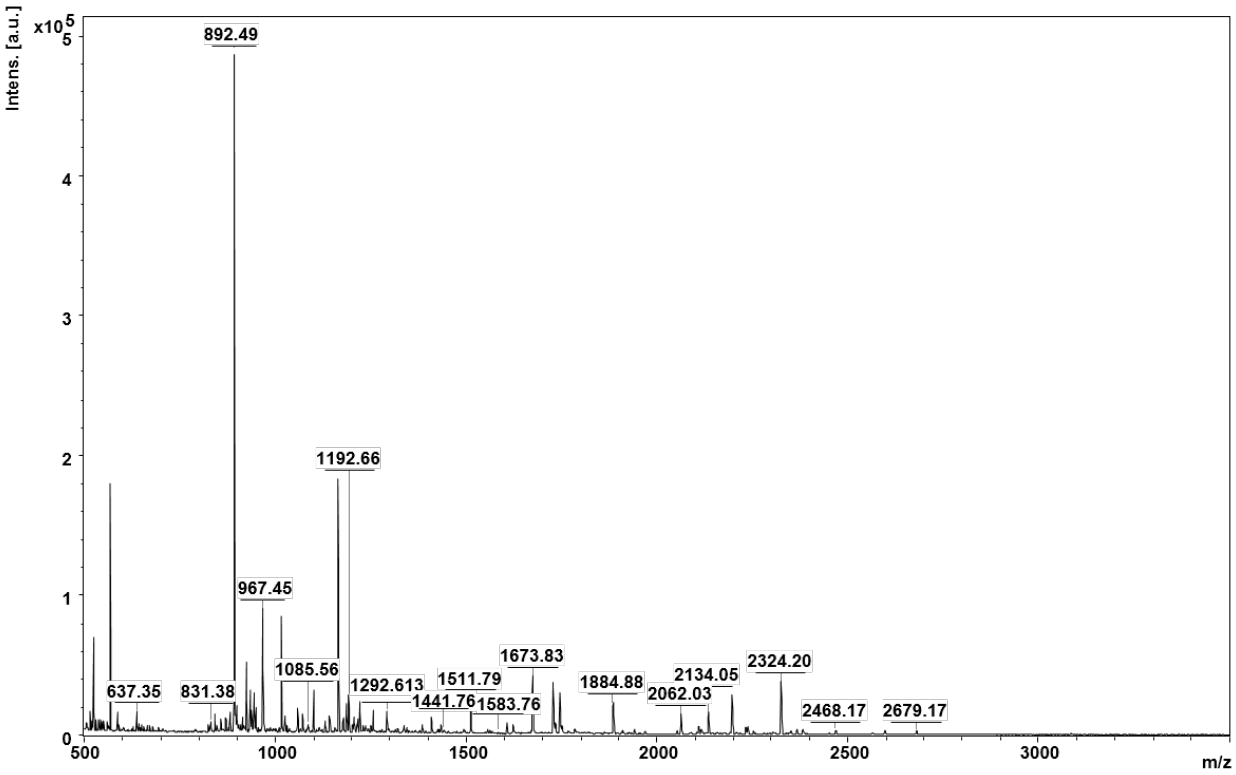


Figure 89: Mass spectrum of sample 3C of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

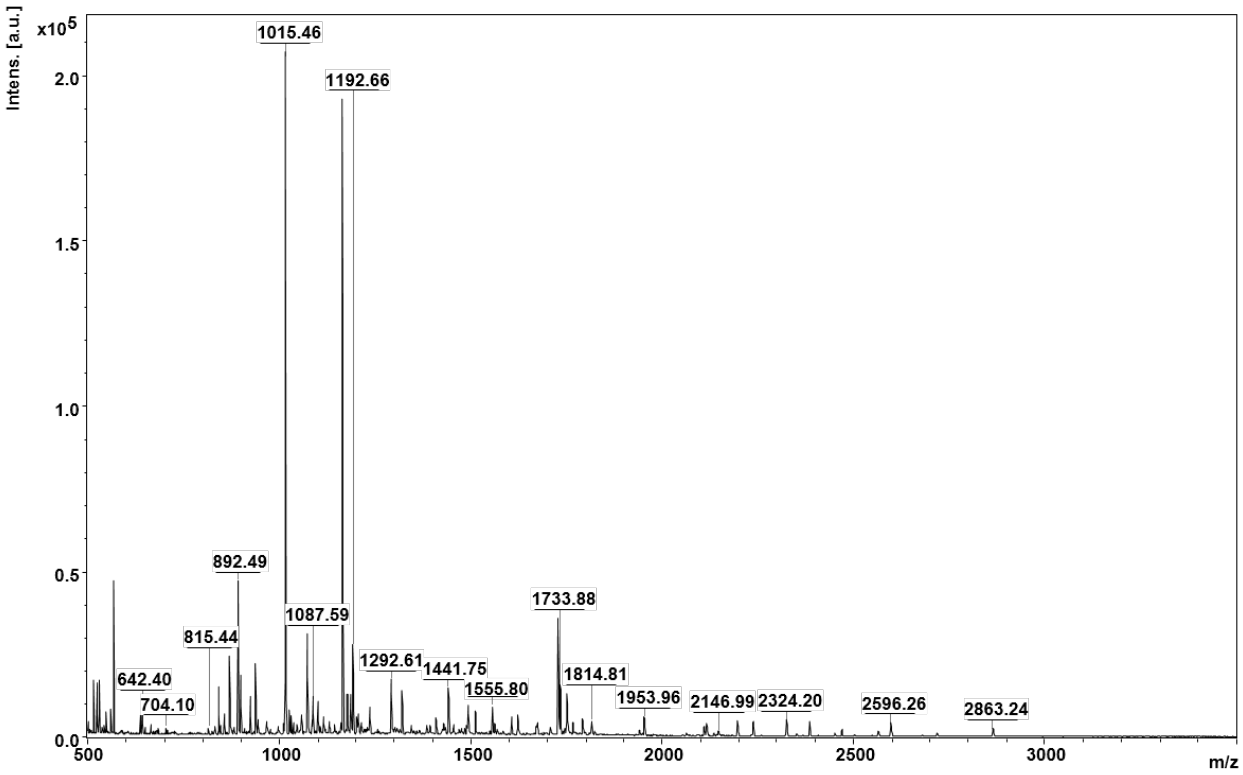


Figure 90: Mass spectrum of sample 3DE of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

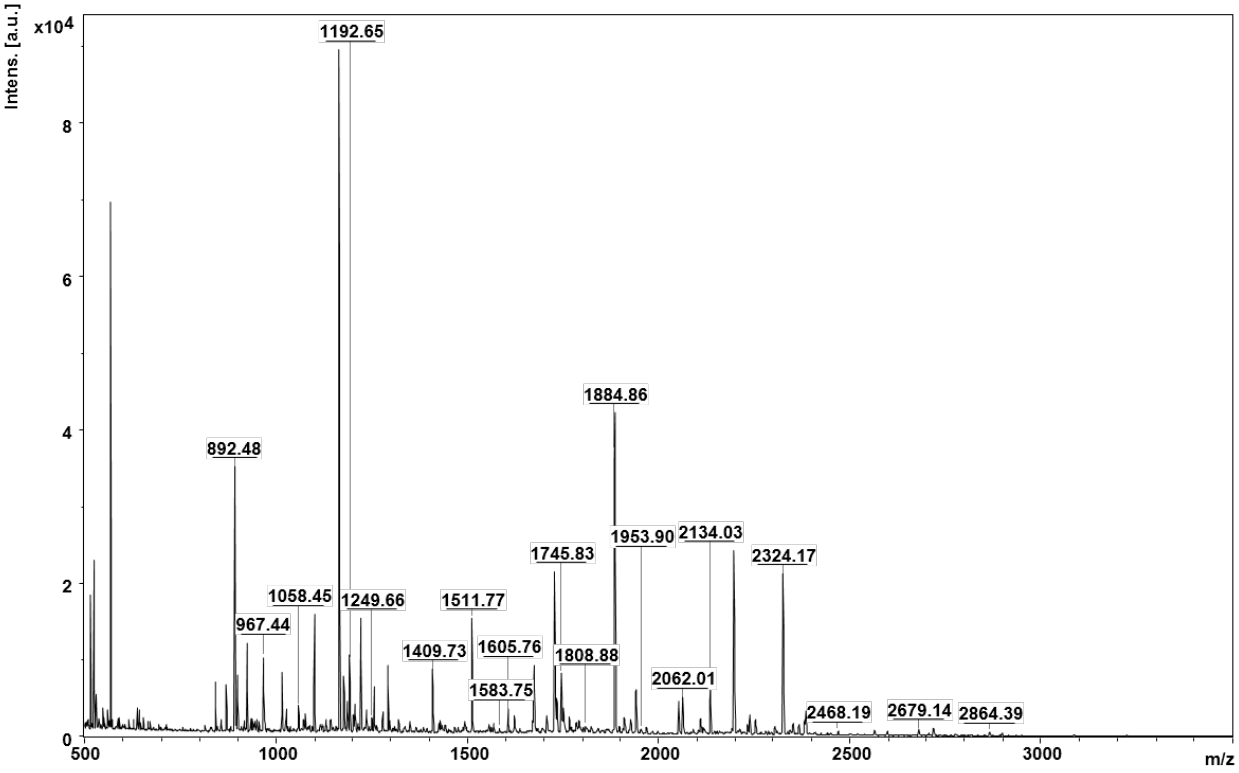


Figure 91: Mass spectrum of sample 3FGH of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

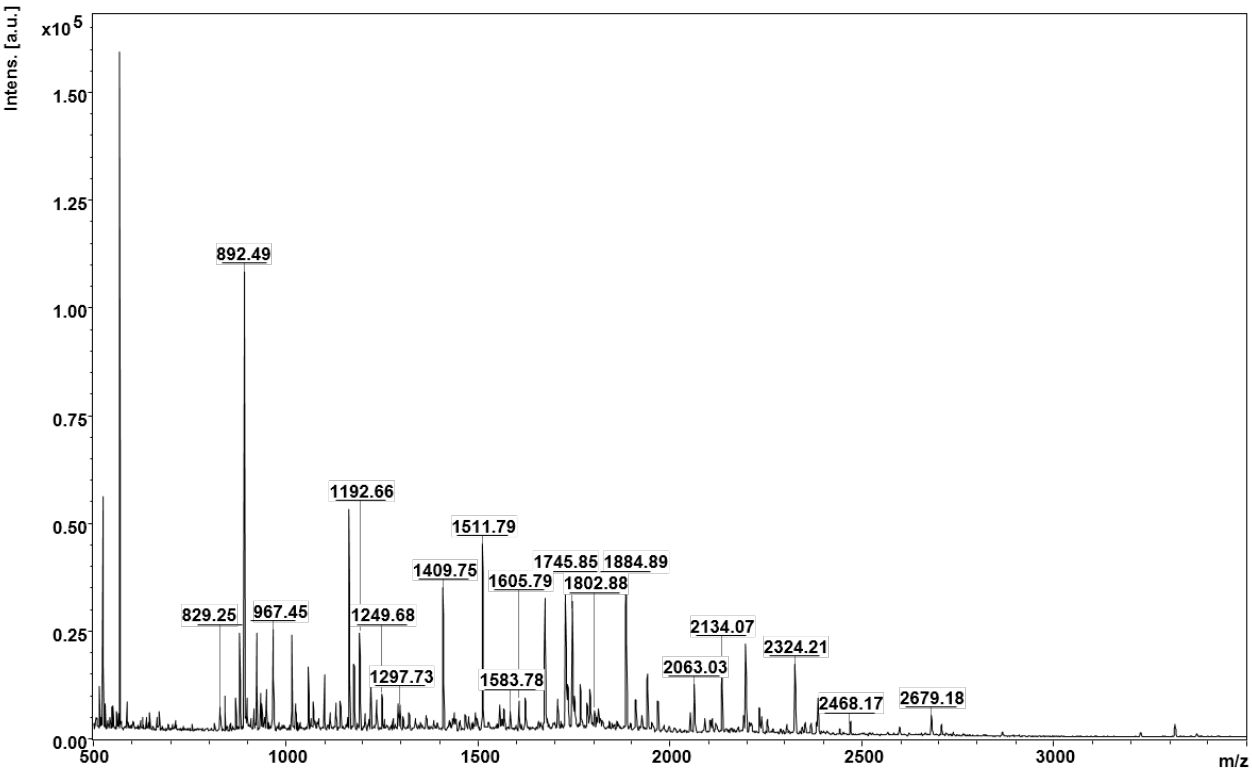


Figure 92: Mass spectrum of sample 3A of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

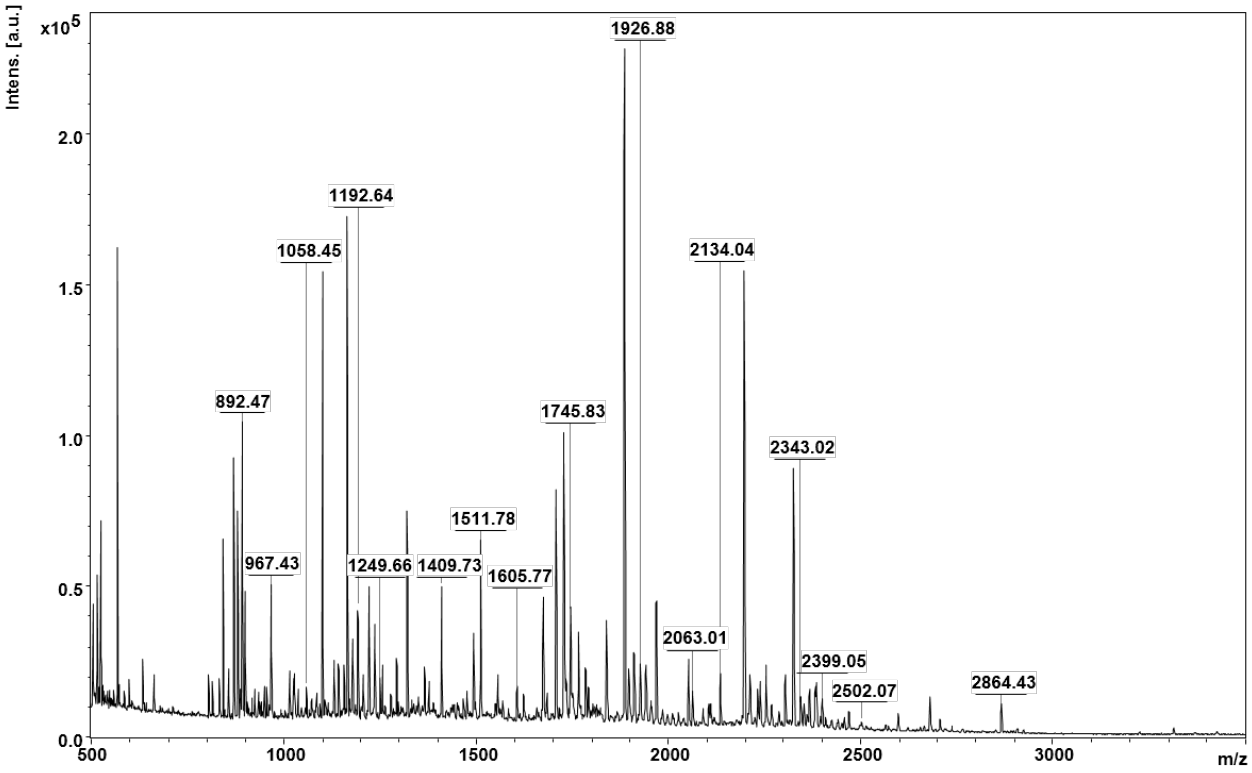


Figure 93: Mass spectrum of sample 3BC of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

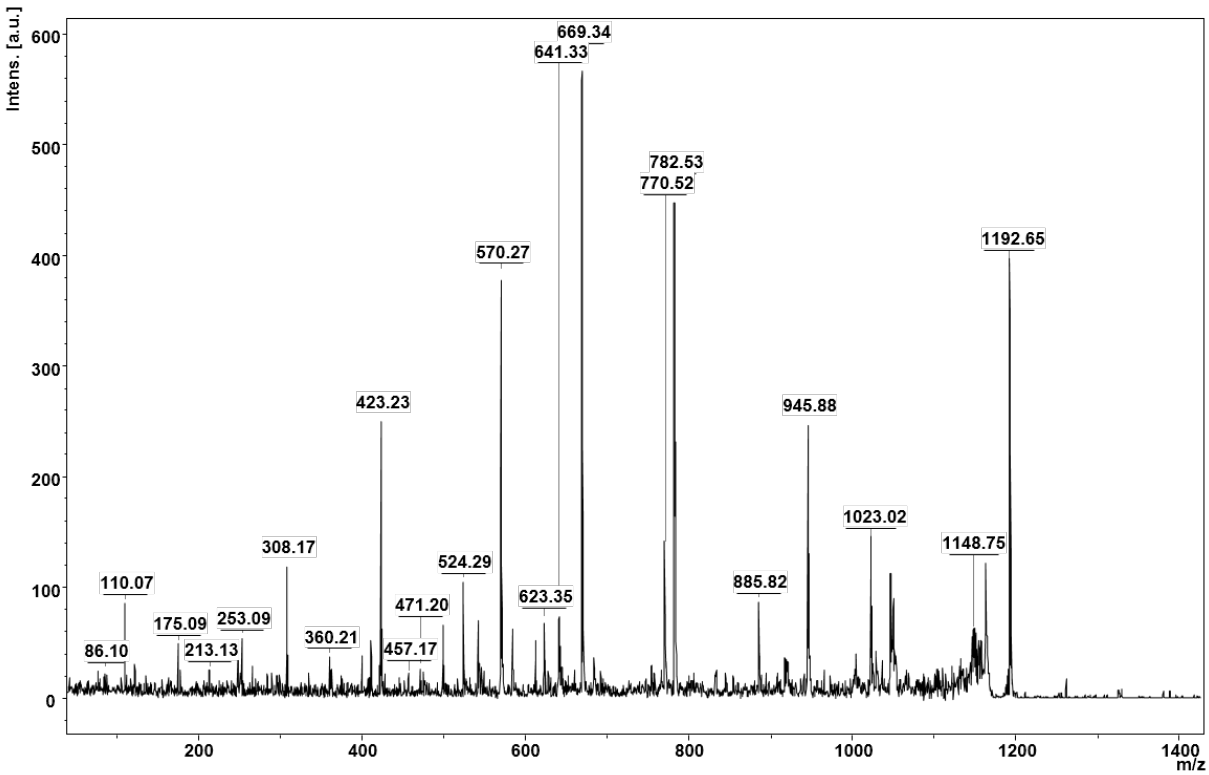


Figure 94: MS/MS spectrum of sample 3A_1192 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

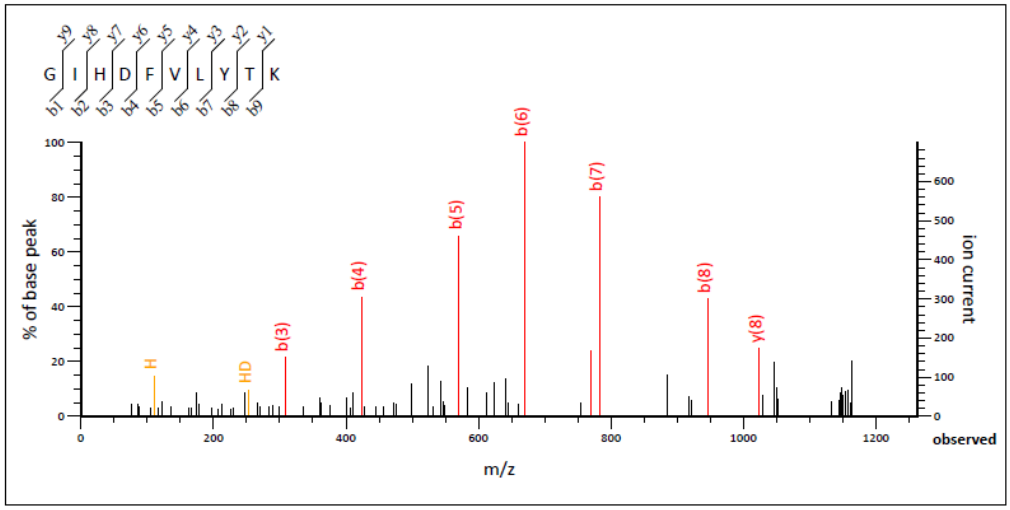


Figure 95: Peptide fragmentation and sequence of sample 3A_1192. The sequence is displayed in the left top corner.

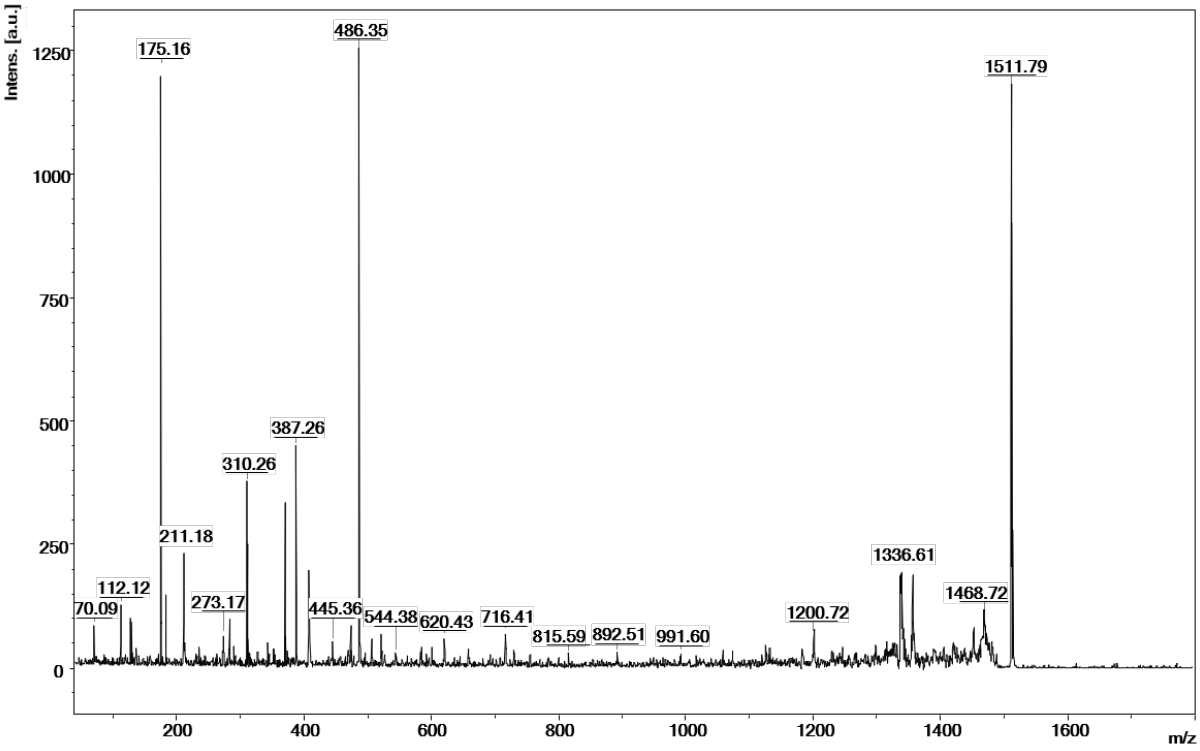


Figure 96: MS/MS spectrum of sample 3A_1511 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

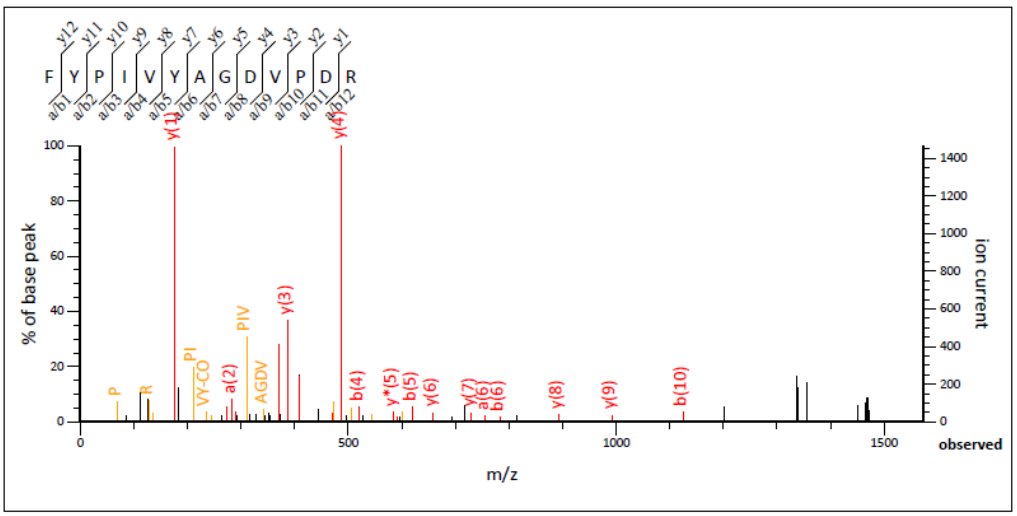


Figure 97: Peptide fragmentation and sequence of sample 3A_1511. The sequence is displayed in the left top corner.

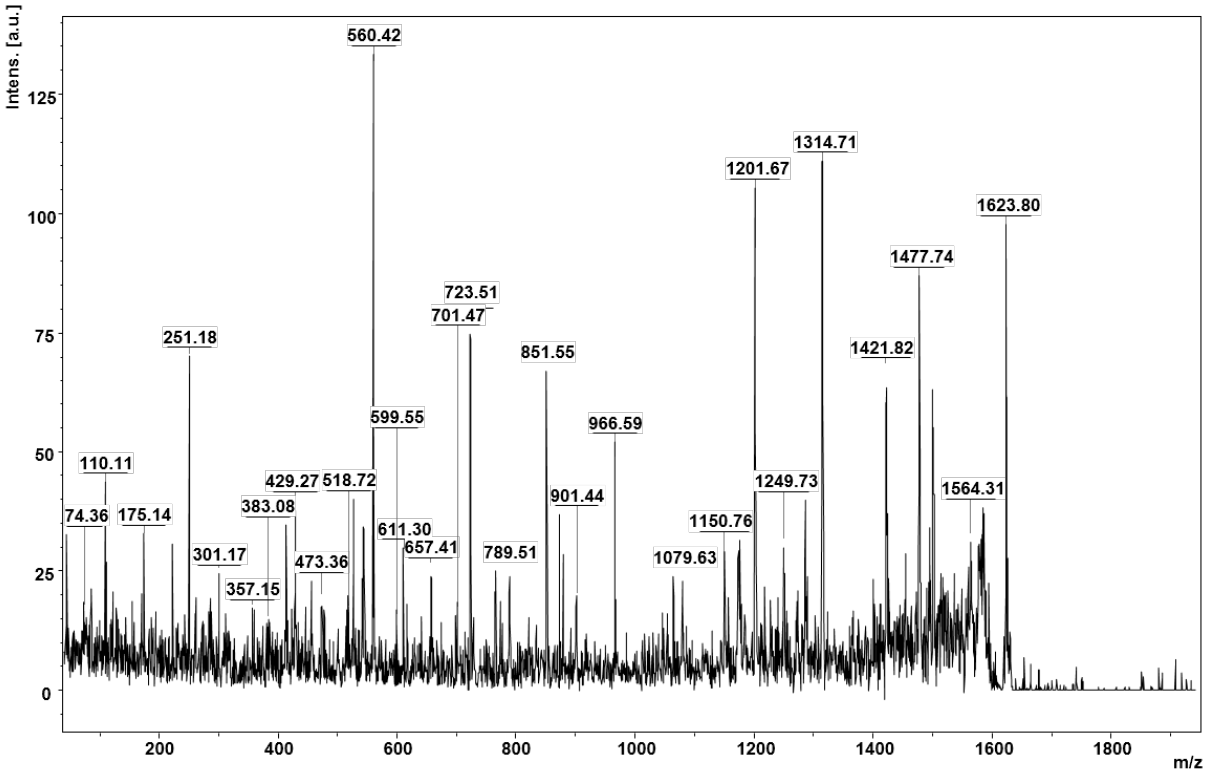


Figure 98: MS/MS spectrum of sample 3A_1623 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

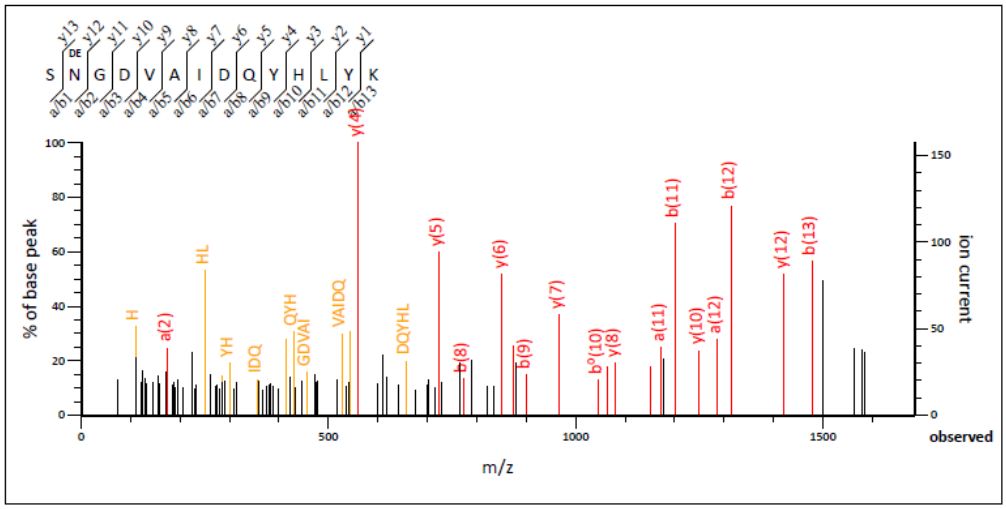


Figure 99: Peptide fragmentation and sequence of sample 3A_1623. The sequence is displayed in the left top corner.

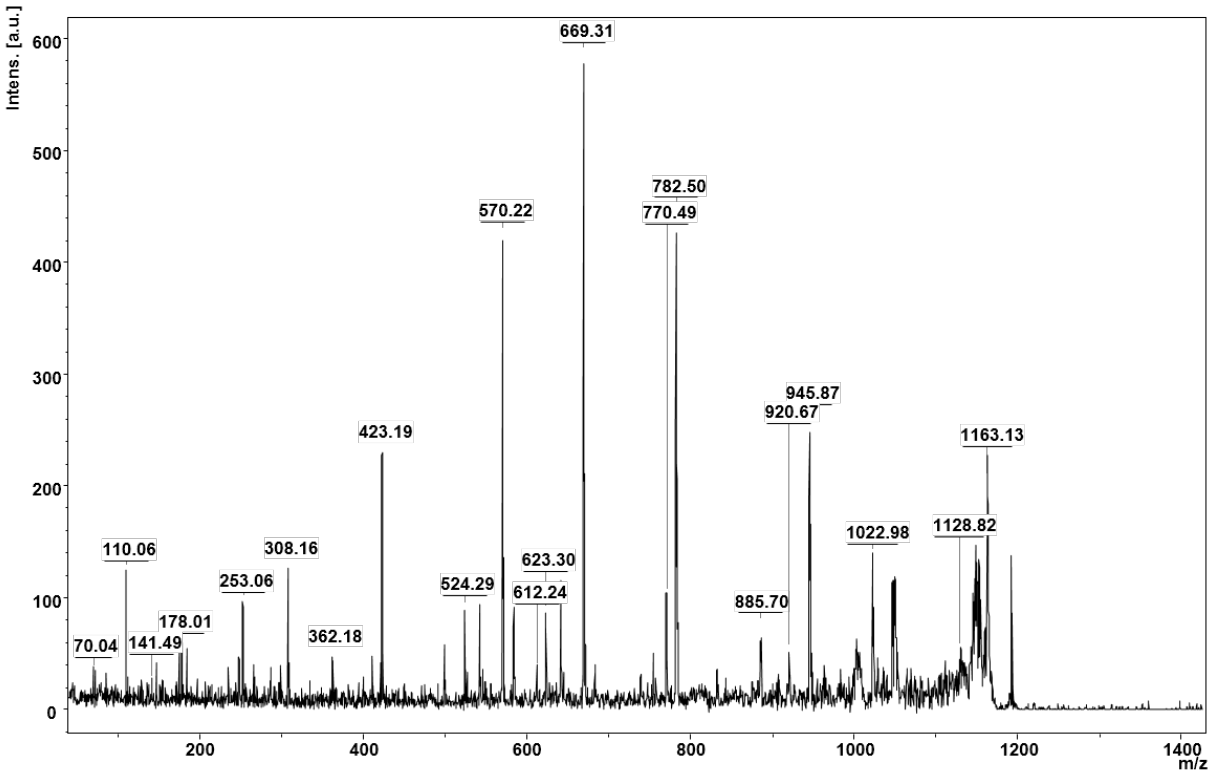


Figure 100: MS/MS spectrum of sample 3B_1192 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

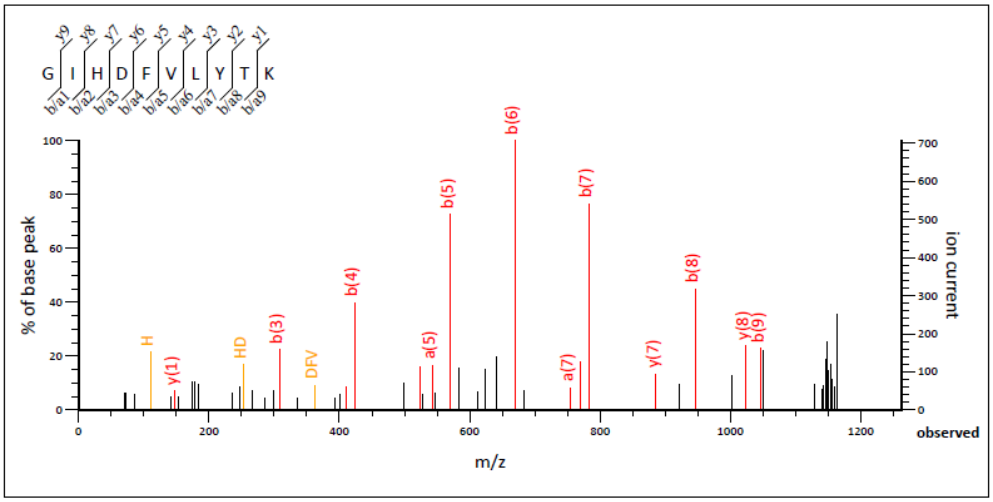


Figure 101: Peptide fragmentation and sequence of sample 3B_1192. The sequence is displayed in the left top corner.

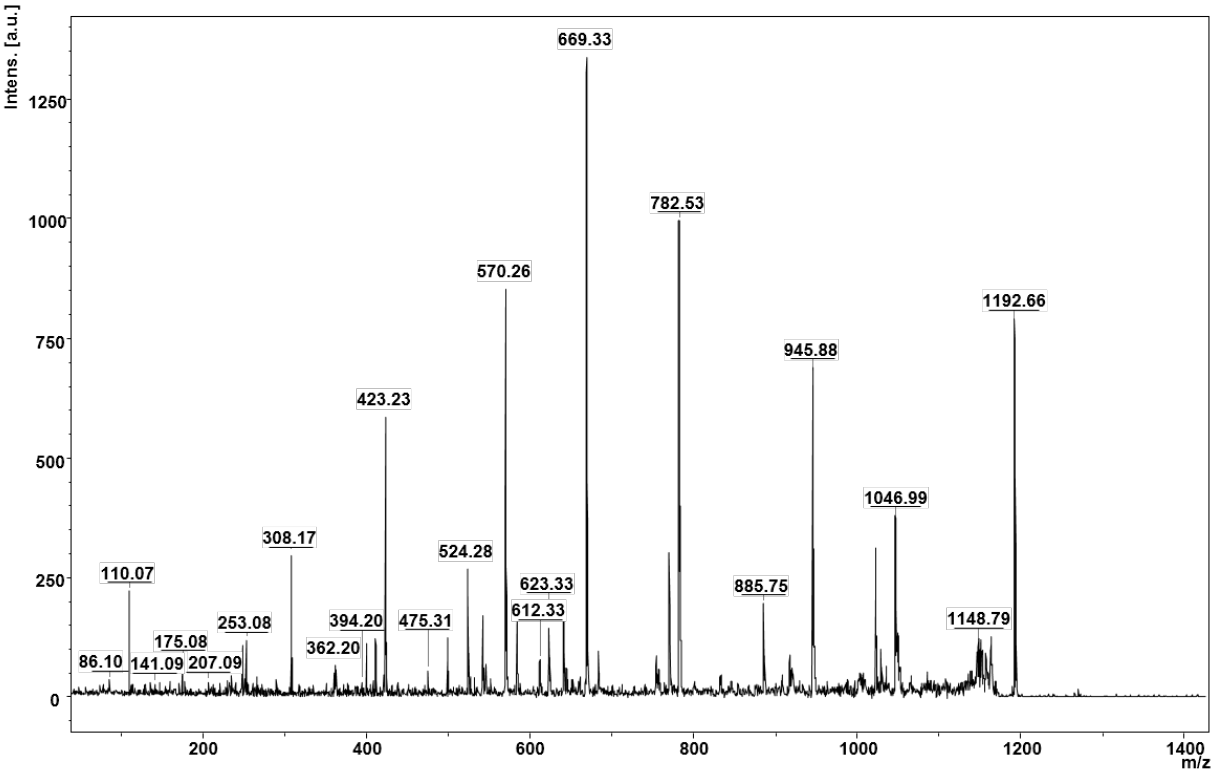


Figure 102: MS/MS spectrum of sample 3C_1192 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

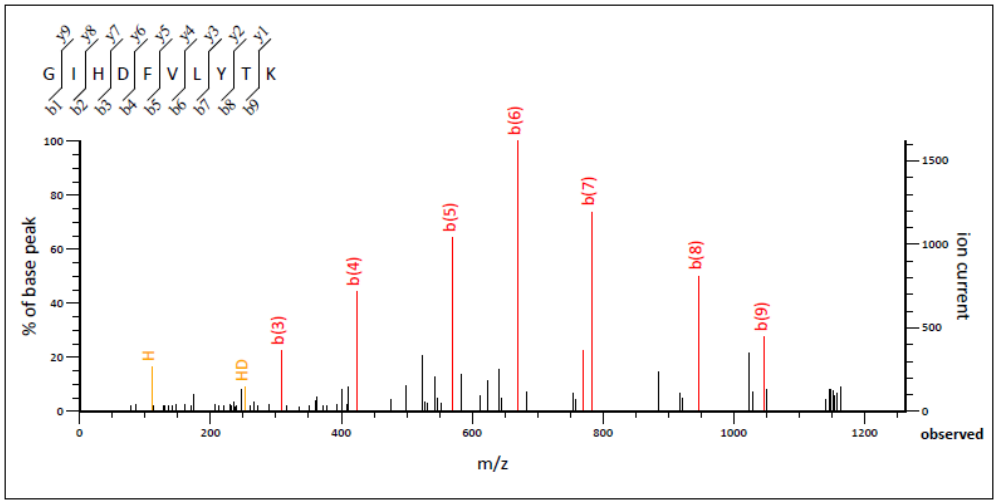


Figure 103: Peptide fragmentation and sequence of sample 3C_1192. The sequence is displayed in the left top corner.

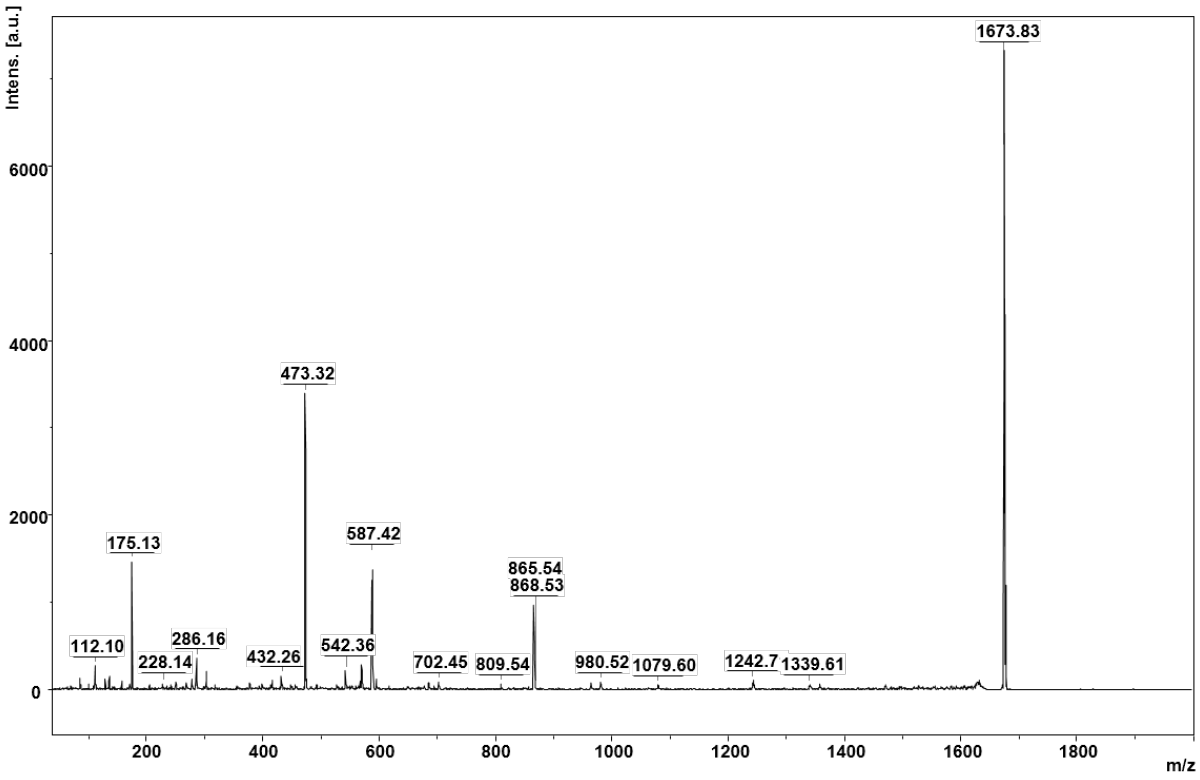


Figure 104: MS/MS spectrum of sample 3C_1673 of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

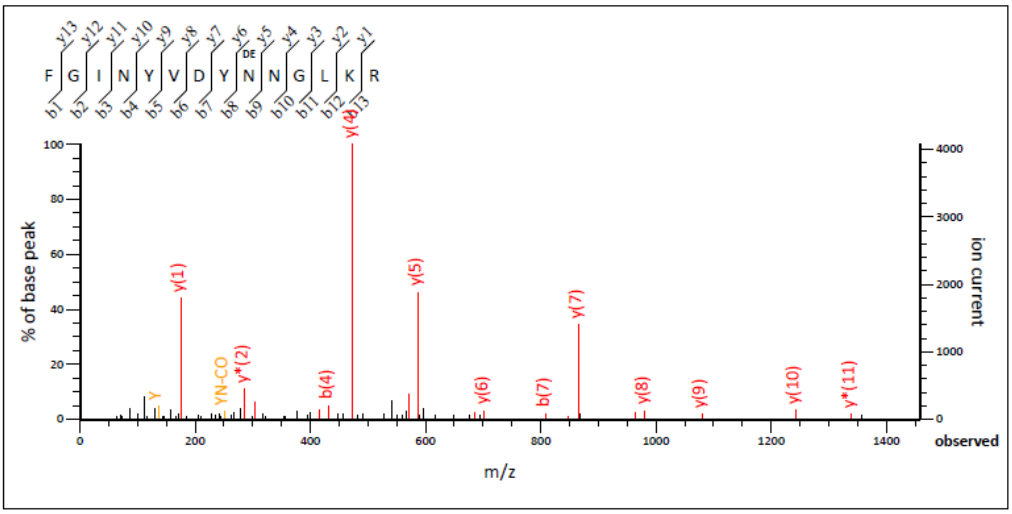


Figure 105: Peptide fragmentation and sequence of sample 3C_1673. The sequence is displayed in the left top corner.

The Mascot search results are listed in table 38. Only mass spectra with a relevant theoretical weight, similar to the estimation from the gel band (59 kDa) are shown. Due to the proteins being isolated and extracted from *M. domestica*, only results from this species were considered for the protein identification.

The MS^2 spectra of 3A_1192, 3A_1511, 3A_1623, 3B_1192, 3C_1192 and3C_1673 all identified the protein as beta-glucosidase 12-like and beta-glucosidase 12-like precursor from *M. domestica* or related species, which are classified as unique peptides and therefore make the result very reliable.

The identification of 3A_1511 yielded the result cucumisin-like isoform X1 and X2 as well as partial, was already confirmed for the Gel band 2, therefore this could be carryover to Gel band 3.

The analyzed Gel band 3 more than likely consists of multiple proteins, which would explain the very high numbers of obtained MS peaks, but beta-glucosidase 12-like and beta-glucosidase 12-like precursor from *M. domestica* can be considered as a correct result for the identification with 95 % probability.

Table 38: Mascot search results of Gel Band 3 for all samples. Only significant hits with relevant theoretical molecular weights, compared to the rough estimate of the gel bands, are listed.

Sample	Taxonomy	Name of protein and Species	Matches	Score/ Thres-hold	th. MW [Da]	th. pI	Protein coverage [%]
3A_1192	Rosaceae	Beta-glucosidase 12-like precursor [*Malus domestica*] NP_001280983.1	1/1	57/45	61259	5.56	1
3A_1192	Rosaceae	Beta-glucosidase 12-like [*Malus domestica*] XP_008357590.1	1/1	57/45	61521	5.88	1

3A_1511	Rosaceae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	66/46	79768	6.76	1
3A_1511	Rosaceae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	66/46	76338	6.74	1
3A_1511	Rosaceae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	66/46	64987	8.39	2
3A_1623	Rosaceae	Beta-glucosidase 12-like precursor [*Malus domestica*] NP_001280983.1	1/1	61/46	61259	5.56	2
3A_1623	Rosaceae	Beta-glucosidase 12-like [*Malus domestica*] XP_008357590.1	1/1	61/46	61521	5.88	2
3B_1192	Rosaceae	Beta-glucosidase 12-like precursor [*Malus domestica*] NP_001280983.1	1/1	63/45	61259	5.56	1
3B_1192	Rosaceae	Beta-glucosidase 12-like [*Malus domestica*] XP_008357590.1	1/1	63/45	61521	5.88	1
3C_1192	Rosaceae	Beta-glucosidase 12-like precursor [*Malus domestica*] NP_001280983.1	1/1	65/45	61259	5.56	1
3C_1192	Rosaceae	Beta-glucosidase 12-like [*Malus domestica*] XP_008357590.1	1/1	65/45	61521	5.88	1
3C_1673	Rosaceae	Beta-glucosidase 12-like precursor [*Malus domestica*] NP_001280983.1	1/1	94/45	61259	5.56	1
3C_1673	Rosaceae	Beta-glucosidase 12-like [*Malus domestica*] XP_008357590.1	1/1	94/45	61044	6.10	2
3C_1673	Rosaceae	Beta-glucosidase 12-like isoform X1 [*Malus domestica*] XP_008384719.1	1/1	93/45	58993	6.57	2
3C_1673	Rosaceae	Beta-glucosidase 12-like isoform X2 [*Malus domestica*] XP_017191209.1	1/1	93/45	47656	7.19	3

3A_1192	Viridiplantae	Beta-glucosidase 12-like precursor [*Malus domestica*] NP_001280983.1	1/1	67/58	61259	5.56	1
3A_1192	Viridiplantae	Beta-glucosidase 12-like [*Malus domestica*] XP_008357590.1	1/1	67/58	61521	5.88	1
3A_1511	Viridiplantae	cucumisin-like isoform X1 [*Malus domestica*] XP_008351270.1	1/1	66/58	79768	6.76	1
3A_1511	Viridiplantae	cucumisin-like isoform X2 [*Malus domestica*] XP_008351271.1	1/1	66/58	76338	6.74	1
3A_1511	Viridiplantae	cucumisin-like, partial [*Malus domestica*] XP_017184694.1	1/1	66/58	64987	8.39	2
3A_1623	Viridiplantae	Beta-glucosidase 12-like precursor [*Malus domestica*] NP_001280983.1	1/1	61/58	61259	5.56	2
3A_1623	Viridiplantae	Beta-glucosidase 12-like [*Malus domestica*] XP_008357590.1	1/1	61/58	61521	5.88	2
3B_1192	Viridiplantae	Beta-glucosidase 12-like precursor [*Malus domestica*] NP_001280983.1	1/1	63/58	61259	5.56	1
3C_1192	Viridiplantae	Beta-glucosidase 12-like precursor [*Malus domestica*] NP_001280983.1	1/1	65/58	61259	5.56	1
3C_1673	Viridiplantae	Beta-glucosidase 12-like [*Malus domestica*] XP_008357590.1	1/1	94/58	61157	5.88	1
3C_1673	Viridiplantae	Beta-glucosidase 12-like [*Malus domestica*] XP_008357590.1	1/1	94/58	61259	5.56	1

4.5.4 Gel Band 4

In the figures 106-112 the mass spectra of the filtered peptide peaks are shown. The peaks used for the search in Mascot are listed in the Appendix (see Chapter 7.1.4).

The mass spectra show diversity in the peak pattern over the whole m/z range, especially for sample 4BC*, which was extracted with a more hydrophobic approach. Consequently, it could be that more tryptic self-digested peptides as well as membrane proteins are extracted.

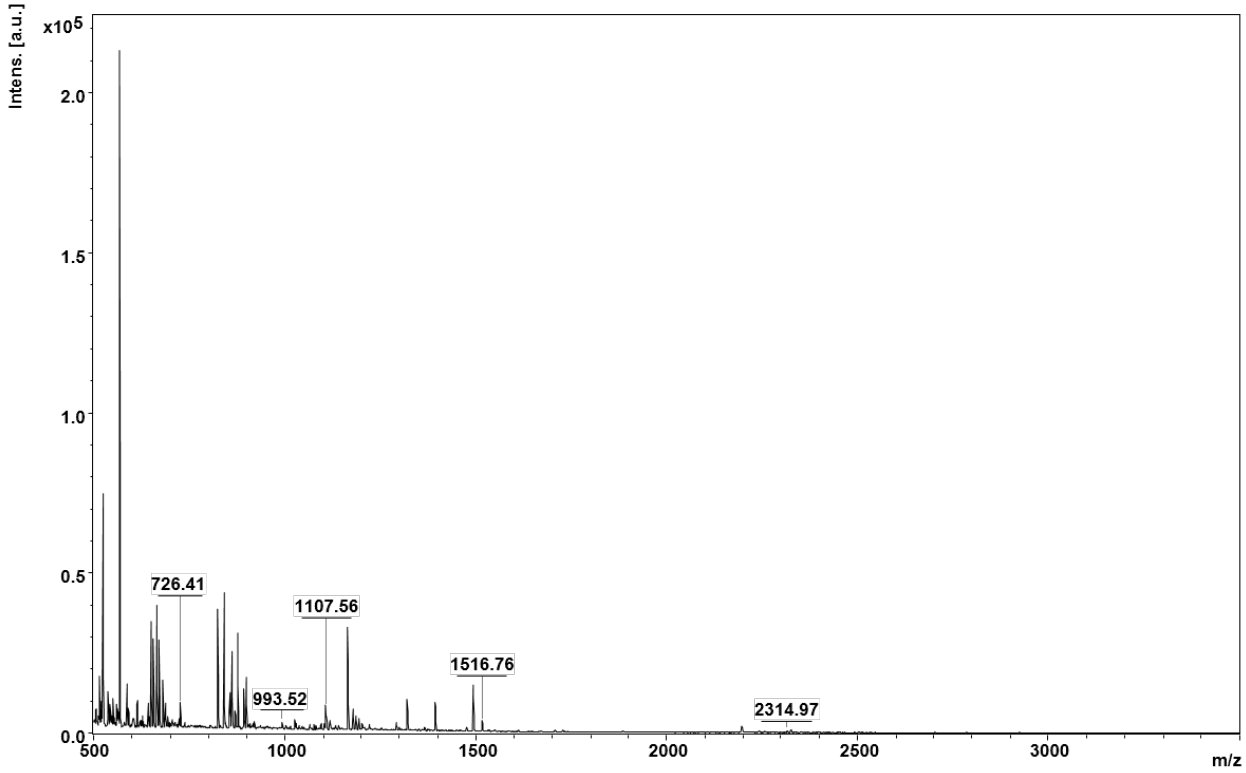


Figure 106: Mass spectrum of sample 4A of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

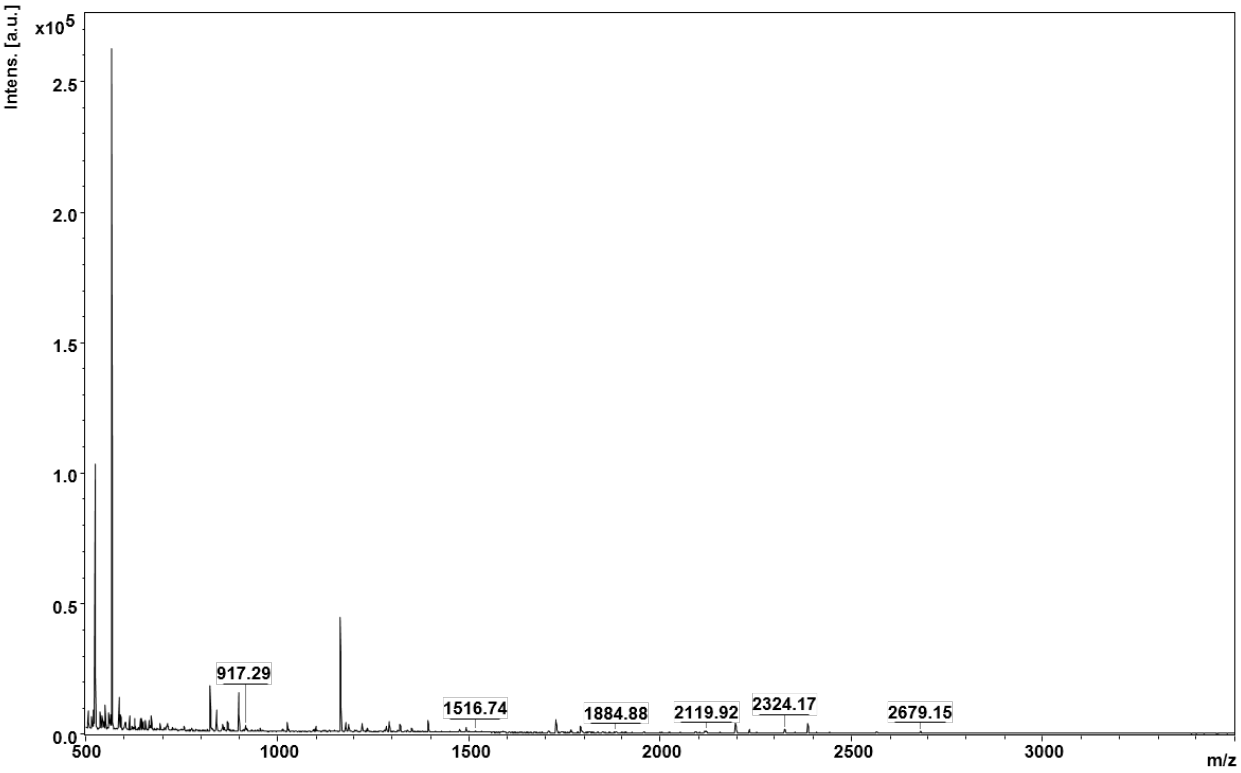


Figure 107: Mass spectrum of sample 4B of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

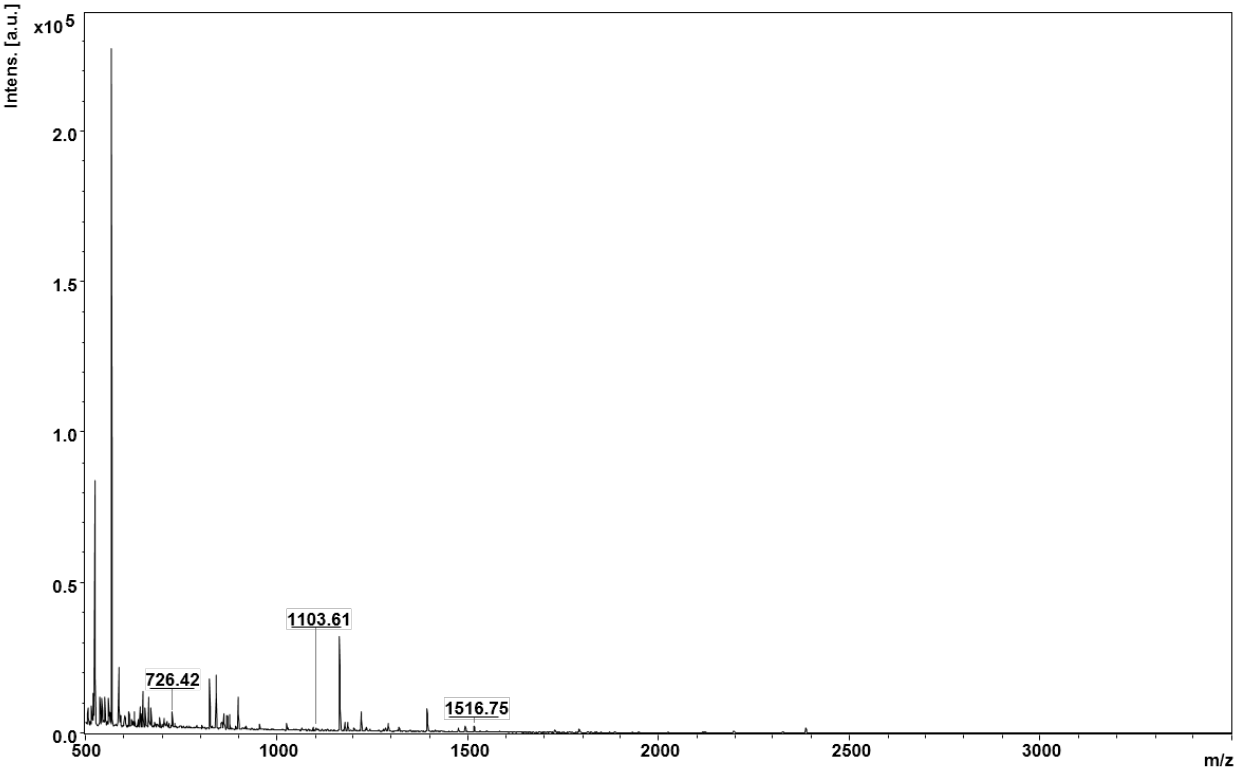


Figure 108: Mass spectrum of sample 4C of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

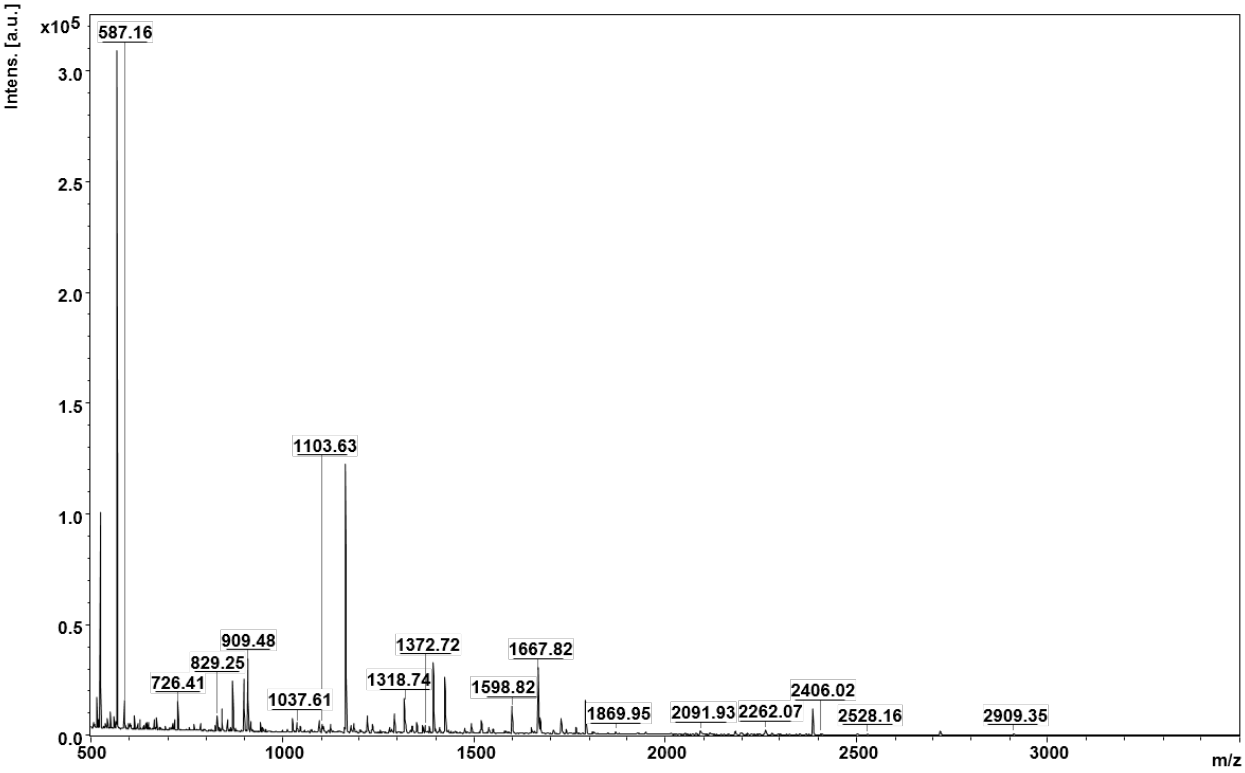


Figure 109: Mass spectrum of sample 4DE of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

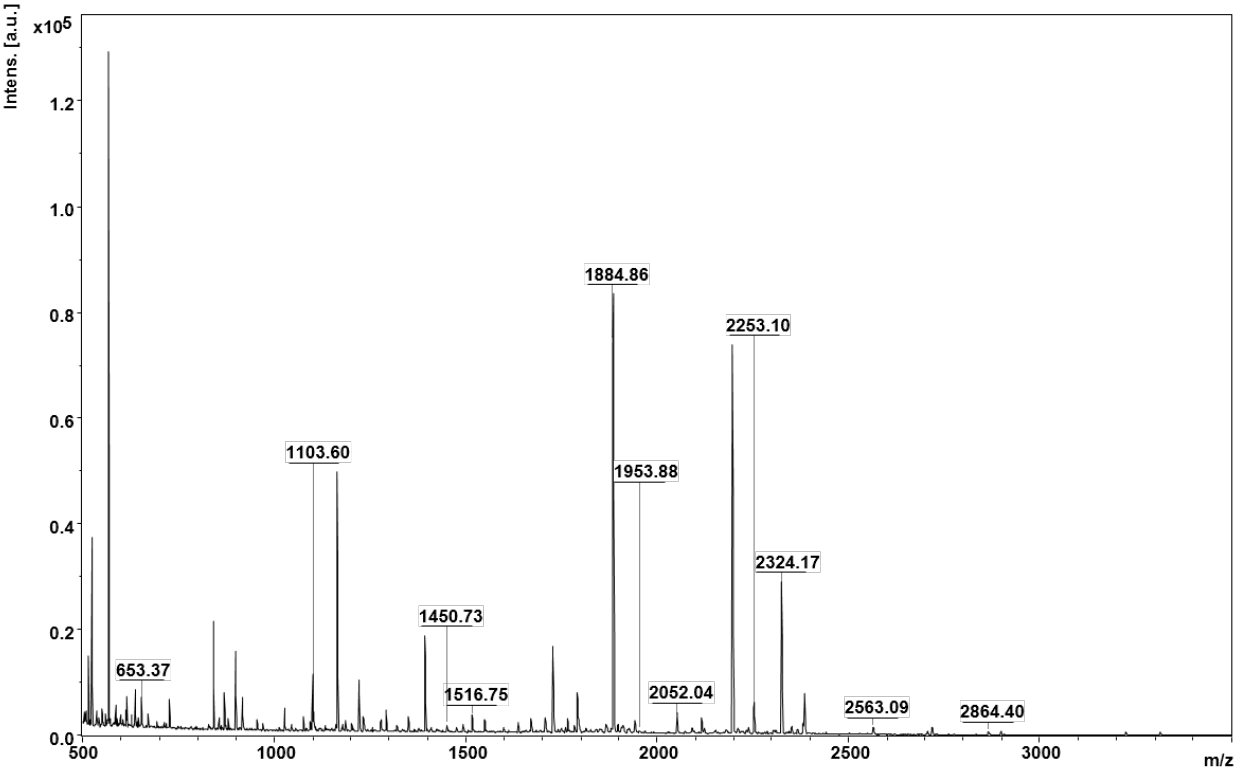


Figure 110: Mass spectrum of sample 4FGH of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

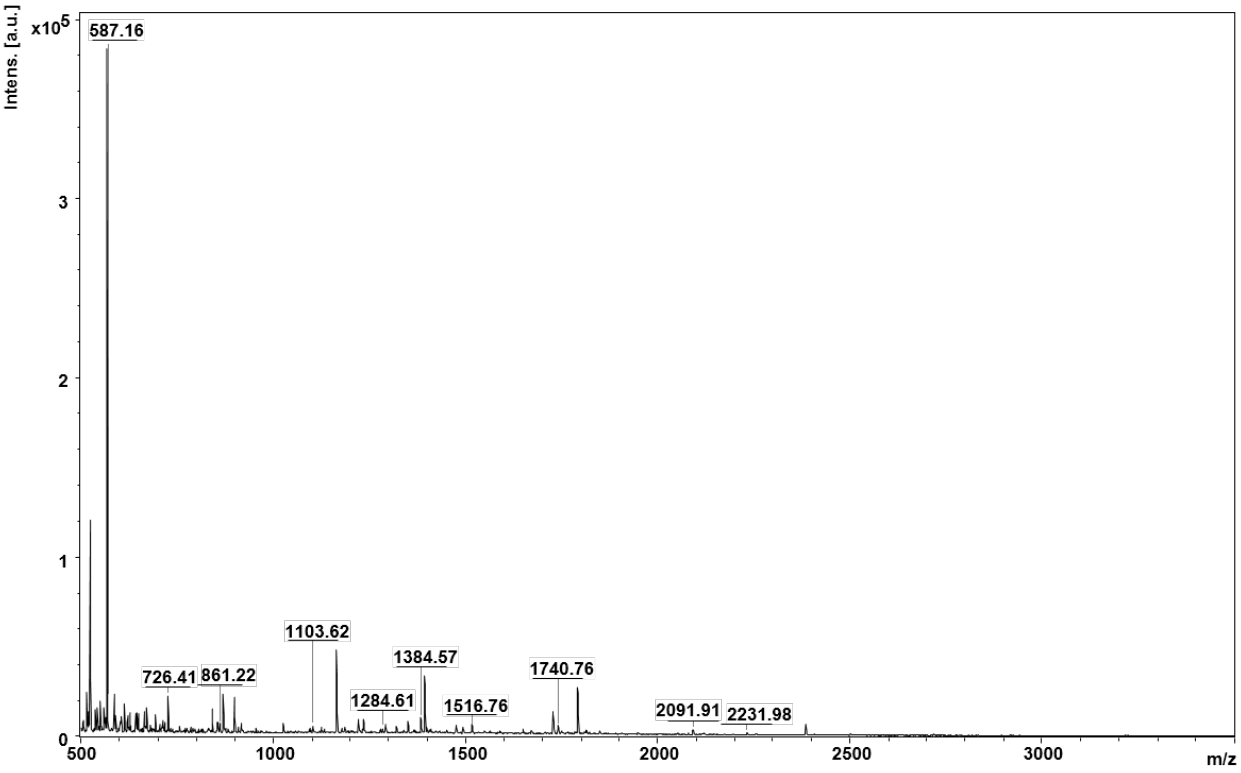


Figure 111: Mass spectrum of sample 4A of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

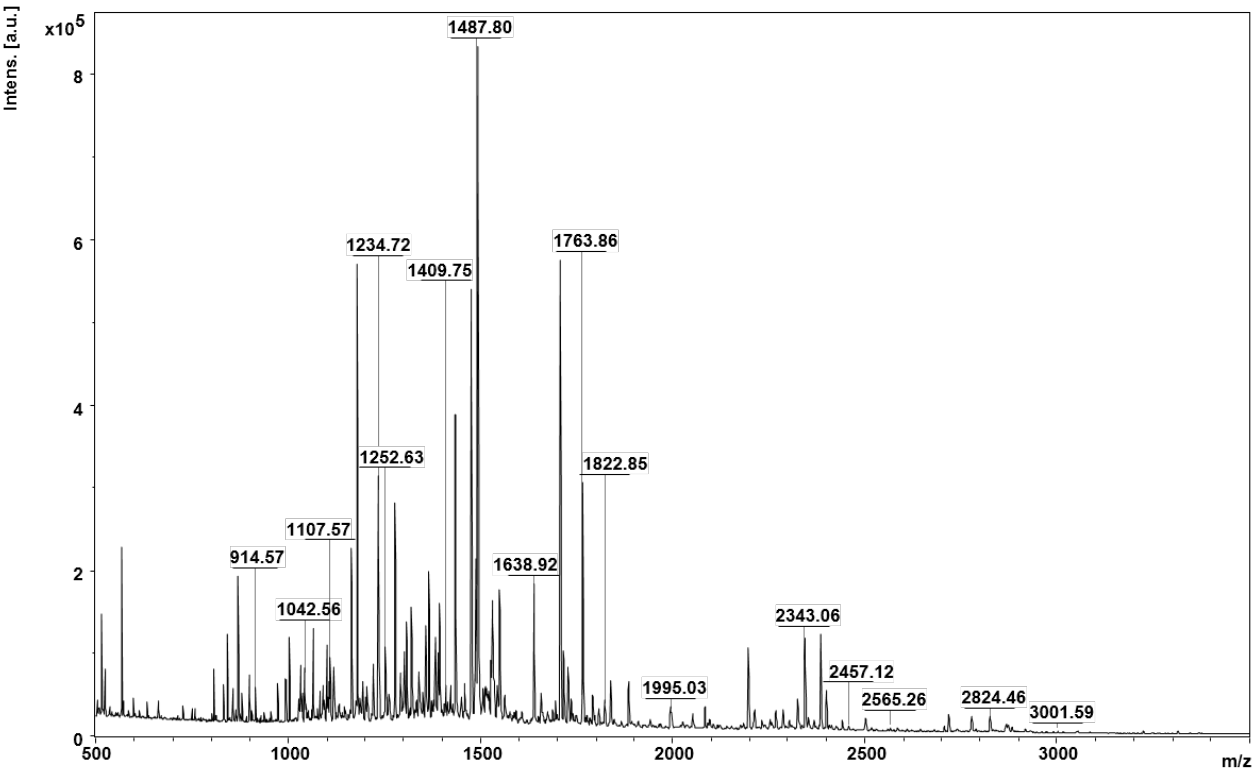


Figure 112: Mass spectrum of sample 4BC of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

The Mascot search results are listed in table 39. Only mass spectra with a relevant theoretical weight, similar to the estimation from the gel band (53 kDa) are shown. Due to the proteins being isolated and extracted from *M. domestica*, only results from this species were considered for the protein identification.

No MS^2 spectra of relevance could be obtained for the Gel band 4, therefore, no information about the peptides and an accurate identification could be achieved.

For the Gel band 4 only sample 4DE could be identified with a relevant molecular weight and a significant probability. The identified protein is an uncharacterized protein from *M. domestica*, which belongs already to the original organism. Due to it being an uncharacterized protein it could be a possible candidate for correct identification, but there were no other indications from the other samples of Gel band 4.

Table 39: Mascot search results of Gel Band 4 for all samples. Only significant hits with relevant theoretical molecular weights, compared to the rough estimate of the gel bands, are listed.

Sample	Taxonomy	Name of protein and Species	Matches	Score/ Thres-hold	th. MW [Da]	th. pI	Protein coverage [%]
4BC*	Rosaceae	Uncharacterized protein LOC103406332 [*Malus domestica*] XP_008343551.1	22/47	75/68	64652	6.04	42

4.5.5 Gel Band 5

In the figures 113-119 the mass spectra of the filtered peptide peaks are shown. The peaks used for the search in Mascot are listed in the Appendix (see Chapter 7.1.5).

The mass spectra show a high diversity in the peak pattern over the whole m/z range, especially for sample 5FGH. The probability of a reliable identification of the digested protein is therefore rather low.

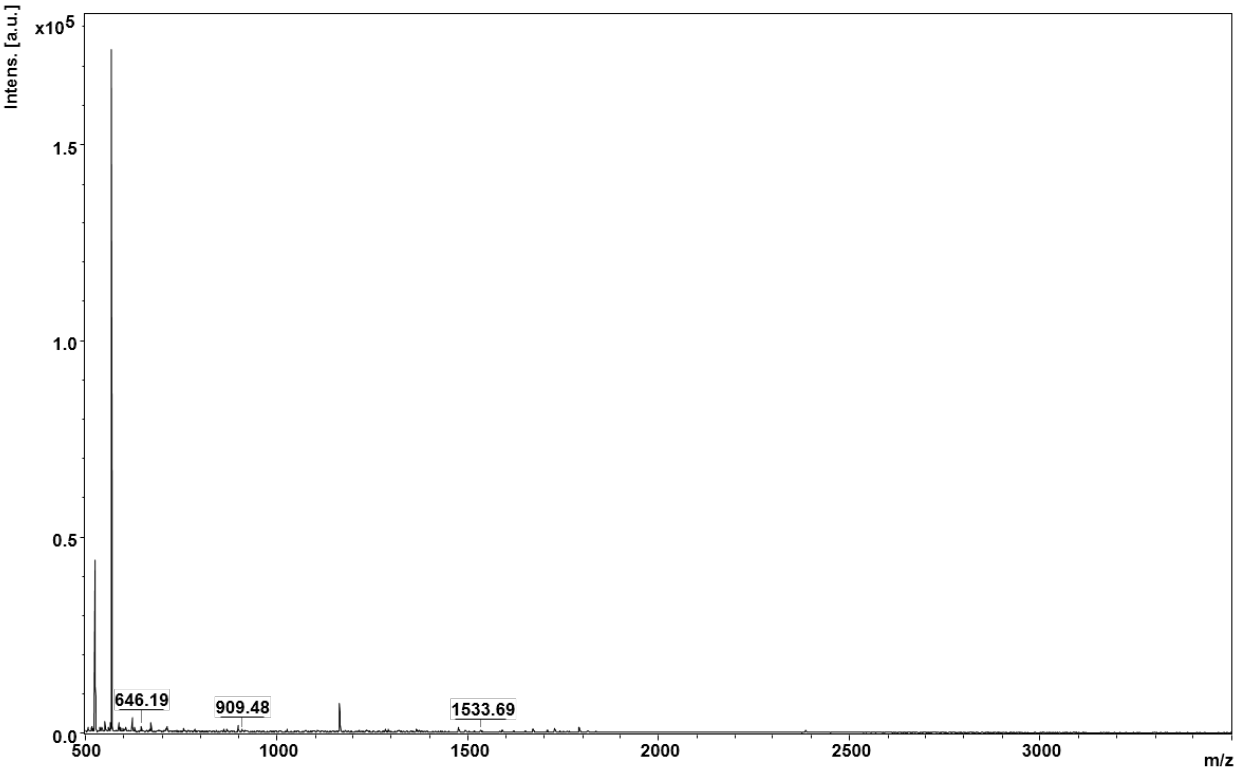


Figure 113: Mass spectrum of sample 5A of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

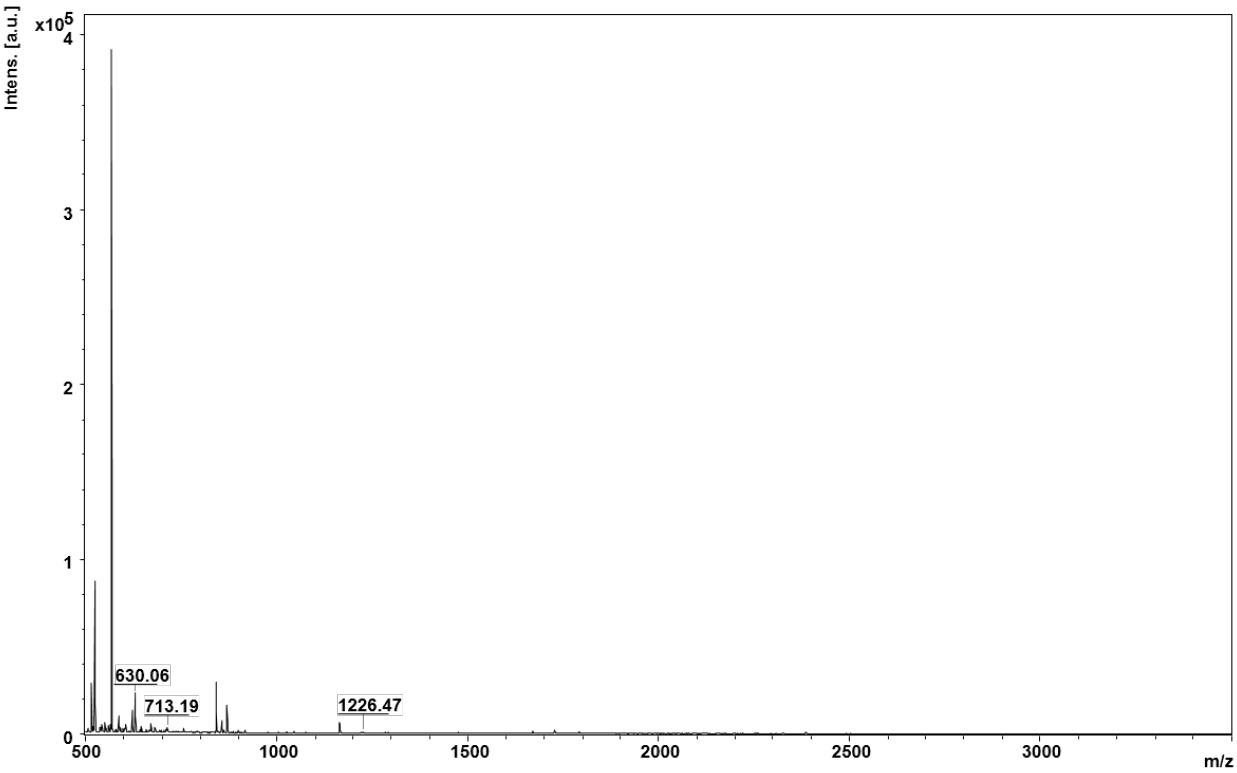


Figure 114: Mass spectrum of sample 5B of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

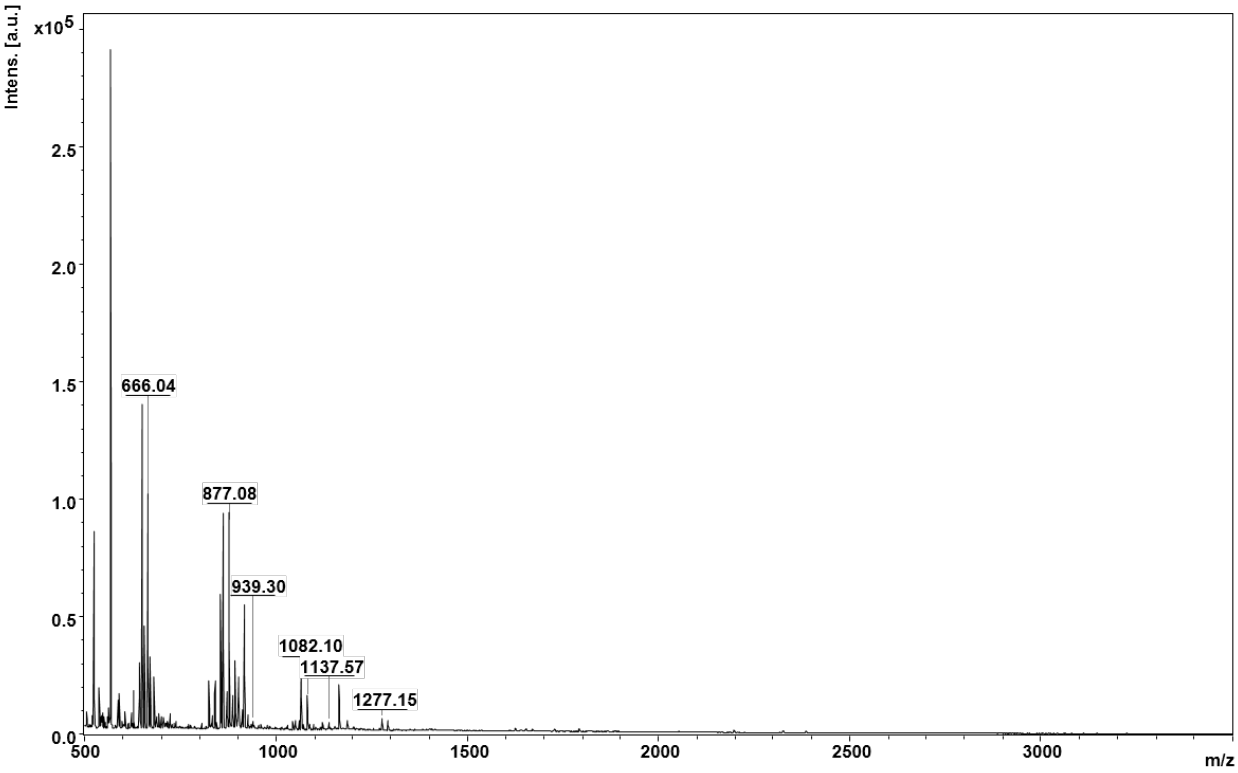


Figure 115: Mass spectrum of sample 5C of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

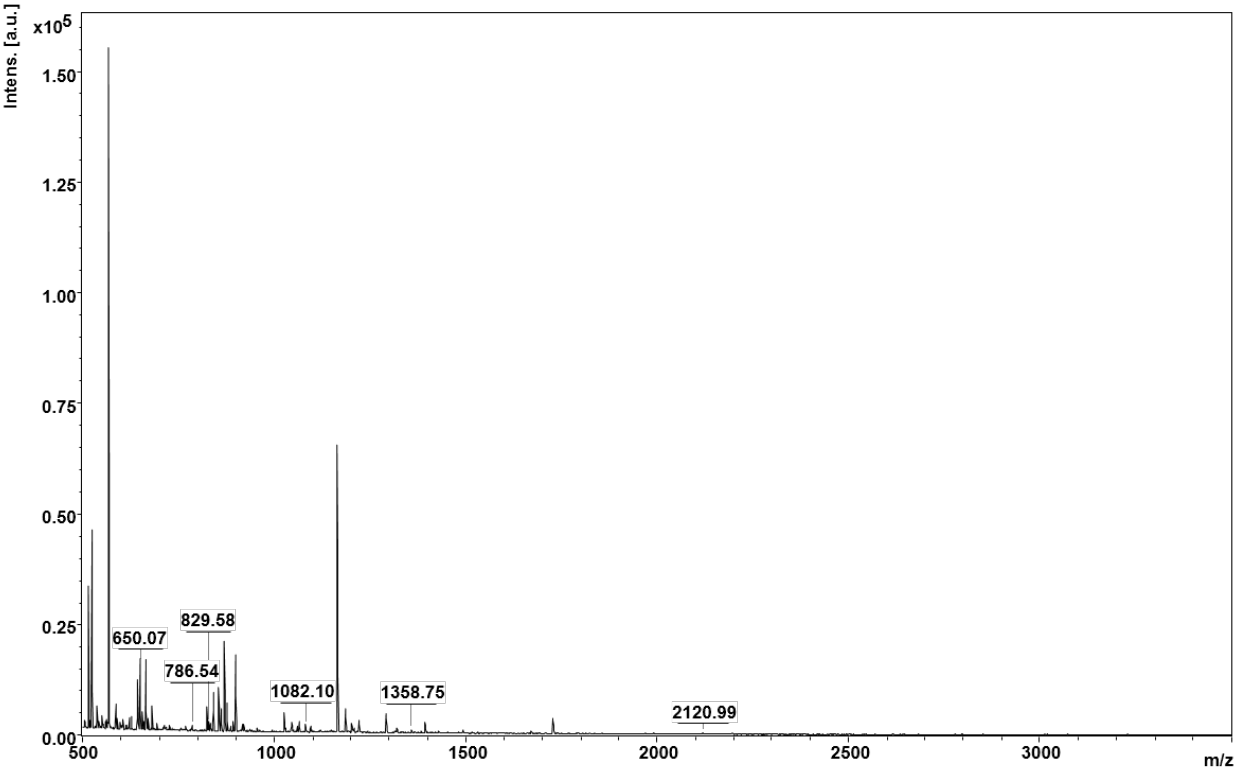


Figure 116: Mass spectrum of sample 5DE of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

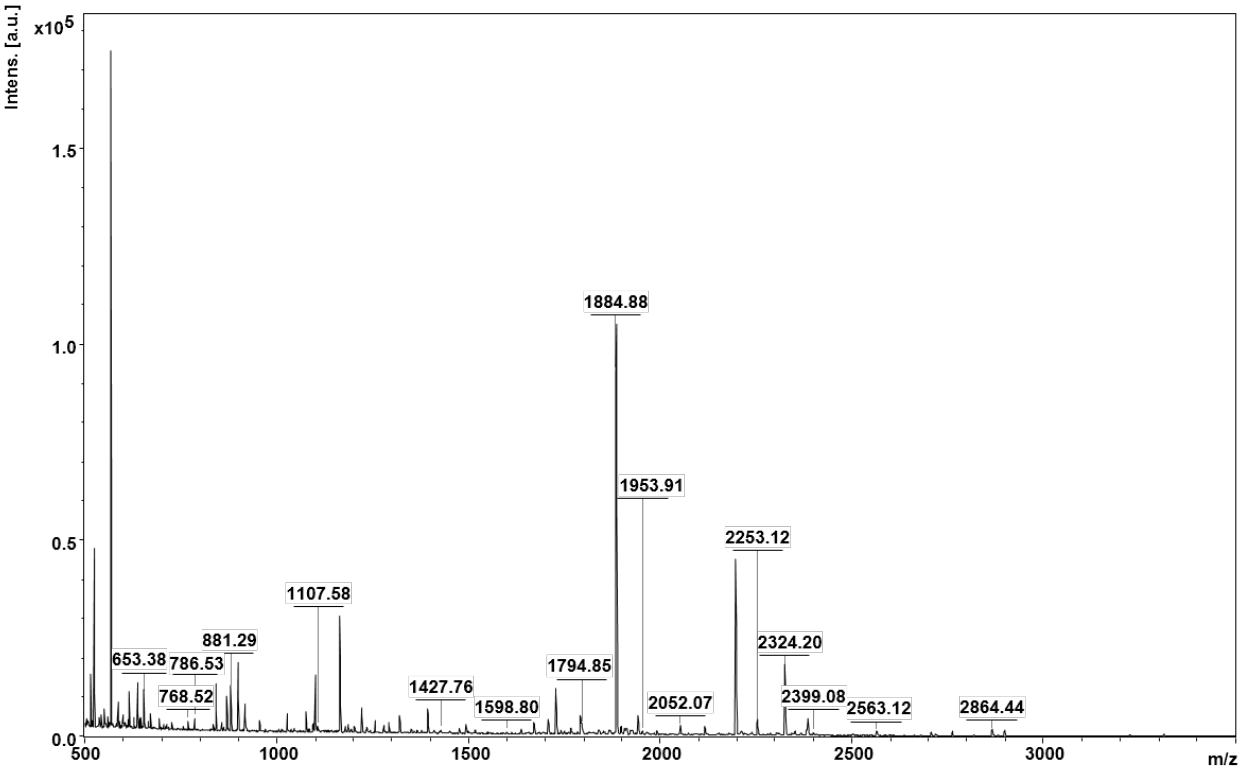


Figure 117: Mass spectrum of sample 5FGH of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

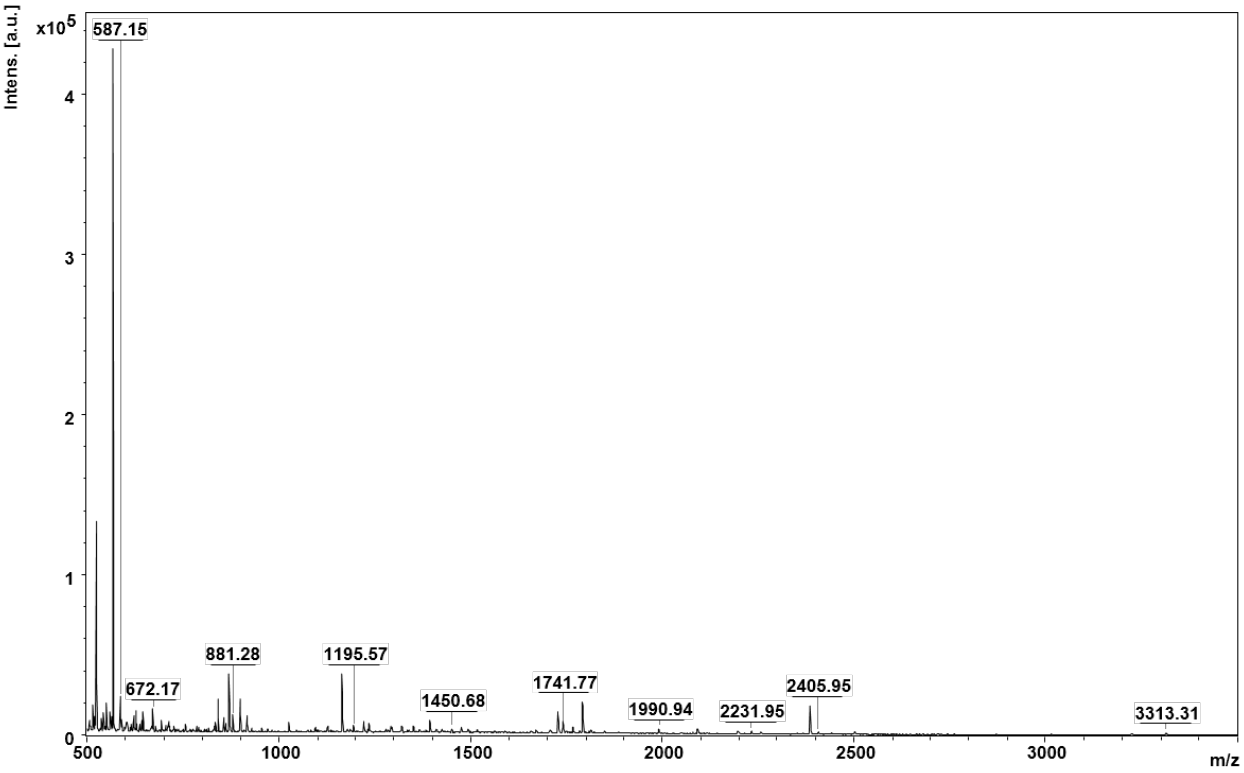


Figure 118: Mass spectrum of sample 5A of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

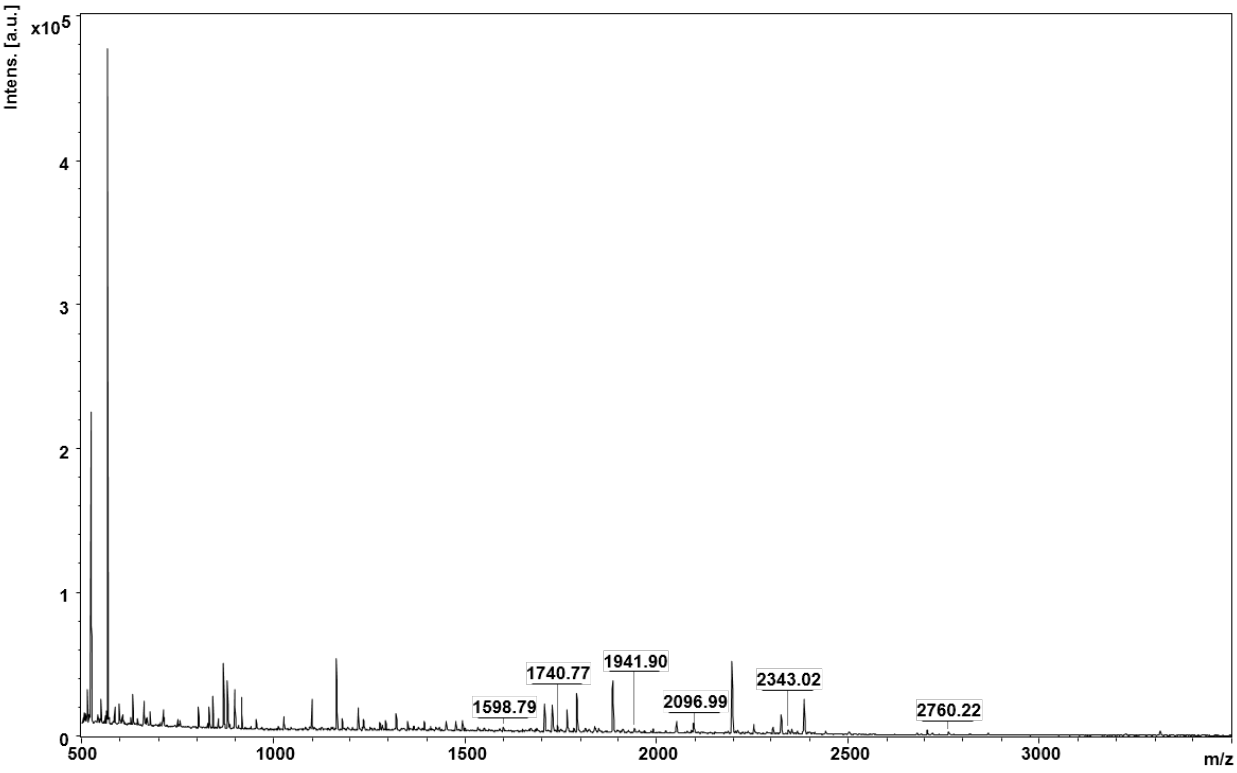


Figure 119: Mass spectrum of sample 5BC of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

No identification for the mass spectra and no MS^2 spectra of relevance could be obtained for the Gel band 5, therefore, no information about the peptides and no accurate identification could be achieved.

4.5.6 Gel Band 6

In the figures 120-126 the mass spectra of the filtered peptide peaks are shown. The peaks used for the search in Mascot are listed in the Appendix (see Chapter 7.1.6).

The mass spectra show a high diversity in the peak pattern over the whole m/z range, from low intensity and low peak numbers to high intensity and high peak numbers. Especially sample 6BC* shows a high peak numbers in the lower m/z range, which could be due to the more hydrophobic extraction and thus an extraction of more tryptic self-digested peptide, which are common in low m/z ranges. Therefore, the likelihood of a reliable identification of the digested protein is rather small.

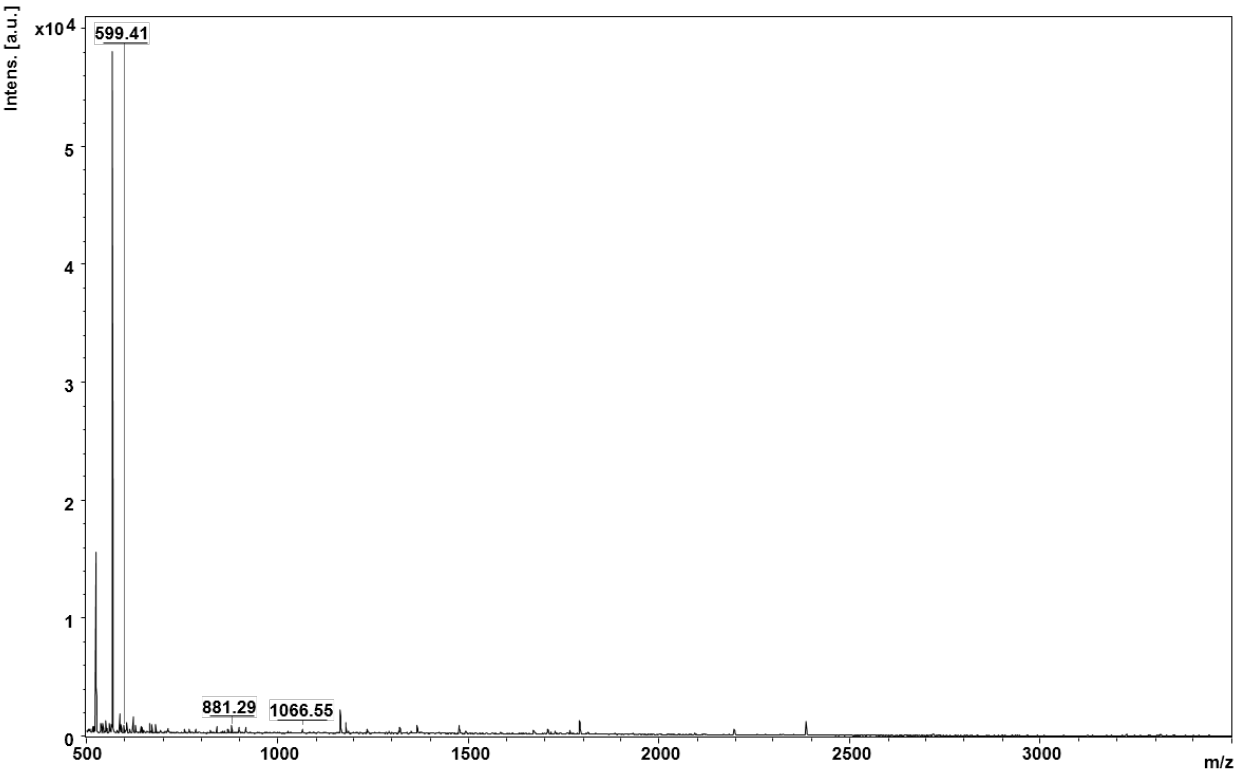


Figure 120: Mass spectrum of sample 6A of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

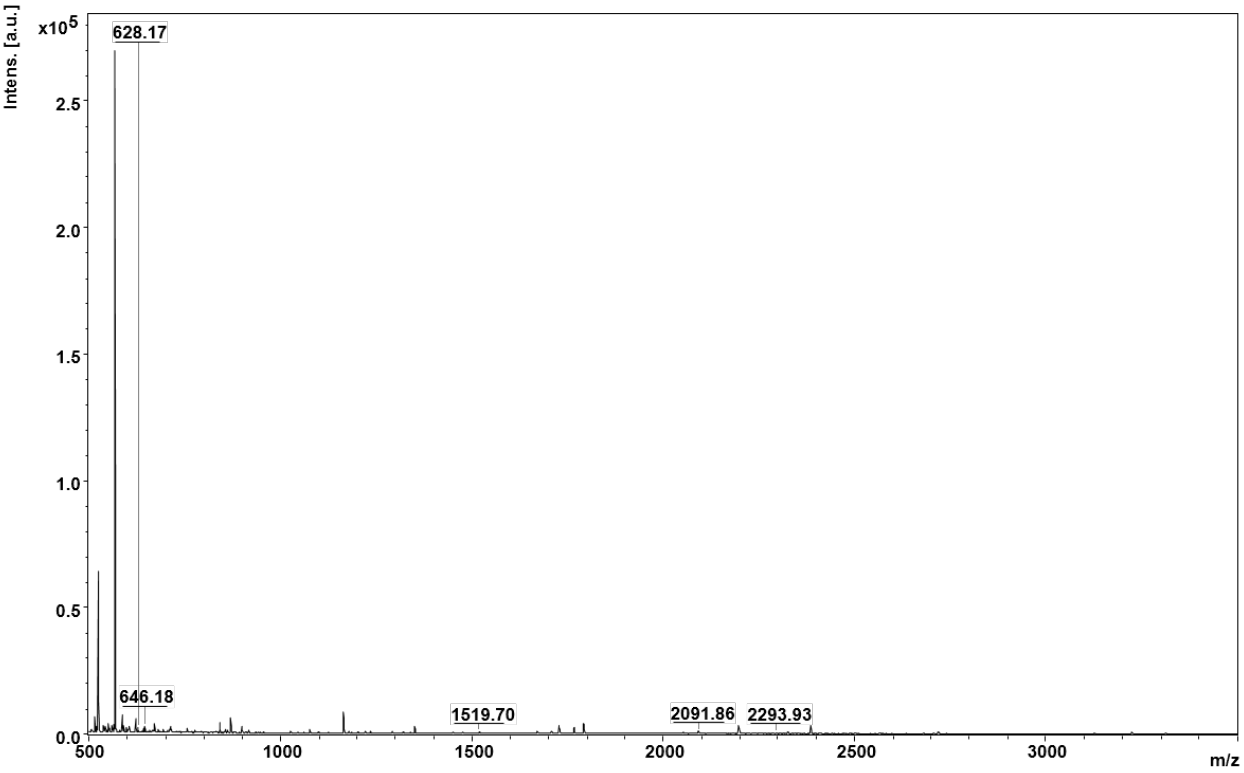


Figure 121: Mass spectrum of sample 6B of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

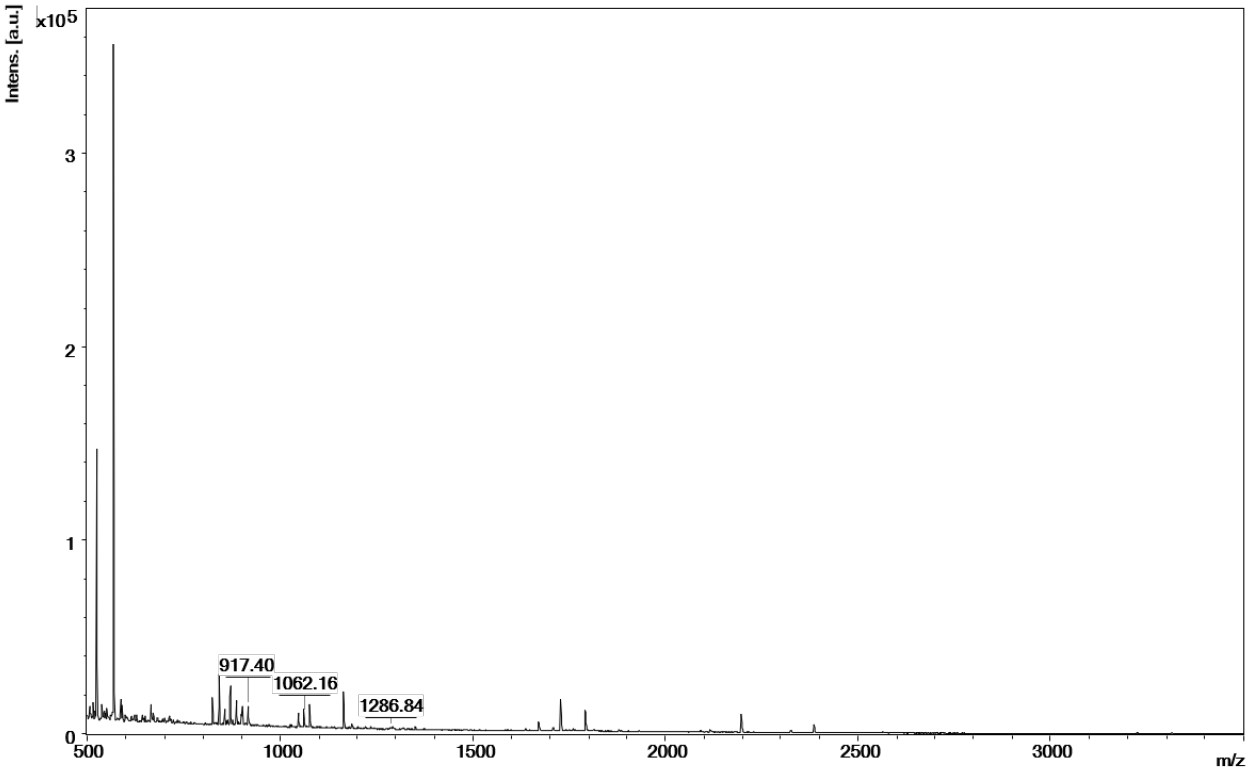


Figure 122: Mass spectrum of sample 6C of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

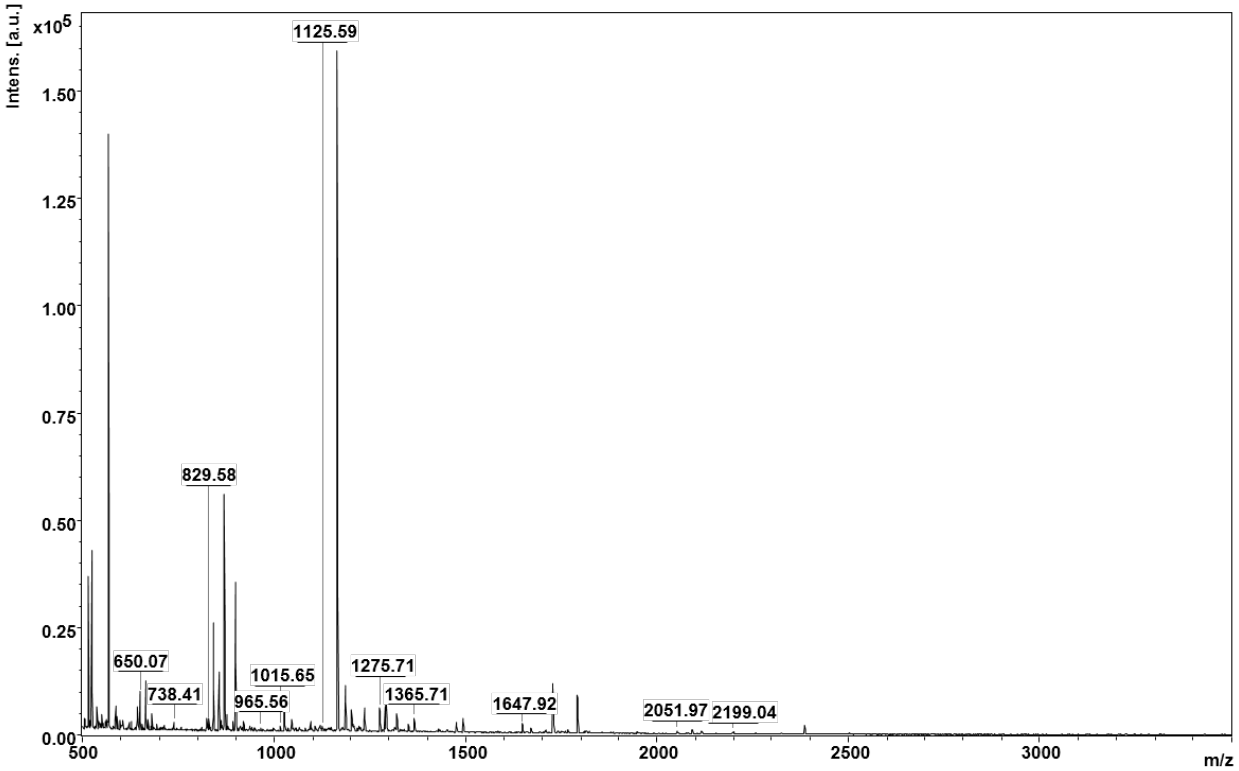


Figure 123: Mass spectrum of sample 6DE of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

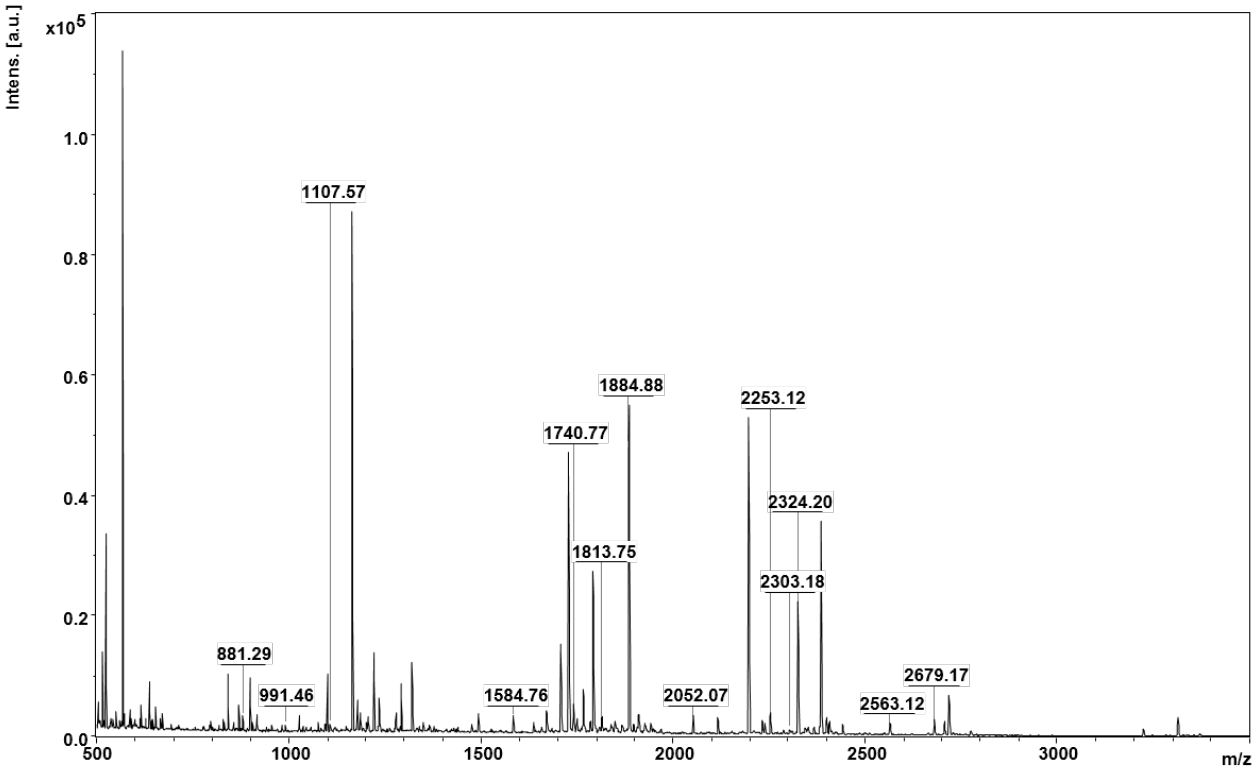


Figure 124: Mass spectrum of sample 6FGH of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

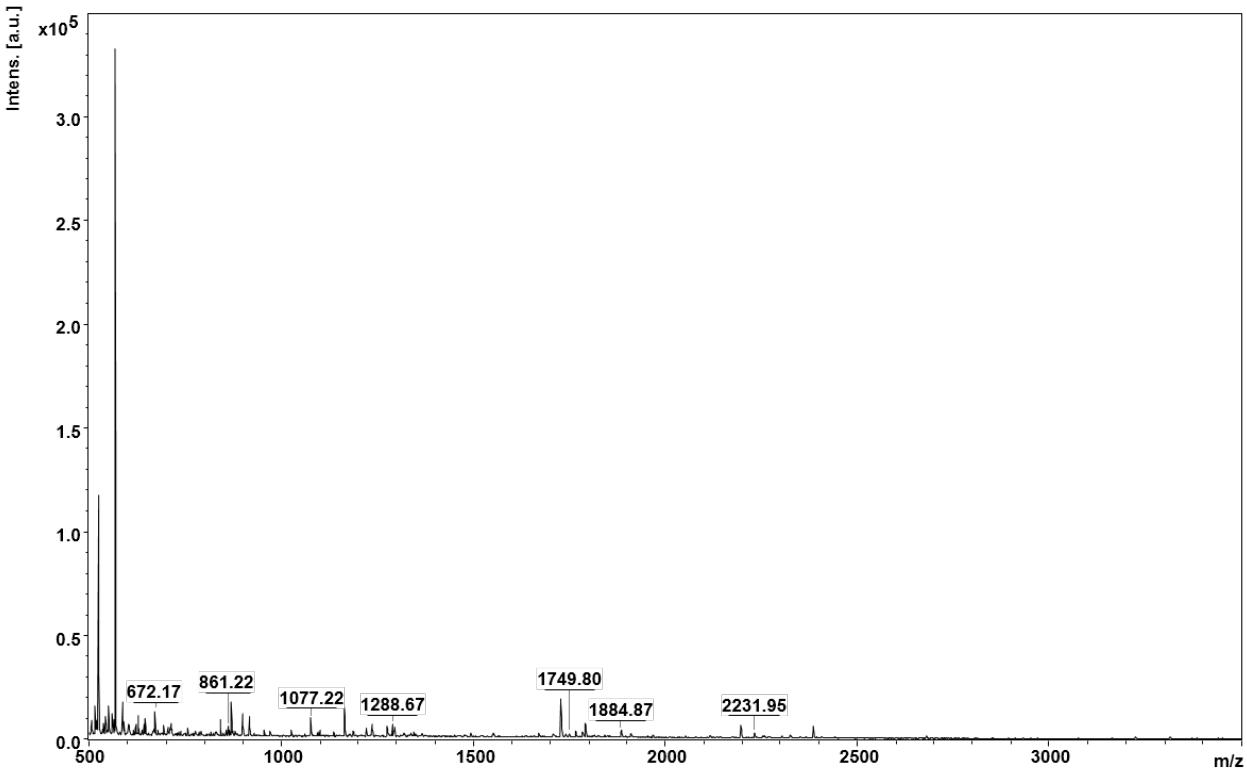


Figure 125: Mass spectrum of sample 6A of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

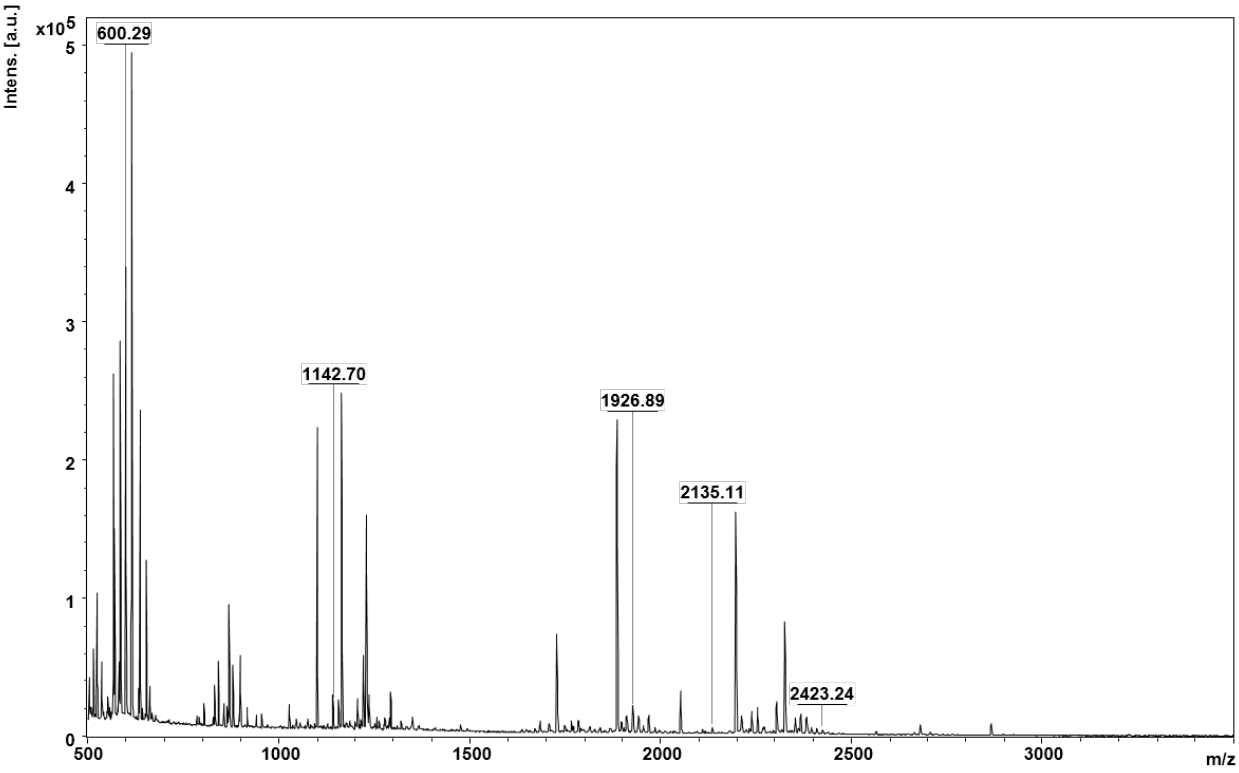


Figure 126: Mass spectrum of sample 6BC of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

For the Gel band 6 no identification of proteins at all could be achieved. Therefore, no statement to the possible protein can be made.

4.5.7 Gel Band 7

In the figures 127-133 the mass spectra of the filtered peptide peaks are shown. The peaks used for the search in Mascot are listed in the Appendix (see Chapter 7.1.7).

The mass spectra show a high diversity in the peak pattern over the whole m/z range, from low intensity and low peak number to high intensity and high peak number. Especially samples 7DE, 7FGH and 7BC* shows a high peak number over the whole m/z range. Therefore, the likelihood of a reliable identification of the digested protein is rather small.

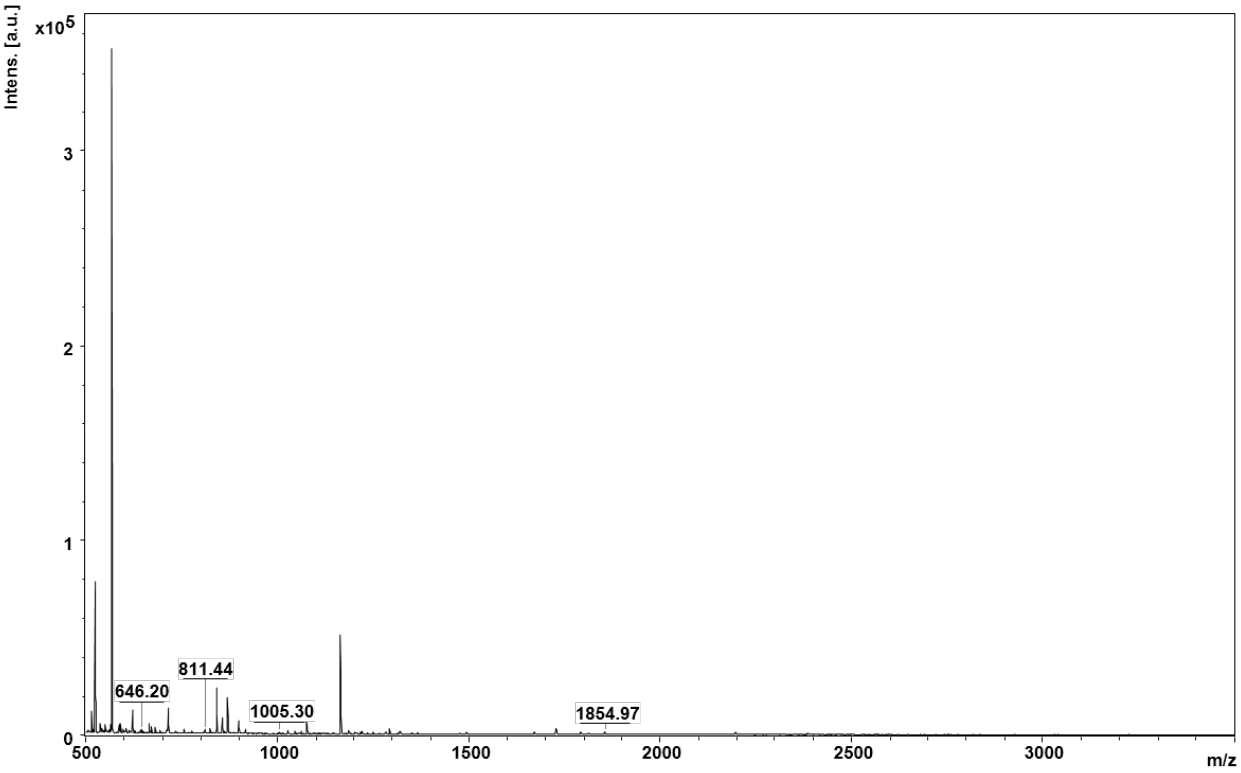


Figure 127: Mass spectrum of sample 7A of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

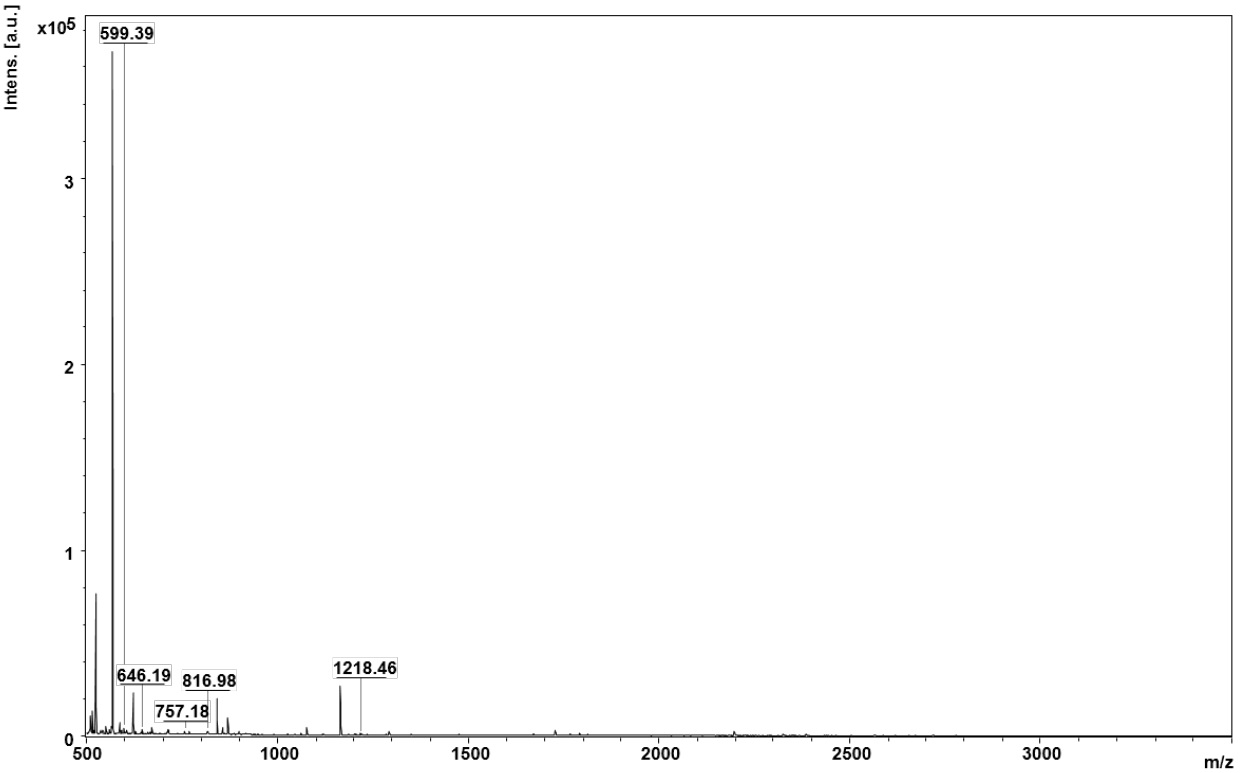


Figure 128: Mass spectrum of sample 7B of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

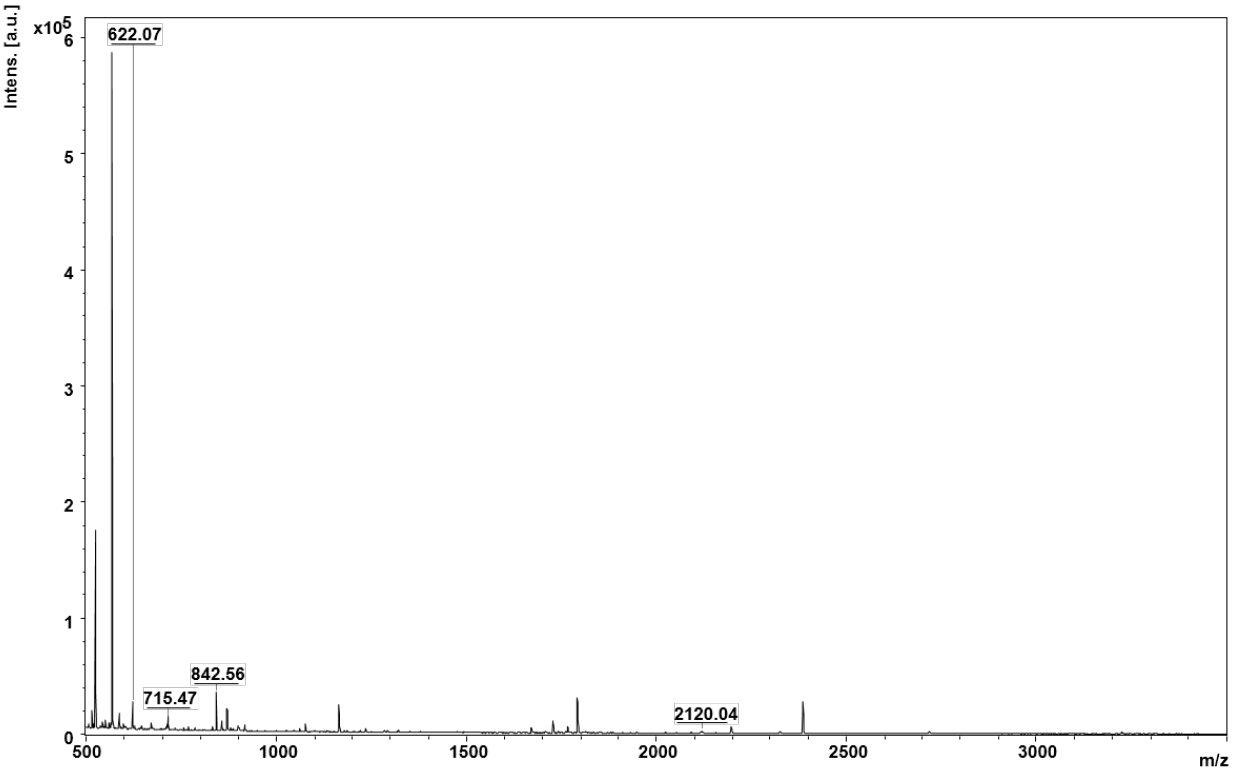


Figure 129: Mass spectrum of sample 7C of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

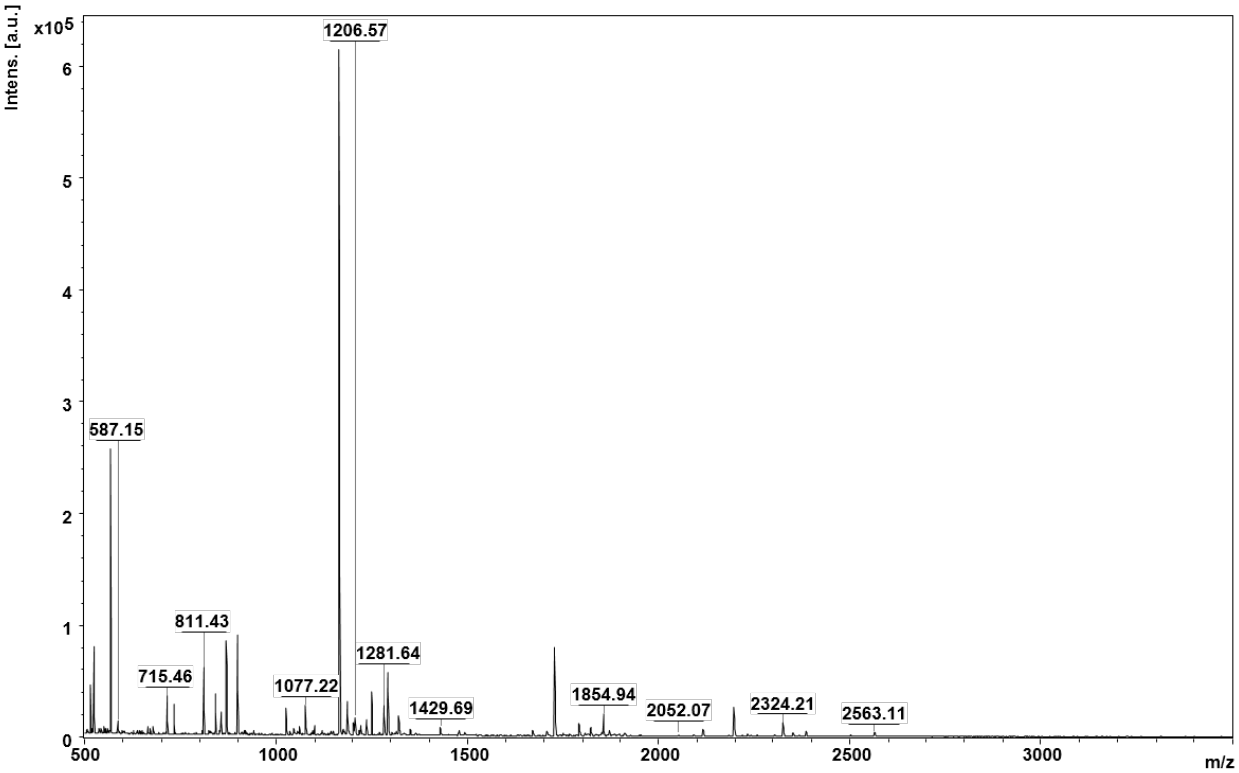


Figure 130: Mass spectrum of sample 7DE of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

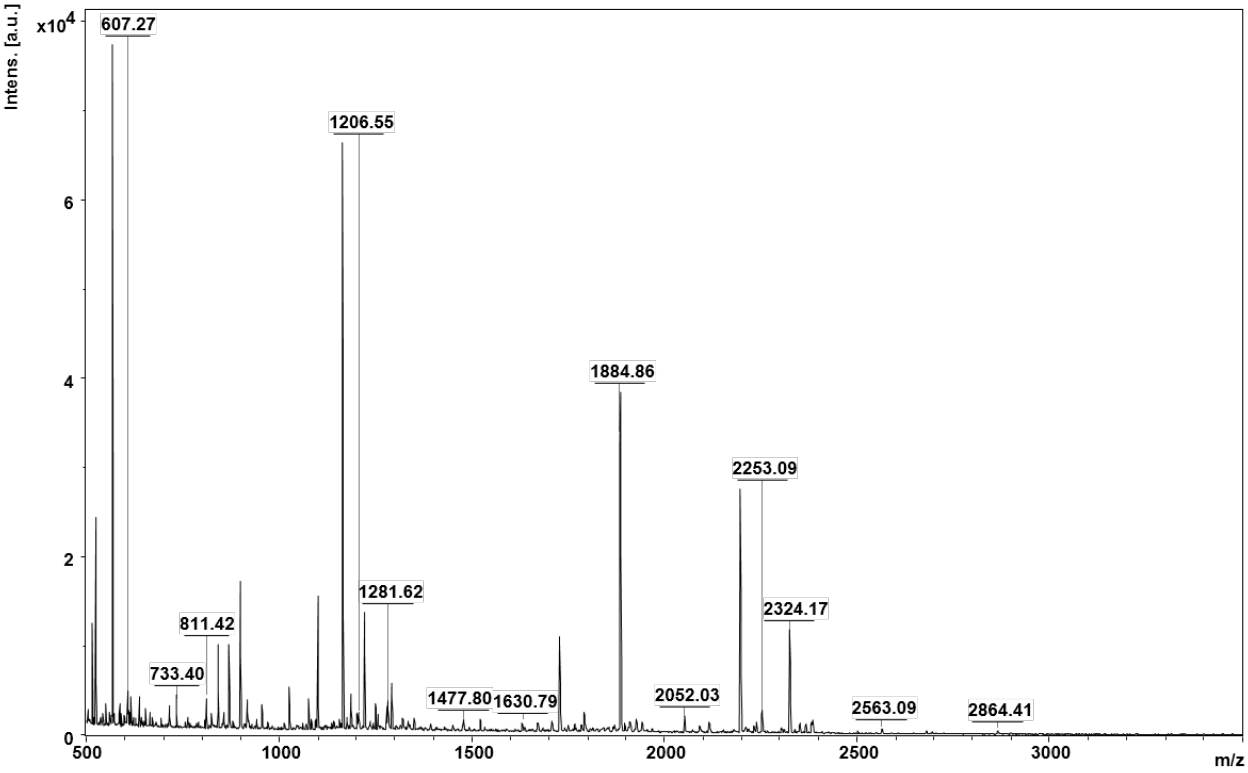


Figure 131: Mass spectrum of sample 7FGH of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

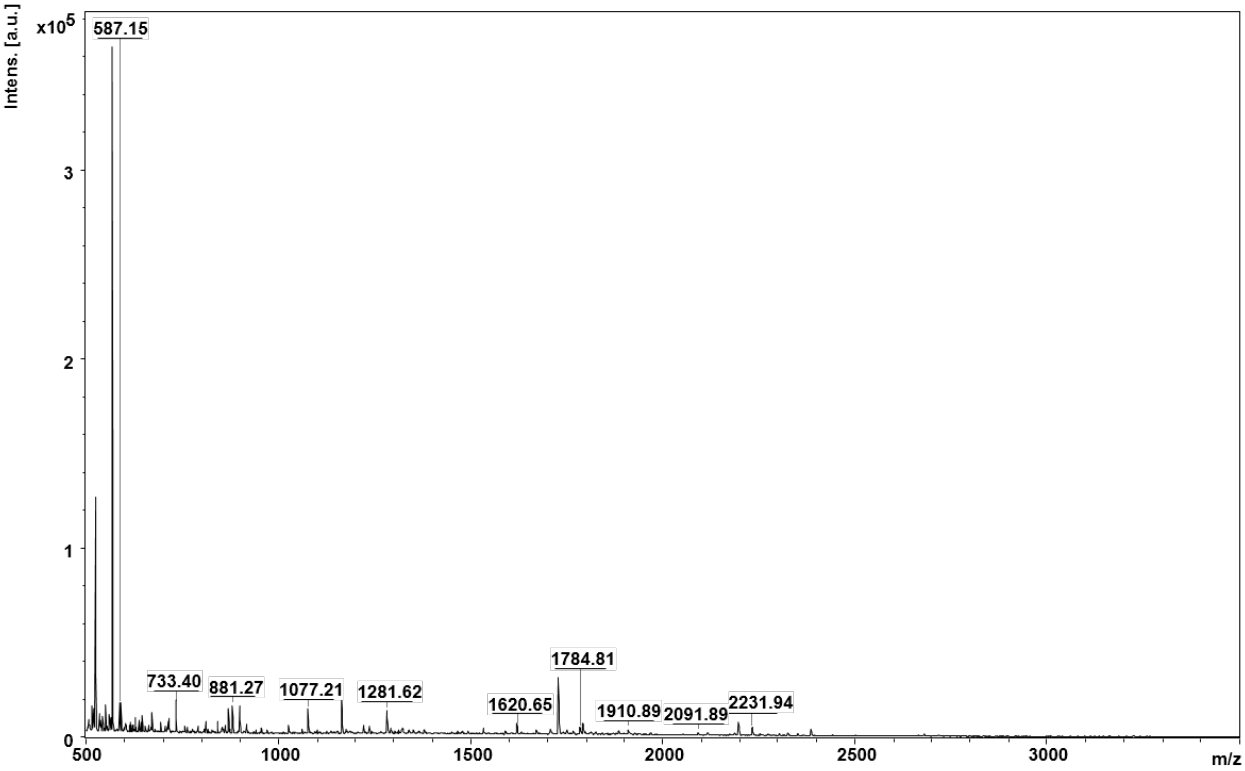


Figure 132: Mass spectrum of sample 7A of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

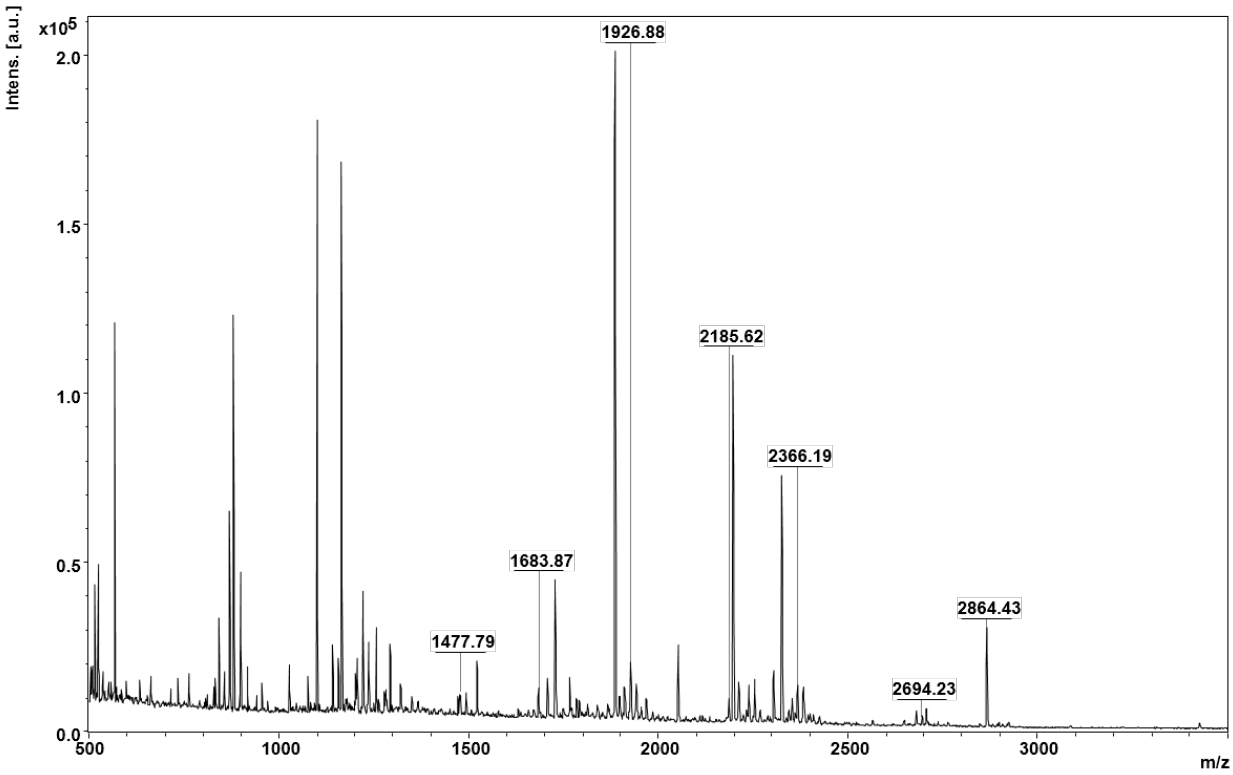


Figure 133: Mass spectrum of sample 7BC of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

None of the mass spectra yielded any significant results. No MS^2 spectra of relevance could be obtained for the Gel band 7, therefore, no information about the peptides and an accurate identification could be achieved.

4.5.8 Gel Band 8

In the figures 134-140 the mass spectra of the filtered peptide peaks are shown. The peaks used for the search in Mascot are listed in the Appendix (see Chapter 7.1.8).

The mass spectra show a high diversity in the peak pattern over the whole m/z range, from low intensity and low peak numbers to high intensity and high peak numbers. Especially samples 8DE, 8FGH and 8BC* show high peak numbers over the whole m/z range. Therefore, the possibility of a reliable identification of the digested protein is rather low.

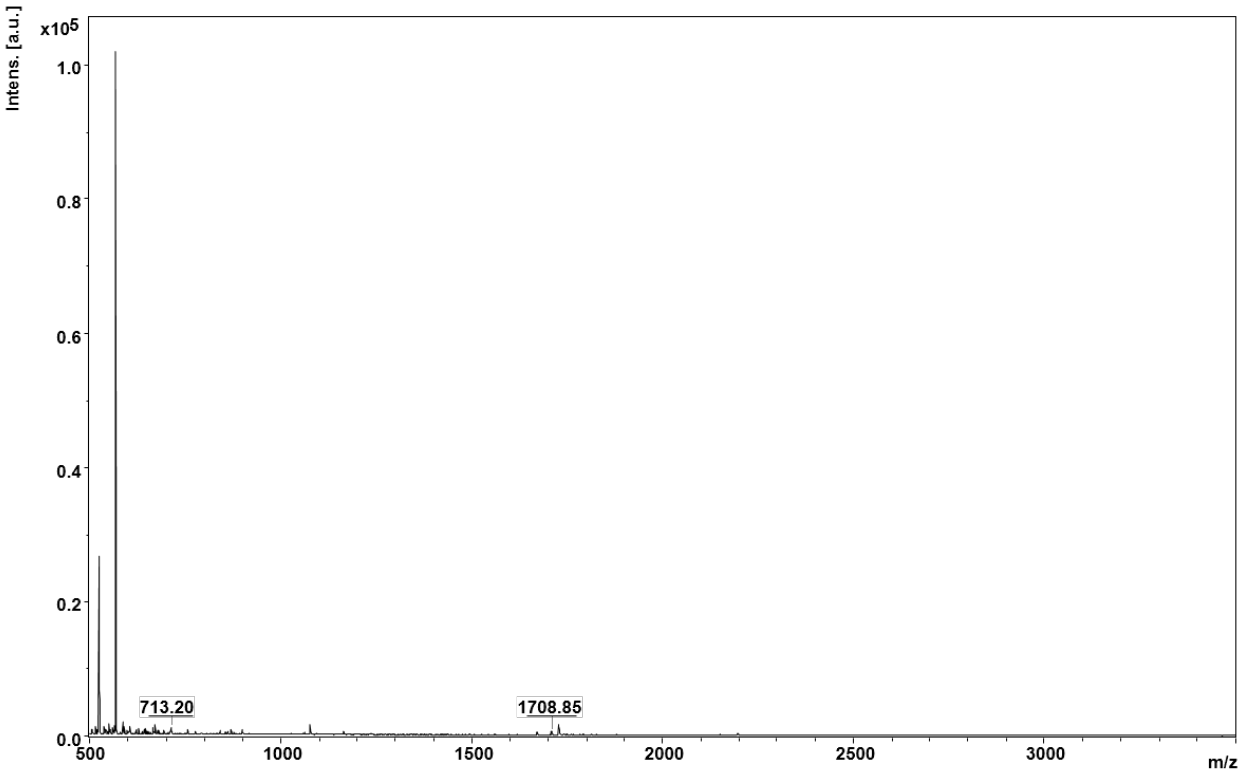


Figure 134: Mass spectrum of sample 8A of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

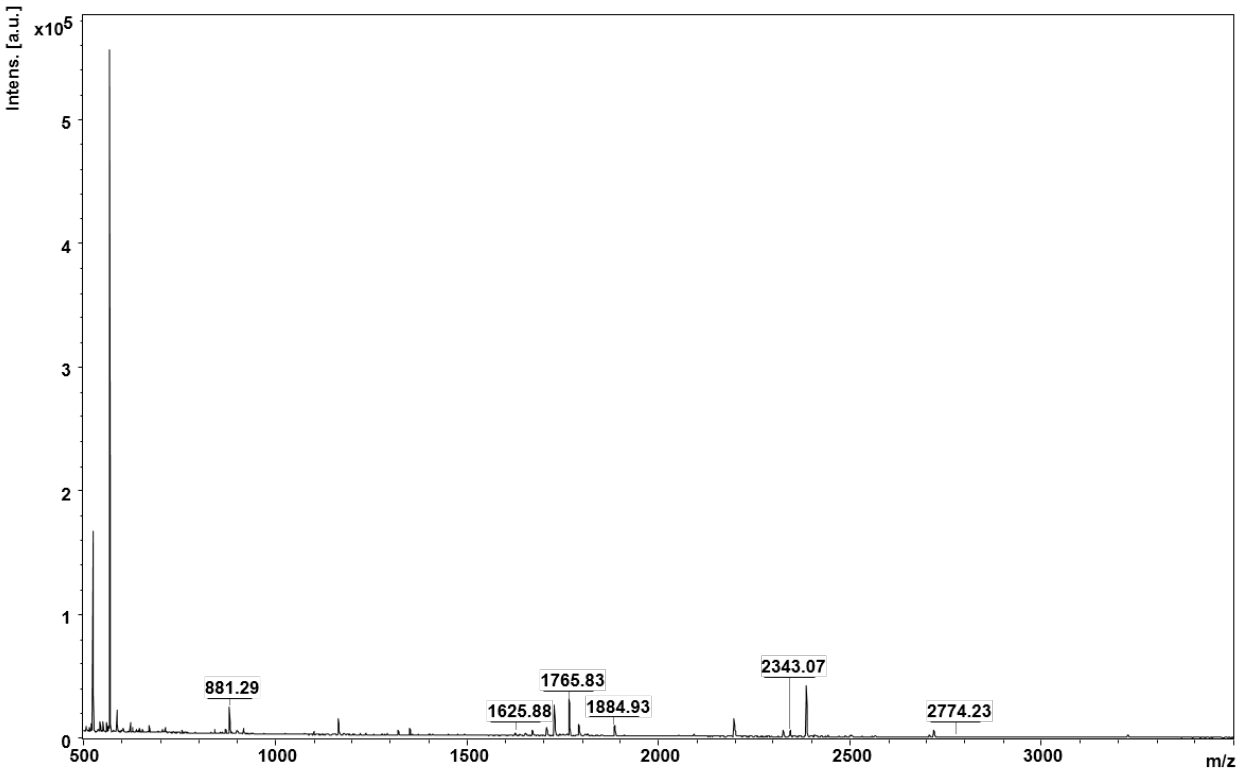


Figure 135: Mass spectrum of sample 8B of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

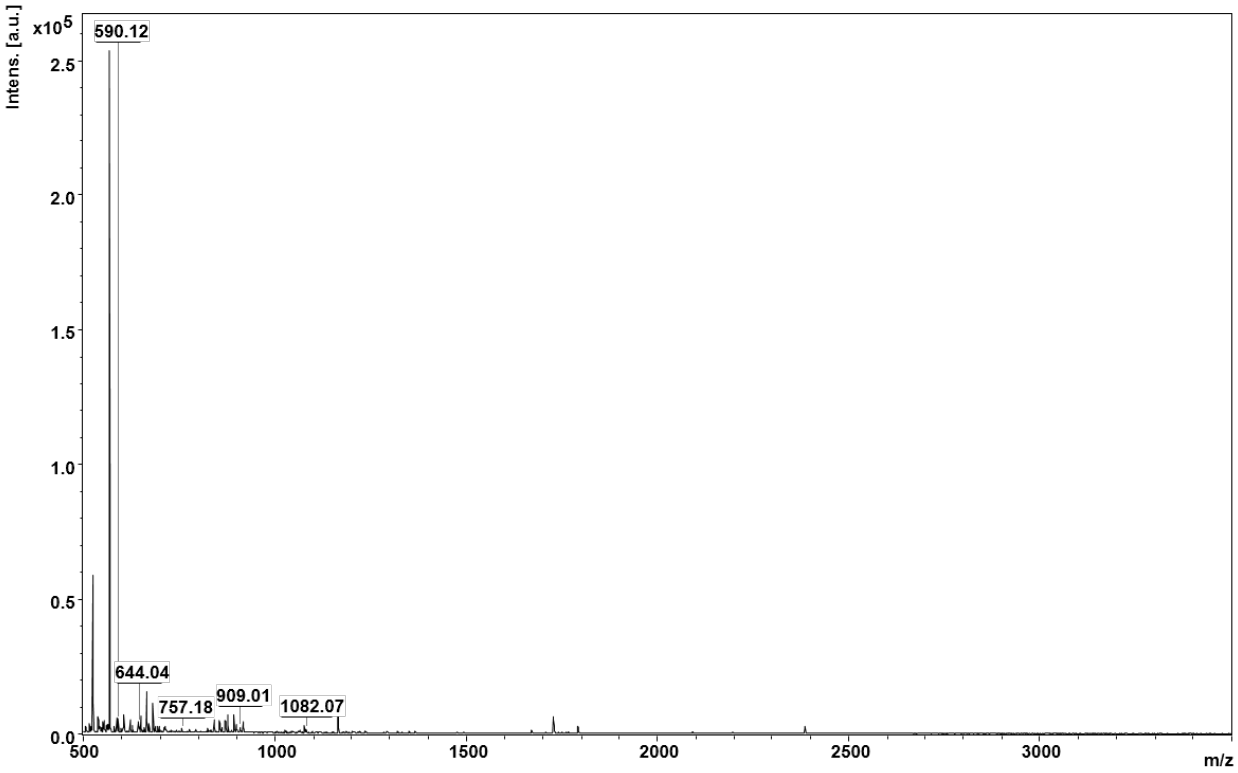


Figure 136: Mass spectrum of sample 8C of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

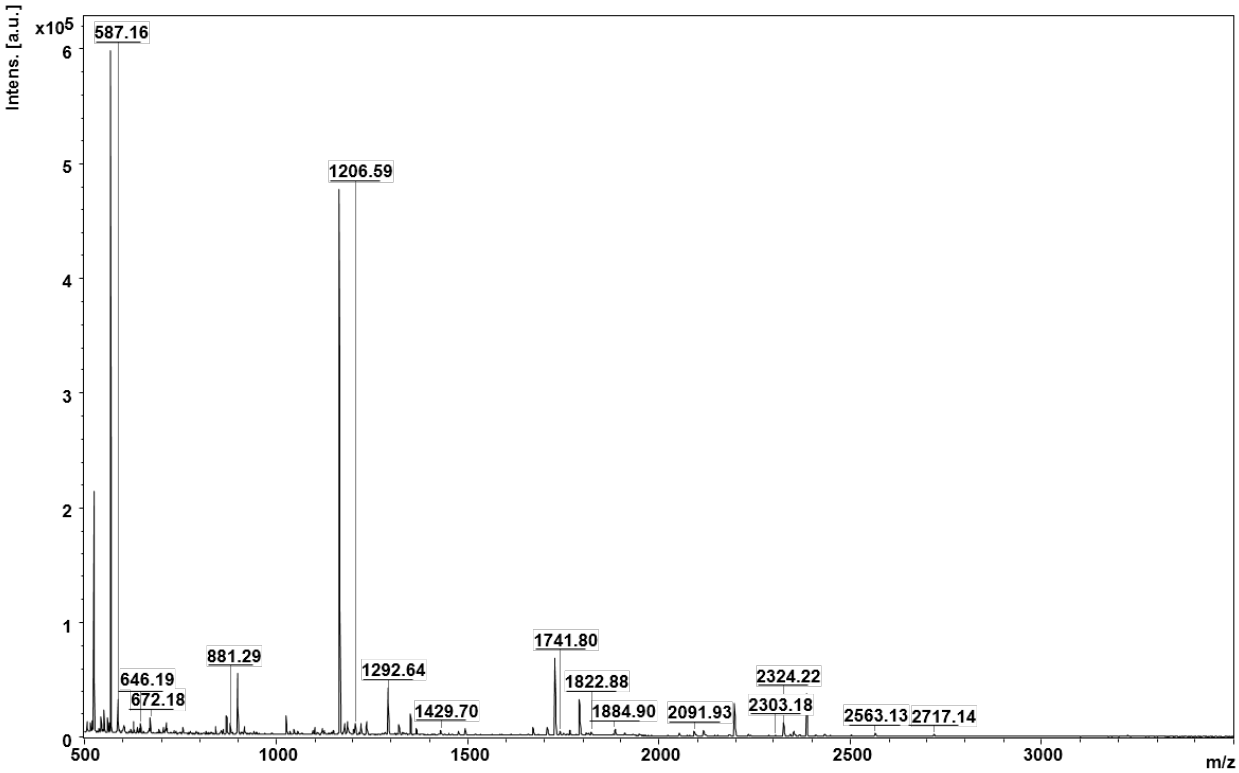


Figure 137: Mass spectrum of sample 8DE of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

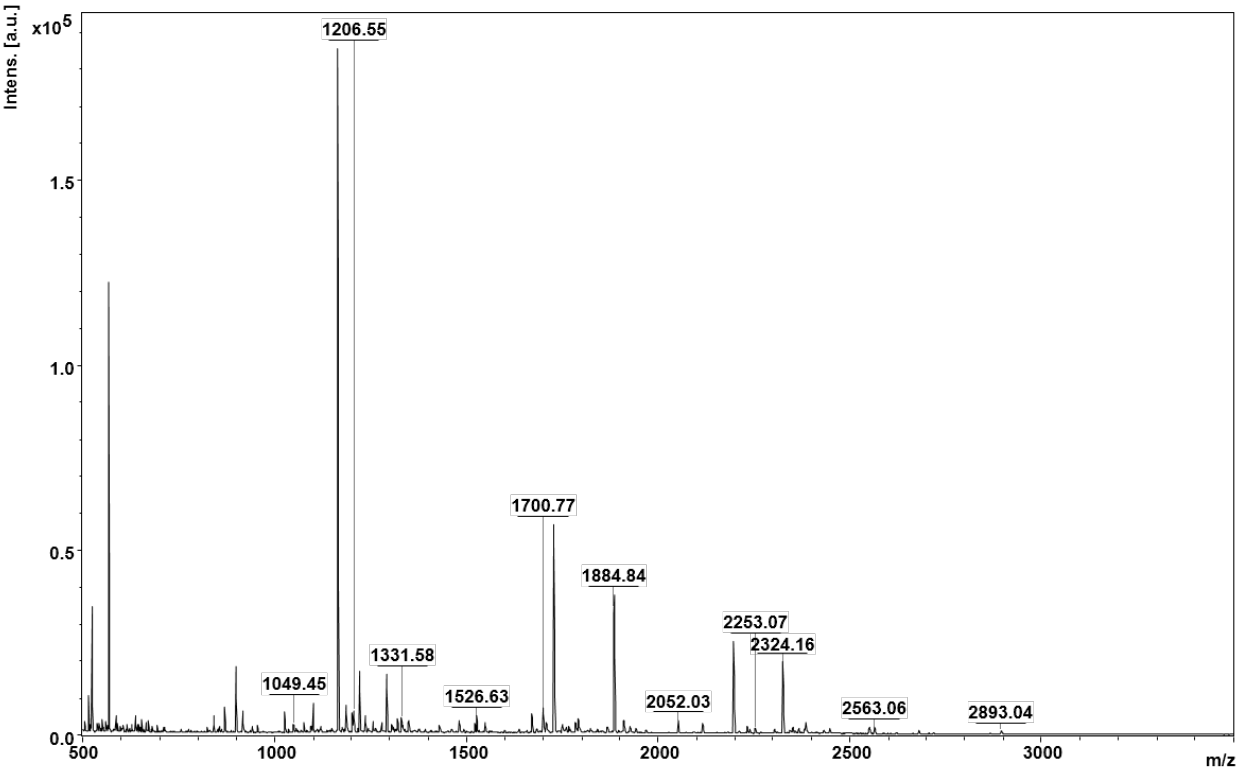


Figure 138: Mass spectrum of sample 8FGH of Gel 1. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].

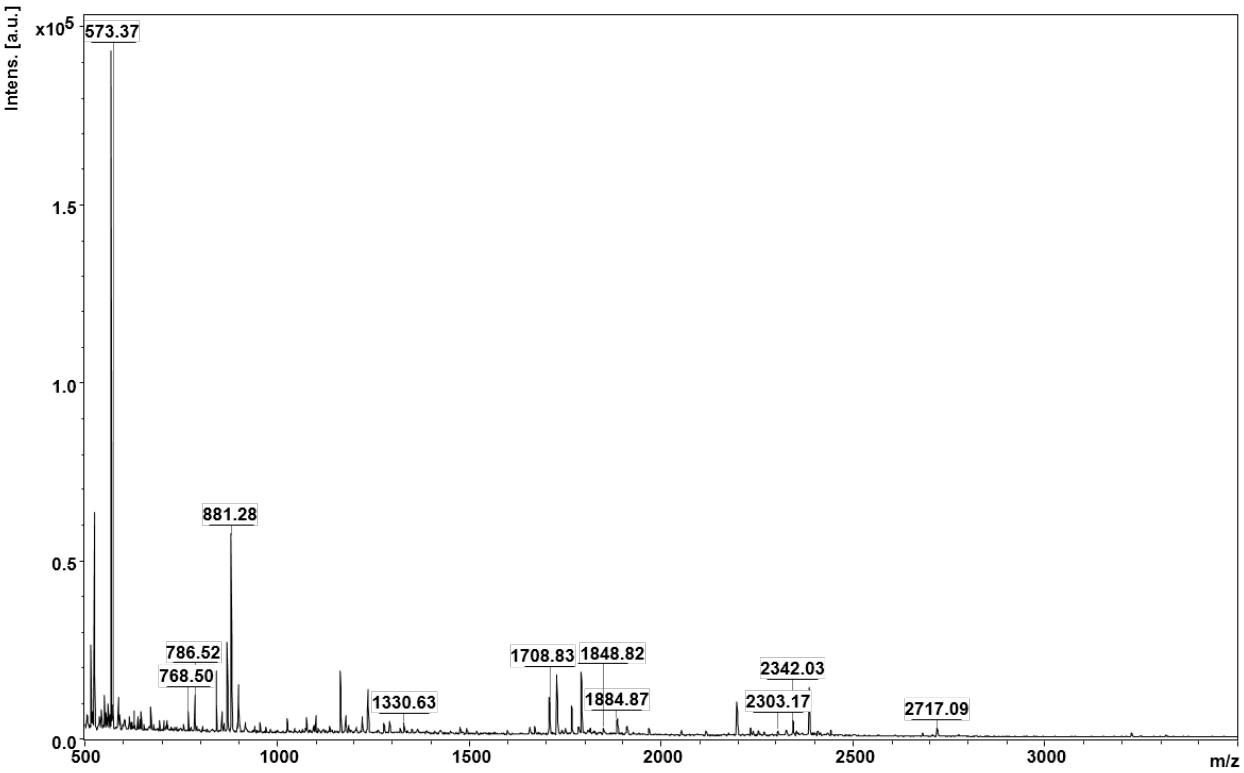


Figure 139: Mass spectrum of sample 8A of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

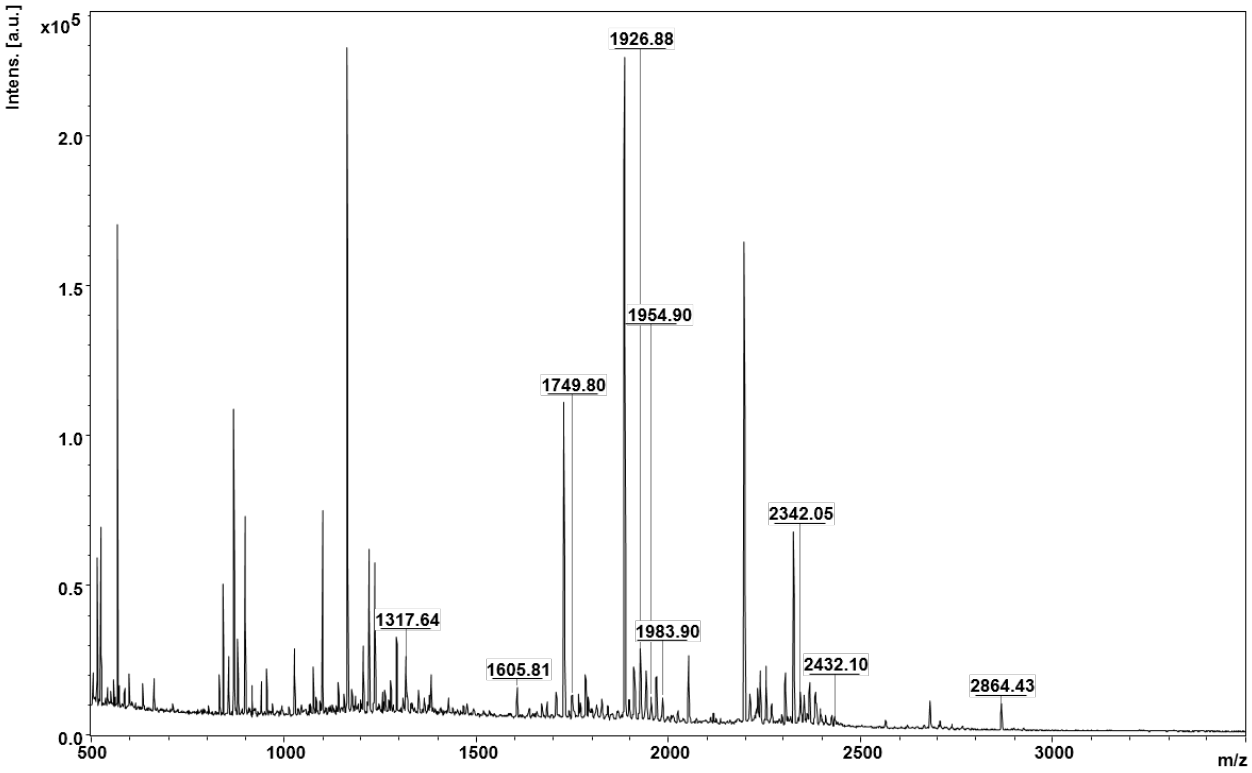


Figure 140: Mass spectrum of sample 8BC of Gel 2. The listed peaks in the figure are part of the ones used in the Mascot search. The m/z values of the peaks are plotted against the intensity in arbitrary units [a.u.].*

The Mascot search results are listed in table 40. Only mass spectra with a relevant theoretical weight, similar to the estimation from the gel band (20 kDa) are shown. Due to the proteins being isolated and extracted from *M. domestica*, only results from this species were considered for the protein identification.

No MS^2 spectra of relevance could be obtained for the Gel band 8, therefore, no information about the peptides and an accurate identification could be achieved.

For the Gel band 8 only sample 8DE could be identified with a relevant molecular weight and a significant probability. The identified protein is an uncharacterized protein from *M. domestica*, which makes it a very likely candidate, but no further information about the protein is known. In addition, the protein hit could not be confirmed with any of the other samples, thus rendering the result not very reliable. It is possible that Gel band 8 consists of multiple proteins and this uncharacterized protein could be one of them.

Table 40: Mascot search results of Gel Band 8 for all samples. Only significant hits with relevant theoretical molecular weights, compared to the rough estimate of the gel bands, are listed.

Sample	Taxonomy	Name of protein and Species	Matches	Score/ Thres-hold	th. MW [Da]	th. pI	Protein coverage [%]
8DE	Rosaceae	Uncharacterized protein LOC103401616 [*Malus domestica*] XP_008338550.1	10/41	70/68	22539	5.39	45

4.5.9 Identification Overview

For a better overview and comprehension all the significant protein hits obtained with Mascot search are summarized in table 41.

From 8 Gel bands only 3 Gel bands could be identified with certainty. Most likely the Gel bands each consist of multiple proteins and thus complicates the accurate identification. The protein of interest, putative *p*-coumaroyl-CoA reductase, could not be identified.

The extraction of the gel bands with a more hydrophobic approach did not yield any results at all. Most likely more tryptic self-digested peptides and membrane protein and similar proteins were extracted instead of the protein of interest.

Table 41: Protein identification of the Gel bands.

Gel Band	Protein name [Species]	Protein ID number	th. MW [Da]	th. pI	pract. MW [Da]
1	alpha-xylosidase 1-like [*M. domestica*]	XP_008368212.1	103503	5.51	95
	alpha-xylosidase 1 [*M. domestica*]	XP_008368212.1	103779	5.93	95
	PREDICTED: alpha-xylosidase 1-like [*M. domestica*]	XP_008368212.1	103616	5.51	95
2	cucumisin-like isoform X1 [*Malus domestica*]	XP_008351270.1	79768	6.76	72
	cucumisin-like isoform X2 [*Malus domestica*]	XP_008351271.1	76338	6.74	72
	cucumisin-like, partial [*Malus domestica*]	XP_017184694.1	64987	8.39	72
3	Beta-glucosidase 12-like precursor [*Malus domestica*]	NP_001280983.1	61259	5.56	59
	Beta-glucosidase 12-like [*Malus domestica*]	XP_008357590.1	61521	5.88	59
	Beta-glucosidase 12-like isoform X1 [*Malus domestica*]	XP_008384719.1	58993	6.57	59
	Beta-glucosidase 12-like isoform X2 [*Malus domestica*]	XP_017191209.1	47656	7.19	59
4	No identification				
5	No identification				
6	No identification				
7	No identification				
8	No identification				

5. Conclusion and Outlook

In this thesis, the characterization of proteins derived from the common apple, *M. domestica*, was performed using a PMF approach with MALDI-RTOF and MALDI-TOF-RTOF mass spectrometry.

The digested peptides require dedicated sample preparation via ZipTip prior to MS analysis. The protocol for the *in-gel* digestion as well as the ZipTip purification are already known and established methods used for peptide sample preparation in the proteomics field [31].

For the characterization eight Gel bands with various molecular weights, with six replicates and one additional experiment with an extraction protocol for hydrophobic peptides[33], were excised from the gels and digested.

Three of the eight Gel bands could be accurately identified with Mascot search. Unfortunately, the distribution of the protein bands and their broadness in the gels, indicates the presence of multiple proteins or proteoforms with similar molecular weight in each Gel band. Hence, during the *in-gel* digestion peptides from various proteins are produced, which leads to a high number of peaks in the mass spectra. Not all of those peaks belong to the protein of interest, i.e. one protein, and are complicating the Mascot search. This observation interferes with the identification of the target protein in the gel bands.

The protein of interest, a putative *p*-coumaroyl-CoA reductase, could not be identified. This enzyme could still be within one of the gel bands but due to the amount of the other proteins and their digested peptides the relevant signals could be overlapped by the others. Furthermore, it also depends on the amount of the target protein being present in the gel. As the identification of fractions in question is only based on biological activity, the enzyme concentration is supposed to be rather low. Recorded data from FPLC purifications and Agilent 2100 Bioanalyzer results support this assumption.

A possible way to identify more proteins would be an approach using 2D gel electrophoresis, were the proteins are separated not just by molecular weight but also by their isoelectric point.

All in all, to gain more insight and establish a more efficient way of target protein identification, the methods employed for isolation, extraction and especially the purification of the target protein require further development.

6. References

1. Tu, S.H., L.C. Chen, and Y.S. Ho, An apple a day to prevent cancer formation: Reducing cancer risk with flavonoids. J Food Drug Anal, 2017. 25(1): p. 119-124.
2. Gosch, C., H. Halbwirth, and K. Stich, Phloridzin: biosynthesis, distribution and physiological relevance in plants. Phytochemistry, 2010. 71(8-9): p. 838-43.
3. Gosch, C.H., H.; Kuhn, J.; Miosic, S.; Stich, K., Biosynthesis of phloridzin in apple (Malus domestica Borkh.). Plant Science, 2009. 176: p. 223-231.
4. Gerhauser, C., Cancer chemopreventive potential of apples, apple juice, and apple components. Planta Med, 2008. 74(13): p. 1608-24.
5. Hunter, M.D.H., L. A., Variation in concentrations of phloridzin and phloretin in apple foliage. Phytochemistry, 1993. 34(5): p. 1251-1254.
6. Iqbal, M., et al., Aqueous two-phase system (ATPS): an overview and advances in its applications. Biol Proced Online, 2016. 18: p. 18.
7. Lottspeich, F.E., J. W., Bioanalytics - Analytical Methods and Concepts in Biochemistry and Molecular Biology. Wiley VCH, 2018.
8. Sheehan, D., Fast protein liquid chromatography (FPLC) methods. Methods Mol Biol, 1996. 59: p. 269-75.
9. Sheehan, D. and S. O'Sullivan, Fast protein liquid chromatography. Methods Mol Biol, 2004. 244: p. 253-8.
10. Westermeier, R.N., T.; Höpker, H.-R.; Proteomics in Practice; A Guide to Successful Experimental Design. Wiley VCH, 2008. 2. Edition: p. 120-124.
11. Westermeier, R., Electrophoresis in Practice - A Guide to Methods and Applications of DNA and Protein Separations. Wiley VCH, 2016. 5. Edition.
12. Shapiro, A.L., E. Vinuela, and J.V. Maizel, Jr., Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. Biochem Biophys Res Commun, 1967. 28(5): p. 815-20.
13. Neuhoff, V., et al., Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis, 1988. 9(6): p. 255-62.
14. Hempelmann, E. and K. Krafts, The mechanism of silver staining of proteins separated by SDS polyacrylamide gel electrophoresis. Biotech Histochem, 2017. 92(2): p. 79-85.
15. Heukeshoven, J.D., R., Simplified method for silver staining of proteins in polyacrylamide gels and the mechanism of silver staining. Electrophoresis, 1985. 6: p. 103-112.
16. Shevchenko, A., et al., Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem, 1996. 68(5): p. 850-8.
17. Görg, A., et al., The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis, 2000. 21(6): p. 1037-53.
18. Bousse, L., et al., Protein sizing on a microchip. Anal Chem, 2001. 73(6): p. 1207-12.
19. de Hoffman, E.S., V., Mass Spectrometry - Principles and Applications. Wiley, 2007. 3. Edition.
20. Hillenkamp, F.P.-K., J., MALDI MS - A Practical Guide to Instrumentation, Methods and Applications. Wiley VCH, 2007.
21. http://what-when-how.com/proteomics/time-of-flight-mass-spectrometry-proteomics/.
22. Kaufmann, R., et al., Post-source decay and delayed extraction in matrix-assisted laser desorption/ionization-reflectron time-of-flight mass spectrometry. Are there trade-offs? Rapid Commun Mass Spectrom, 1996. 10(10): p. 1199-208.
23. Wiedenbeck, M.B., R.; Duke, M. J. M.; Zoubir, A., GGR Biennial Critical Review: Analytical Developments Since 2010. Geostandards and Geoanalytical Research, 2012. 36(4): p. 337-398.
24. Laforge, A.C., et al., Three-body dynamics in single ionization of atomic hydrogen by 75 keV proton impact. Phys Rev Lett, 2009. 103(5): p. 053201.
25. Roepstorff, P. and J. Fohlman, Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom, 1984. 11(11): p. 601.

26. Henzel, W.J., et al., Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc Natl Acad Sci U S A, 1993. 90(11): p. 5011-5.
27. Pappin, D.J., P. Hojrup, and A.J. Bleasby, Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol, 1993. 3(6): p. 327-32.
28. Keller, B.O., et al., Interferences and contaminants encountered in modern mass spectrometry. Anal Chim Acta, 2008. 627(1): p. 71-81.
29. ThermoScientific, Pierce™ 660nm Protein Assay. https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0016386_Pierce660nmProteinAssay_PI.pdf&title=Instructions:%20Pierce%20660nm%20Protein%20Assay.
30. Agilent, Agilent Protein 230 Kit Quick Start Guide. https://www.agilent.com/cs/library/usermanuals/public/Protein-230_Assay_QSG_RevC.pdf.
31. Pluskal, M.G., Microscale sample preparation. Nature Biotechnology, 2000. 18: p. 104-105.
32. Shevchenko, A., et al., In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc, 2006. 1(6): p. 2856-60.
33. Asler, I.L.Z., M.; Kovacic, F.; Müller, R.; Abramic, M.; Allmaier, G.; Kojic-Prodic, B., Mass spectrometric evidence of covalent-bound tetrahydrolipstatin at the catalytic serine of Streptomyces rimosus lipase. Biochimica et Biophysica Acta - General Subjects, 2007. 1170(2): p. 163-170.

7. Appendix

7.1 Peak lists

7.1.1 Gel Band 1

Table 42: Peak list of sample 1A used for Mascot search.

m/z value	Intensity
917.30	13354
921.41	8457
1106.52	29009
1695.82	6550
1785.86	2230

Table 43: Peak list of sample 1B used for Mascot search.

m/z value	Intensity
524.15	56610
526.17	10036
1106.52	6431
1695.84	1154

Table 44: Peak list of sample 1C used for Mascot search.

m/z value	Intensity
1062.07	3256
1077.22	7902
1106.53	19663
1695.84	1802
2257.98	817
2502.06	977
2717.11	1795

Table 45: Peak list of sample 1DE used for Mascot search.

m/z value	Intensity
768.51	2738
774.43	3164
786.52	3444
816.97	2066
1061.50	2210
1106.52	17192
1148.53	2341
1695.81	3340

Table 46: Peak list of sample 1FGH used for Mascot search.

m/z value	Intensity
881.27	2059
1106.51	5303
1118.09	2034
1133.27	6317
1176.55	2318
1206.55	4934
1221.57	27488
1292.60	12196
1404.71	3272
1429.66	2701
1436.70	1444
1523.84	1119
1678.79	2717
1695.81	10494
1822.83	1817
1884.86	40595
1896.87	1177
1910.87	1642
1926.85	1189
1941.87	1419
1967.91	785
2052.05	5143
2238.08	1040
2253.08	1648
2303.14	863
2324.17	27944
2351.14	1822
2366.16	1298

2381.17	2041
2563.08	842
2679.17	609
2864.39	445

Table 47: Peak list of sample 1A used for Mascot search.*

m/z value	Intensity
1106.52	10821
1678.79	6461
1695.82	15305
1784.84	4262
1884.86	3289
1910.88	3366
2231.94	2910
2441.00	1151
2679.20	1774
2717.15	1686

Table 48: Peak list of sample 1BC used for Mascot search.*

m/z value	Intensity
522.35	46215
1148.60	62847
1177.78	65105
1248.64	65565
1631.86	55518
1695.86	66591
1765.82	54575
1790.96	67893
2052.10	33094
2081.00	31407
2253.17	25848
2296.25	24755
2399.12	17009
2864.48	7918

Table 49: Peak list of sample 1A_1695 used for Mascot search.

m/z value	Intensity
112.14	62
136.14	80
159.17	35
175.20	243
184.19	29
216.22	29
246.24	55
258.18	66
292.22	46
316.24	51
333.27	119
354.28	23
371.27	69
387.27	20
405.41	23
425.58	19
444.27	19
462.32	28
491.41	21
517.36	28
533.45	46
541.36	17
551.44	18
645.46	31
662.48	25
697.37	17
719.56	107
831.60	20
848.58	20
863.89	18
959.58	21
1034.63	25
1125.98	26
1134.18	26
1315.10	75
1438.82	61
1504.04	101
1590.89	39
1612.62	43
1632.52	47
1635.86	37
1641.33	39

1646.38	50
1650.67	56
1656.20	56
1660.21	35
1695.83	131

7.1.2 Gel Band 2

Table 50: Peak list of sample 2A used for Mascot search.

m/z value	Intensity
813.44	50679
938.53	39333
1087.56	13518
1178.67	49661
1409.73	24522
1511.78	55116
1562.79	3649
1939.96	3910
2052.06	2158
2077.89	2399
2324.17	6098
2351.14	1873
2563.09	1295
2679.14	1512
2717.10	2176

Table 51: Peak list of sample 2B used for Mascot search.

m/z value	Intensity
813.44	15905
938.53	18382
1000.46	3633
1077.20	4765
1087.55	8776
1135.59	4073
1178.65	5911
1349.61	1925
1378.70	2161
1409.71	4331
1511.77	7060

1562.77	1794
2231.93	1502
2679.13	734
2717.08	725

Table 52: Peak list of sample 2C used for Mascot search.

m/z value	Intensity
716.44	9295
813.46	30398
938.54	20111
951.50	6451
1018.53	4838
1077.22	6970
1087.57	15656
1107.57	4782
1178.67	21676
1409.74	9142
1511.79	11897
1561.80	1869
1884.89	3975
1939.97	1664
2304.19	1392
2324.20	4546
2679.17	1046
2717.13	2656

Table 53: Peak list of sample 2DE used for Mascot search.

m/z value	Intensity
643.37	10724
716.42	12207
813.44	16077
829.57	6762
938.54	38013
1022.43	6297
1036.45	5250
1068.57	4186
1087.58	4667
1177.62	8071
1178.69	169087

1186.56	17945
1206.58	3998
1220.69	6479
1250.67	3402
1297.73	7307
1335.75	12807
1351.74	6461
1364.72	3238
1409.75	50394
1429.70	3046
1511.80	28287
1544.80	2753
1560.79	4347
1562.81	17531
1584.80	3796
1600.79	2275
1690.91	3623
1822.86	2157
1884.87	3146
1895.91	1873
1909.92	2198
1939.98	20125
1953.96	27036
1975.96	3991
1991.94	2087
2052.07	1733
2074.98	2440
2077.92	19536
2281.00	2314
2324.20	6823
2351.18	2804
2563.12	1679
2815.24	860
2863.25	9688
2879.22	1156
2961.37	1050
3066.32	948

Table 54: Peak list of sample 2FGH used for Mascot search.

m/z value	Intensity
813.43	2636
938.52	1874
1087.55	1031
1178.66	5298
1409.72	6864
1511.77	10999
1561.78	885
1568.79	763
1583.75	770
1884.85	4176
1939.95	1890
2052.04	979
2238.08	500
2253.09	519
2303.14	573
2324.17	2457

Table 55: Peak list of sample 2A used for Mascot search.*

m/z value	Intensity
665.35	12454
813.44	44586
899.43	16487
938.53	34925
1087.57	13575
1103.61	7082
1178.69	71210
1220.70	12161
1221.59	15584
1235.67	7257
1292.62	8412
1297.73	6287
1364.72	9868
1409.75	108240
1437.74	7747
1451.75	5883
1466.76	10143
1511.80	116259
1562.81	18994
1568.81	10747

1583.78	8181
1727.83	52331
1749.82	4704
1784.85	7608
1884.88	6654
1910.91	6341
1939.98	12730
1953.95	7011
1967.94	5221
2077.92	6576
2231.97	8607
2293.06	3042
2679.18	3363

Table 56: Peak list of sample 2BC used for Mascot search.*

m/z value	Intensity
993.51	39511
1029.52	65405
1042.54	29518
1052.55	17198
1133.56	18568
1207.61	24614
1234.70	156949
1236.64	88732
1252.61	29961
1291.72	46432
1409.73	45229
1422.70	41368
1440.73	35629
1458.76	52462
1485.76	93369
1487.78	421785
1511.79	93390
1526.76	96964
1530.79	315900
1542.79	36165
1544.81	119905
1550.79	92630
1587.80	46496
1638.89	558649
1695.91	101397
1735.81	47598
1806.82	48735

1822.82	48985
1884.88	136788
1941.96	27987
1995.01	98854
1996.98	22492
2025.97	24152
2034.09	34671
2097.01	20026
2212.11	27222
2268.03	27248
2324.18	31394
2343.03	420358
2352.04	27820
2371.03	20467
2399.07	256584
2441.04	22986
2456.08	29484
2558.29	12823
2573.30	12995
2584.25	27618
2643.34	15688
2717.14	22997
2739.13	18566
2776.44	81898
2787.17	16883
2824.44	110969
2866.45	52810
2873.45	108525
2881.46	33356
2915.45	24631
2959.46	12014
2972.47	8849
2989.57	9111
3000.53	12882
3048.55	11800
3084.55	9940
3097.46	7923
3141.57	6363
3515.63	3373
3536.62	3159
3614.60	2833
3623.57	2949
3705.51	4590
3731.67	6064

Table 57: Peak list of sample 2A_1409 used for Mascot search.

m/z value	Intensity
60.06	20
70.06	37
86.05	33
112.11	62
120.11	20
129.13	51
157.16	32
169.14	28
175.16	269
178.08	30
180.11	20
185.16	55
187.14	32
197.17	19
217.76	20
226.16	21
233.21	77
258.19	19
261.19	161
269.20	20
272.18	31
278.21	31
284.22	30
289.18	26
300.23	134
314.21	30
332.26	146
349.27	45
360.27	139
367.22	17
371.28	139
397.25	17
403.29	21
415.21	20
419.28	29
426.37	19
431.34	21
468.36	70
479.33	19
489.32	24
496.33	37
523.33	18
583.33	29
632.45	24
638.42	27
663.41	130
680.45	135
730.44	19

762.53	20
779.52	24
1039.68	78
1294.73	450
1332.95	33
1342.34	33
1346.57	33
1348.71	47
1350.70	46
1360.05	34
1363.49	53
1366.73	76
1368.93	54
1371.09	49
1409.73	225

Table 58: Peak list of sample 2A_1511 used for Mascot search.

m/z value	Intensity
60.07	27
70.10	127
86.12	30
112.12	237
120.12	36
129.14	112
136.11	63
158.15	53
175.16	2454
183.19	336
211.19	537
228.21	37
230.15	62
235.16	66
244.14	70
261.16	59
263.19	70
273.16	125
283.21	159
290.20	46
292.19	31
310.26	631
313.21	31
315.20	33
327.23	68
334.22	42
343.22	60
345.23	43
370.24	752

374.28	67
387.26	921
398.20	29
408.26	363
445.35	76
457.38	77
469.30	52
473.35	187
478.33	37
486.34	2822
496.32	55
506.32	98
521.35	168
544.39	68
584.35	69
592.40	48
601.39	86
620.46	121
658.38	83
712.37	27
716.46	136
729.43	93
783.48	30
788.40	31
815.41	35
854.48	35
875.38	32
892.48	64
991.54	56
1026.57	31
1125.72	43
1201.74	107
1511.78	2803

Table 59: Peak list of sample 2C_1511 used for Mascot search.

m/z value	Intensity
70.09	97
86.12	46
112.12	173
129.15	99
136.11	69
158.14	40
175.16	1428
183.19	188
197.15	61
211.18	356
230.15	51

235.16	56
244.16	35
261.17	48
263.18	70
273.17	70
283.21	103
290.20	48
310.27	429
315.22	32
317.22	31
328.24	36
334.23	41
341.30	44
343.22	36
370.23	462
374.26	62
387.27	585
408.26	267
425.17	32
445.33	60
457.34	55
473.36	107
486.34	1439
493.30	33
506.30	59
521.34	79
544.40	38
584.35	63
592.41	43
601.39	63
620.40	98
658.41	55
711.18	29
716.48	113
729.44	48
783.44	40
892.51	38
1201.68	71
1447.57	70
1451.78	48
1463.34	81
1466.58	80
1468.77	119
1473.26	99
1475.46	52
1511.80	134

*Table 60: Peak list of sample 2A*_1409 used for Mascot search.*

m/z value	Intensity
57.99	343
70.03	2425
73.90	515
86.07	2799
100.06	438
112.07	4525
115.07	987
120.06	1091
129.09	1829
136.06	740
141.08	493
154.06	574
158.08	1127
167.07	620
169.09	1202
175.11	9675
178.04	964
185.09	1958
187.05	1279
192.08	310
197.11	709
201.09	306
207.08	398
210.05	331
229.08	352
233.12	2274
239.16	397
245.11	929
256.15	408
261.11	4485
272.13	652
278.07	640
284.13	654
300.15	3883
304.14	481
310.15	376
314.14	563
316.17	723
332.16	2643
349.17	604
360.18	1422
371.18	2148
403.21	337
419.17	299
431.20	301
468.23	586
548.18	241
663.28	1172

680.30	1130
730.34	306
762.30	319
779.40	502
1022.51	1338
1039.55	3868
1110.63	1871
1223.77	1638
1258.42	724
1276.76	693
1294.81	63734
1332.05	729
1335.69	619
1338.94	874
1342.25	1253
1344.65	3630
1348.76	4356
1355.94	1556
1359.15	2238
1361.31	3243
1364.58	4312
1367.80	2061
1370.08	523
1375.28	967
1379.38	2918
1383.29	429

*Table 61: Peak list of sample 2A*_1511 used for Mascot search.*

m/z value	Intensity
70.06	2090
73.86	316
86.06	563
112.07	3161
115.07	369
120.06	604
129.09	1353
136.05	1049
154.06	321
158.06	524
175.09	22694
178.01	584
183.12	3190
197.09	325
211.11	4781
230.08	434
235.08	513
244.07	429

255.05	243
261.08	436
263.10	478
273.07	777
283.11	892
290.09	289
310.17	2723
327.12	284
343.12	263
370.12	1891
374.15	259
387.14	2240
408.14	1043
445.21	263
473.21	294
486.19	3290
506.18	232
521.19	301
538.21	204
620.23	211
658.27	169
716.25	307
729.26	186
892.36	328
991.41	424
1026.42	234
1125.50	278
1201.64	1774
1364.80	977
1404.76	927
1417.21	392
1433.27	393
1438.77	439
1441.07	391
1444.23	582
1446.31	586
1448.51	625
1450.62	543
1453.94	344
1456.85	557
1460.07	1208
1462.20	1386
1464.40	1311
1466.56	1841
1469.82	1358
1473.21	346
1477.19	391
1480.34	523
1482.51	720

7.1.3 Gel Band 3

Table 62: Peak list of sample 3A used for Mascot search.

m/z value	Intensity
550.14	10628
587.17	17979
622.07	26909
715.47	13060
842.56	35633
856.57	9292
917.34	7241
1179.68	3250
1379.71	2626
1930.99	1479
2120.04	2292

Table 63: Peak list of sample 3B used for Mascot search.

m/z value	Intensity
617.13	14115
653.37	9604
881.28	26101
882.46	7662
892.48	128499
917.29	7586
924.47	29422
938.52	10289
944.50	17667
967.44	71974
1015.44	59958
1024.46	7133
1031.43	5817
1058.45	12048
1062.04	6318
1072.46	7892
1077.20	22000
1085.54	6209
1130.57	6211
1142.57	11923
1178.64	9919
1192.63	10458
1202.50	13754
1206.55	6870
1214.63	4982
1292.60	20172
1349.63	3119
1385.65	3740

1409.71	5670
1429.66	3167
1511.76	8696
1533.66	2842
1555.77	2301
1561.74	1591
1584.74	1943
1605.76	4776
1623.77	3228
1673.80	5647
1727.80	22588
1733.85	3376
1745.82	4222
1884.85	3099
1910.87	2502
1947.91	1301
1953.94	1288
2155.87	1406
2185.90	1315
2231.93	3711
2253.06	1654
2260.98	1641
2303.15	1604
2324.16	6043
2351.13	1629
2405.91	1107
2440.98	1005
2451.14	1288
2468.13	2189
2563.06	1485
2596.22	2174
2679.12	2031
2717.08	3043

Table 64: Peak list of sample 3C used for Mascot search.

m/z value	Intensity
530.28	10320
637.35	15834
831.38	9213
881.29	6520
882.47	15414
892.49	488904
896.49	12643
904.48	6989
908.49	6918
914.47	11512
918.50	5865

924.48	49897
934.50	30773
938.53	15962
944.51	29863
949.51	18628
967.45	87274
976.50	5063
981.50	4535
986.52	5317
989.43	4335
1001.53	4847
1009.47	5791
1015.45	81653
1024.47	12983
1031.44	6194
1058.46	17557
1072.47	14756
1085.56	7892
1115.53	5358
1130.58	8808
1142.58	12609
1144.60	6023
1156.60	5696
1178.67	11046
1192.65	27416
1202.51	6659
1206.57	11842
1214.64	7462
1216.66	8716
1221.58	22387
1230.65	5484
1249.67	6713
1292.61	16474
1297.72	4513
1337.67	5345
1344.76	5256
1349.64	3600
1385.68	6647
1409.73	11744
1441.76	3281
1453.76	3031
1511.79	24858
1555.78	3668
1561.78	3028
1569.80	2502
1583.76	1824
1605.78	8482
1623.80	6019
1673.83	34902
1695.81	1914

1699.83	1864
1727.82	34741
1733.87	7280
1745.85	28787
1751.88	4907
1765.83	2930
1784.84	3872
1802.87	2200
1808.90	2190
1817.84	2034
1822.87	2095
1862.94	1845
1884.88	21381
1906.84	2220
1910.90	2794
1926.89	1932
1939.96	3314
1967.94	2193
2052.07	2635
2062.03	11716
2063.98	2712
2090.05	2239
2104.07	1901
2134.05	12956
2157.02	1616
2191.08	1880
2231.96	4729
2237.14	4557
2253.11	2093
2303.17	1780
2324.20	28144
2336.19	1284
2351.18	2114
2366.20	2743
2381.21	2830
2450.21	1038
2468.17	2760
2563.11	1176
2596.26	2285
2679.17	2177
3525.59	975

Table 65: Peak list of sample 3DE used for Mascot search.

m/z value	Intensity
530.29	16368
637.34	6726
642.40	10856
665.35	2798
684.41	2559
704.10	2586
815.44	2319
846.48	2857
882.47	2703
892.49	46608
908.49	2320
924.48	11811
938.53	21576
944.51	4948
967.45	4607
976.50	2097
997.44	2890
1011.52	3979
1015.46	204836
1025.49	2616
1031.45	6099
1037.42	3848
1057.47	7116
1073.57	30285
1085.56	4109
1087.58	10635
1101.57	3304
1102.43	3852
1107.57	3081
1115.58	5905
1130.59	4318
1144.61	3543
1160.55	4031
1178.69	12265
1192.66	27243
1214.65	4241
1292.63	16932
1385.70	3362
1409.74	5862
1429.69	4038
1441.75	14236
1455.76	2906
1469.76	2224
1482.75	2299
1489.78	3260
1511.79	7712
1555.80	9031
1562.80	3612

1569.81	2196
1584.78	1658
1605.79	6132
1623.80	5194
1673.83	3488
1733.88	14433
1745.85	2600
1751.90	10498
1814.81	4426
1822.87	1629
1910.90	892
1930.91	726
1939.97	1729
1947.95	892
1953.95	5143
2062.02	813
2134.07	883
2146.99	1230
2237.15	3671
2324.20	3819
2351.16	771
2450.21	635
2468.17	1593
2563.10	1167
2596.26	2906
2717.12	619
2863.24	1589

Table 66: Peak list of sample 3FGH used for Mascot search.

m/z value	Intensity
530.28	5274
642.40	3015
815.43	1451
892.48	34477
924.47	11690
938.52	2251
949.49	2011
967.44	10327
1015.45	7994
1024.46	2178
1058.45	3961
1072.47	2163
1130.58	2283
1142.57	2295
1178.67	7080

1192.64	9915
1206.56	4039
1249.66	2259
1385.67	1309
1409.73	8075
1429.66	2082
1441.75	1537
1466.75	1140
1511.77	15201
1555.78	1610
1568.79	1464
1583.75	951
1605.76	3182
1623.78	2260
1673.80	7337
1675.77	1724
1733.85	4443
1745.83	8009
1751.85	2591
1802.83	1241
1808.88	1424
1822.82	1404
1884.86	37297
1910.87	2138
1926.86	1835
1939.95	5527
1953.90	923
1967.90	1138
2052.04	3655
2062.00	4004
2104.06	862
2134.03	4813
2191.04	898
2237.10	2117
2253.09	1722
2294.10	640
2303.15	960
2324.17	15202
2336.16	580
2343.01	842
2351.14	1196
2366.15	1044
2381.19	1581
2468.19	639
2563.08	563
2596.22	528
2679.14	664
2864.39	462
2896.33	301

Table 67: Peak list of sample 3A used for Mascot search.*

m/z value	Intensity
829.25	6695
881.28	24759
892.49	106316
924.48	23334
934.50	9652
938.53	7303
949.51	10428
967.45	25026
1015.46	23205
1024.47	7947
1058.46	15742
1072.48	7865
1115.49	6033
1130.58	6644
1142.59	8777
1178.67	16767
1192.66	24233
1249.68	10054
1297.73	7513
1306.71	5065
1364.72	5341
1385.69	4213
1409.75	33770
1466.78	5626
1511.79	43592
1526.82	3588
1555.79	6906
1568.81	5875
1583.78	5088
1605.79	8826
1623.80	8314
1673.83	25285
1675.81	8151
1733.88	10490
1745.85	29632
1751.88	8103
1784.85	7256
1802.88	6460
1812.89	5428
1884.89	32572
1910.91	7991
1939.98	13222
1967.94	7721
2052.07	4093
2063.03	8874
2090.08	3993
2104.10	4238
2134.07	12685

2191.09	3428
2206.09	3349
2231.98	5734
2237.13	3680
2253.12	3177
2303.19	2259
2324.21	12404
2343.03	2545
2351.21	3195
2366.23	2915
2381.22	2699
2468.17	2979
2596.27	2289
2679.18	3459

Table 68: Peak list of sample 3BC used for Mascot search.*

m/z value	Intensity
892.47	102869
967.43	46965
1015.44	19030
1024.46	17808
1058.45	16638
1130.57	23270
1142.57	24025
1178.65	28790
1192.64	43348
1249.66	18817
1409.73	49976
1511.77	60640
1555.77	18766
1568.78	12116
1583.77	9560
1605.77	15764
1623.78	14610
1673.81	38893
1683.88	13752
1733.85	16929
1745.83	41351
1895.93	19644
1926.88	20220
1939.94	19147
1941.88	14337
1996.95	7235
2009.96	7968
2063.01	12023
2104.06	9630
2134.04	16871

2343.02	10253
2399.05	9196
2423.21	5308
2441.00	5298
2456.06	5219
2468.15	6115
2502.07	5244
2563.09	3854
2596.25	5200
2736.20	3201
2763.18	2728
2864.42	7129
3368.52	1620
3424.53	1555

Table 69: Peak list of sample 3A_1192 used for Mascot search.

m/z value	Intensity
77.02	25
86.10	26
88.14	20
105.69	19
110.07	86
117.11	18
122.02	31
136.09	20
163.04	19
166.03	17
175.09	49
178.02	26
197.10	17
207.07	16
213.13	26
225.68	16
230.11	17
248.14	51
253.09	56
266.11	29
270.20	19
284.14	21
290.16	22
299.23	20
308.17	125
335.17	21
360.21	38
362.20	27
375.18	22
400.19	39

407.19	18
411.22	51
423.23	254
428.17	21
446.07	20
457.17	21
471.20	28
475.25	27
499.25	70
524.29	106
532.25	20
542.26	74
546.17	30
549.28	23
570.27	385
584.29	61
612.38	51
623.35	71
641.33	80
644.90	28
660.30	25
669.34	583
754.46	29
770.52	138
782.53	466
885.82	88
917.71	41
920.84	33
945.87	251
1023.02	144
1028.94	44
1047.05	114
1049.71	61
1051.75	37
1132.86	31
1144.51	34
1146.63	49
1148.75	61
1150.99	45
1154.48	52
1157.82	55
1160.67	28
1163.09	118
1192.65	387

Table 70: Peak list of sample 3A_1511 used for Mascot search.

m/z value	Intensity
70.09	87
86.14	28
112.12	128
127.13	102
129.14	92
136.13	36
175.16	1218
183.18	148
211.18	239
235.16	42
244.09	28
263.20	26
273.17	68
283.19	100
290.20	43
292.15	25
310.26	376
315.20	34
327.25	34
343.21	54
345.25	28
352.19	36
354.20	26
370.24	341
374.26	32
387.26	448
408.26	204
445.36	54
469.32	40
473.35	87
486.35	1222
496.33	27
506.29	60
521.37	65
527.25	27
544.38	30
584.35	42
592.47	24
596.63	22
601.36	43
620.43	64
658.42	40
692.50	23
716.41	73
729.43	40
755.43	24
783.50	21
815.59	28

892.51	30
991.60	29
1125.71	44
1200.72	67
1336.61	204
1338.66	149
1356.70	174
1451.68	71
1464.52	83
1468.72	105
1473.22	52
1511.78	1147

Table 71: Peak list of sample 3A_1623 used for Mascot search.

m/z value	Intensity
74.36	17
110.11	43
112.11	27
120.58	15
123.14	21
128.69	17
131.29	15
145.06	16
155.18	19
158.11	15
171.09	21
175.14	32
184.11	15
188.06	16
191.02	13
195.15	17
204.77	13
223.19	30
229.16	13
232.14	14
251.18	70
261.15	16
262.47	20
273.14	14
274.85	14
279.10	13
284.12	16
286.16	19
289.20	16
301.17	25
309.73	12
313.16	16

357.15	17
360.13	17
366.14	12
374.71	14
381.25	14
383.08	15
387.15	14
399.24	13
414.26	37
423.30	18
429.27	40
434.36	13
446.29	16
456.30	21
473.36	19
475.29	16
478.22	17
516.31	15
518.72	17
527.31	39
536.31	14
541.40	16
544.33	40
560.42	131
599.55	15
611.29	29
617.23	18
641.42	14
657.41	26
675.02	12
699.55	14
701.47	17
715.22	13
723.51	78
728.48	16
766.16	25
773.47	18
789.51	26
821.60	14
833.52	14
851.55	68
873.54	34
879.34	25
901.44	20
966.59	48
1046.64	17
1064.59	23

1079.63	25
1150.76	23
1173.65	33
1176.62	27
1201.67	92
1249.73	31
1286.69	36
1314.71	101
1421.82	68
1477.74	74
1499.87	65
1564.31	32
1578.33	31
1585.00	30
1623.79	86

Table 72: Peak list of sample 3B_1192 used for Mascot search.

m/z value	Intensity
70.04	37
72.04	36
86.13	34
110.06	127
141.49	28
147.01	42
154.02	28
175.06	61
178.01	61
184.03	57
235.08	38
248.10	50
253.06	99
266.10	42
286.39	26
299.01	42
308.16	133
335.14	27
362.18	54
394.21	26
400.19	34
411.22	50
423.19	235
499.20	60
524.29	93
527.10	33
542.22	97

546.15	38
570.22	428
584.27	91
612.24	38
623.30	90
641.33	117
669.31	589
683.78	41
754.47	48
770.49	105
782.50	450
885.70	79
920.67	56
945.87	265
1002.56	76
1022.98	141
1047.01	135
1049.64	129
1128.82	56
1141.43	44
1143.34	53
1145.46	110
1148.71	148
1150.86	86
1153.21	100
1155.37	67
1160.29	51
1163.13	209
1192.64	139

Table 73: Peak list of sample 3C_1192 used for Mascot search.

m/z value	Intensity
77.97	27
86.10	34
110.07	223
113.06	27
129.15	29
136.05	29
141.09	28
147.09	32
160.00	32
171.06	27
175.08	81
207.09	32
212.05	28
221.06	29
230.14	36

232.13	31
235.08	47
237.09	25
240.09	30
248.14	111
253.08	120
261.14	29
266.13	45
272.19	25
290.13	36
308.17	302
317.22	27
335.16	24
351.17	29
360.24	52
362.20	74
372.17	29
377.21	31
394.20	34
400.18	112
408.21	37
411.23	123
423.23	596
475.31	57
499.23	131
524.28	277
527.13	45
532.25	38
542.26	175
546.15	68
552.22	39
570.26	866
584.32	185
612.33	78
623.33	155
641.33	208
644.84	63
669.33	1348
683.72	95
754.50	89
757.68	56
770.53	302
782.53	995
885.75	197
917.84	89
920.62	67
945.88	675
1023.03	292
1029.11	99

1047.00	371
1050.58	108
1140.29	62
1146.53	109
1148.79	111
1151.04	104
1153.15	75
1157.77	93
1163.06	124
1192.66	781

Table 74: Peak list of sample 3C_1673 used for Mascot search.

m/z value	Intensity
64.21	34
70.05	48
72.08	38
86.11	128
101.07	63
112.10	278
116.10	30
129.13	125
136.10	156
143.13	34
146.08	35
158.10	107
167.10	29
171.12	68
175.13	1491
184.09	30
205.11	56
209.94	31
228.14	68
235.15	46
242.15	60
244.14	35
251.13	98
263.11	42
268.12	87
278.12	128
286.16	372
299.19	36
303.17	216
318.21	64
323.13	32
355.14	41
357.15	34

377.22	97
394.18	44
399.26	75
416.26	117
432.26	168
448.28	65
456.31	68
473.32	3396
482.27	48
492.23	65
527.30	63
542.36	224
550.35	46
560.35	52
567.38	96
570.34	310
587.42	1570
590.33	51
595.38	125
617.29	48
649.44	45
677.39	51
685.41	87
694.46	49
702.45	106
809.54	73
848.50	32
865.54	1170
868.53	57
963.57	79
980.52	103
1079.60	60
1242.71	113
1339.61	63
1357.80	54
1673.83	6211
1676.03	640

7.1.4 Gel Band 4

Table 75: Peak list of sample 4A used for Mascot search.

m/z value	Intensity
726.41	9296
993.52	3449
1107.55	8221
1516.76	3686
1548.74	1309
2240.95	842
2314.97	873

Table 76: Peak list of sample 4B used for Mascot search.

m/z value	Intensity
917.29	3039
1516.74	1192
1884.88	844
2119.92	1242
2231.94	1100
2324.17	1356
2563.04	531
2679.15	730

Table 77: Peak list of sample 4C used for Mascot search.

m/z value	Intensity
726.41	6777
1103.61	1987
1516.75	2178

Table 78: Peak list of sample 4DE used for Mascot search.

m/z value	Intensity
587.16	14611
726.41	14308
768.52	5350
786.53	5653
829.25	8107
909.48	32380
942.58	5410
1037.61	5061
1103.63	5255

1107.59	4259
1125.57	4606
1232.66	2637
1284.60	3052
1318.73	14738
1338.70	3910
1372.71	4322
1384.58	3312
1398.60	2977
1406.69	2372
1410.73	3465
1432.71	2717
1516.79	2806
1519.76	5996
1539.80	3147
1549.83	2714
1598.82	12359
1649.80	3483
1667.82	27733
1672.85	7137
1869.95	1734
1926.96	1484
1947.94	1435
2091.93	1705
2182.12	1601
2214.05	1100
2262.07	2040
2278.06	1117
2294.02	899
2406.02	753
2528.16	625
2909.35	730
2925.29	524

Table 79: Peak list of sample 4FGH used for Mascot search.

m/z value	Intensity
653.37	6629
726.40	6257
1103.60	3969
1160.65	2096
1232.67	3716
1410.71	1622
1450.73	1954
1516.75	3496
1549.82	2534
1598.77	1580
1866.85	2034

1884.86	74417
1896.85	1713
1906.83	1569
1910.87	1686
1922.80	1183
1926.87	1282
1941.87	2413
1953.88	989
1967.91	876
2052.04	3298
2071.89	837
2238.07	1255
2253.10	4645
2303.14	855
2324.17	20116
2351.14	1126
2381.19	1751
2563.09	1208
2864.40	496
2896.39	475

Table 80: Peak list of sample 4A used for Mascot search.*

m/z value	Intensity
587.16	22980
644.04	10675
694.15	11924
713.20	7958
726.40	21970
861.22	6519
1103.62	5797
1125.55	5276
1186.54	5293
1284.61	4378
1384.57	9734
1516.76	5545
1649.84	4212
1740.76	5352
2051.89	2426
2079.96	2137
2091.91	3043
2231.98	1721

Table 81: Peak list of sample 4BC used for Mascot search.*

m/z value	Intensity
914.57	49804
993.53	61832
1042.56	72565
1107.57	65926
1204.57	50658
1234.72	293764
1252.63	93280
1291.73	61100
1409.75	57404
1422.72	48799
1458.78	55973
1485.78	63223
1487.80	178403
1526.78	89675
1530.81	151317
1535.80	51775
1544.82	52495
1550.81	151729
1562.81	43018
1592.82	29965
1607.83	29591
1638.92	179842
1695.94	38413
1735.83	37447
1763.86	45060
1807.85	33291
1822.85	34765
1995.03	29531
2097.04	20313
2212.14	23561
2343.06	87415
2399.09	39350
2441.06	13055
2457.12	10834
2565.26	8990
2584.28	7957
2739.23	7144
2776.47	13128
2824.46	14715
2864.47	8717
2867.49	11981
2881.49	8422
2916.48	7673
3001.59	5540
3048.56	3990
3053.72	5067
3705.57	3585

7.1.5 Gel Band 5

Table 82: Peak list of sample 5A used for Mascot search.

m/z value	Intensity
646.19	1686
711.19	1034
713.20	1608
757.19	1273
909.47	835
1533.69	677

Table 83: Peak list of sample 5B used for Mascot search.

m/z value	Intensity
630.06	22656
632.06	7770
646.19	3896
658.16	2437
713.19	3360
1226.47	1099

Table 84: Peak list of sample 5C used for Mascot search.

m/z value	Intensity
644.04	28199
666.04	98322
682.01	23270
833.09	7125
877.08	89821
887.24	14623
902.19	22245
911.09	7318
927.00	8577
939.30	5584
1050.14	5537
1072.13	4613
1082.10	15652
1088.11	4481
1098.07	4533
1137.57	5278
1271.13	2902
1277.15	6520
1293.12	5100

Table 85: Peak list of sample 5DE used for Mascot search.

m/z value	Intensity
644.05	12133
650.07	16480
666.04	16244
726.42	2488
786.54	2362
829.58	2840
1082.10	2456
1358.75	1348
2120.99	655

Table 86: Peak list of sample 5FGH used for Mascot search.

m/z value	Intensity
653.38	10947
768.52	3944
786.53	4751
834.41	3180
881.29	12164
1107.57	2719
1410.72	1774
1427.76	1842
1450.72	1421
1516.78	1794
1598.80	1313
1794.85	2986
1839.93	1641
1866.88	1446
1884.88	93456
1896.88	2254
1906.85	1776
1911.89	1867
1922.83	1634
1926.89	1826
1941.90	4496
1953.91	1183
1990.94	1363
2052.07	2425
2071.92	781
2238.11	1051
2253.12	3115
2303.18	927
2324.20	13128
2351.19	1136
2381.21	1387
2399.08	887
2563.12	1084

2760.23	894
2864.44	1233
2896.43	931

Table 87: Peak list of sample 5A used for Mascot search.*

m/z value	Intensity
587.15	23328
622.06	11950
628.18	14882
646.19	14039
672.17	15733
694.15	8217
834.40	7384
861.22	7090
881.28	11679
1195.57	5518
1450.68	4427
1516.76	3547
1741.77	6506
1812.89	3456
1990.94	4005
2091.90	3476
2231.95	2448
2405.95	1751
3313.31	1862

Table 88: Peak list of sample 5BC used for Mascot search.*

m/z value	Intensity
1598.79	6658
1740.77	5833
1848.80	5180
1941.90	5566
1990.93	5354
2022.91	4221
2091.91	5023
2096.99	7646
2343.02	4369
2441.03	3395
2717.08	2344
2760.22	2660
2864.44	2005

7.1.6 Gel Band 6

Table 89: Peak list of sample 6A used for Mascot search.

m/z value	Intensity
599.41	855
881.29	797
1066.55	644

Table 90: Peak list of sample 6B used for Mascot search.

m/z value	Intensity
571.15	2099
599.39	2250
628.17	2422
646.18	2784
757.18	2190
1519.70	805
2051.86	534
2091.86	873
2293.93	370
571.15	2099
599.39	2250
628.17	2422
646.18	2784
757.18	2190
1519.70	805
2051.86	534
2091.86	873
2293.93	370

Table 91: Peak list of sample 6C used for Mascot search.

m/z value	Intensity
887.32	14521
902.29	12839
917.40	14739
1047.19	9069
1062.16	11139
1286.84	3752

Table 92: Peak list of sample 6DE used for Mascot search.

m/z value	Intensity
644.05	6449
650.07	9853
738.41	3065
829.58	3414
965.56	1907
1015.65	2274
1107.59	2321
1125.59	2335
1275.71	6408
1288.70	6371
1365.71	4049
1647.92	2513
2051.97	1143
2199.04	829

Table 93: Peak list of sample 6FGH used for Mascot search.

m/z value	Intensity
797.37	2506
829.39	2945
881.29	3347
905.49	2372
991.45	1867
1107.57	2024
1584.76	2999
1740.77	4362
1813.75	2779
1848.80	2002
1867.90	1968
1884.88	48559
1896.87	2097
1910.90	3271
1926.89	1886
1941.89	1779
1947.94	1281
2052.07	2568
2238.10	1551
2253.12	2898
2303.18	1089
2324.20	16035
2343.03	1096
2350.06	997
2366.22	1311
2381.21	1252
2399.05	2198
2405.97	1730

2441.02	1411
2563.12	1621
2679.17	1754
2774.14	578

Table 94: Peak list of sample 6A used for Mascot search.*

m/z value	Intensity
672.17	13347
861.22	6149
1077.22	9424
1187.61	3993
1275.66	5979
1288.67	7135
1295.69	5190
1337.72	3187
1345.69	3531
1749.80	2766
1884.87	3668
2231.95	2585

Table 95: Peak list of sample 6BC used for Mascot search.*

m/z value	Intensity
537.45	43075
570.64	140485
582.29	47963
585.39	269355
597.40	50020
600.29	322354
653.37	125227
1142.70	25844
1926.89	21057
1941.90	12883
2135.11	7050
2423.24	4880

7.1.7 Gel Band 7

Table 96: Peak list of sample 7A used for Mascot search.

m/z value	Intensity
646.20	2647
811.44	2520
887.04	1565
1005.30	1242
1854.97	1565
1870.93	737

Table 97: Peak list of sample 7B used for Mascot search.

m/z value	Intensity
509.30	3403
510.32	9104
599.39	3926
646.19	2854
757.18	2499
816.98	2431
1218.46	1262

Table 98: Peak list of sample 7C used for Mascot search.

m/z value	Intensity
550.14	10628
587.17	17979
622.07	26909
715.47	13060
842.56	35633
856.57	9292
917.34	7241
2120.04	2292

Table 99: Peak list of sample 7DE used for Mascot search.

m/z value	Intensity
587.15	12830
715.46	34709
733.41	26583
811.43	59882
853.44	5888
1056.51	4693

1077.22	26465
1120.57	5700
1206.57	16056
1218.49	5477
1281.63	26286
1303.63	3954
1314.61	4124
1429.69	7751
1477.83	4649
1806.92	2371
1813.80	2413
1822.87	7388
1850.95	3405
1854.94	18792
1870.92	4873
1896.94	2234
1910.92	2809
2052.07	1556
2218.22	1264
2241.00	1398
2303.20	1601
2324.21	9003
2351.18	2775
2502.08	1169
2563.11	2776
2679.16	974

Table 100: Peak list of sample 7FGH used for Mascot search.

m/z value	Intensity
607.27	4335
715.45	2948
733.40	3738
763.42	2020
786.52	1431
790.42	1542
811.42	4110
1176.54	1891
1206.55	2385
1281.62	3493
1392.63	1294
1477.79	1542
1630.79	1421
1884.85	33587
1896.84	1364
1906.82	981
1910.87	1395
1926.86	1618

1941.87	1284
2052.03	1678
2238.07	1062
2253.09	2164
2303.11	801
2324.17	8681
2351.15	910
2366.15	947
2381.18	1083
2563.09	502
2679.10	389
2864.41	473

Table 101: Peak list of sample 7A used for Mascot search.*

m/z value	Intensity
534.18	10536
543.16	10181
587.15	17184
590.21	16715
646.18	10658
733.40	17668
811.42	7257
881.27	16597
1077.21	14415
1281.62	13874
1533.67	4883
1620.65	5797
1749.79	4321
1784.81	4602
1884.88	3556
1910.89	4037
2091.89	2589
2185.88	2366
2231.94	4176
2253.09	2083

Table 102: Peak list of sample 7BC used for Mascot search.*

m/z value	Intensity
1471.74	10498
1477.79	10843
1683.87	11080
1926.88	18726
1941.88	12450
2185.62	6312

2343.02	5662
2366.19	9906
2694.23	3466
2864.43	18507

7.1.8 Gel Band 8

Table 103: Peak list of sample 8A used for Mascot search.

m/z value	Intensity
646.20	1089
678.03	763
713.20	1106
757.19	978
1708.84	642

Table 104: Peak list of sample 8B used for Mascot search.

m/z value	Intensity
881.29	24051
1625.88	3900
1653.88	3521
1765.83	27418
1884.93	8669
2343.07	4460
2774.23	991

Table 105: Peak list of sample 8C used for Mascot search.

m/z value	Intensity
539.09	4841
542.16	2060
554.13	4765
580.11	2545
590.12	5215
644.04	4439
646.18	2422
658.16	2374
697.97	2642
710.12	2099
713.19	2405
757.18	1942
777.15	1745
793.12	1665

909.01	2341
1005.28	1131
1082.07	1621
1098.04	1095
1107.55	849

Table 106: Peak list of sample 8DE used for Mascot search.

m/z value	Intensity
506.16	13646
543.17	15608
550.14	21815
587.16	29873
602.21	9352
628.19	11903
646.19	11002
672.17	15201
706.24	8782
713.20	10770
881.29	9653
917.30	7665
1120.58	7013
1125.56	5149
1176.58	5223
1206.59	10786
1292.64	39963
1429.70	4844
1450.70	3719
1741.80	3870
1808.81	3438
1822.88	4104
1879.84	2758
1884.90	5716
1910.91	2572
1930.93	1916
1947.96	2622
2052.02	2834
2091.93	3896
2238.14	1402
2303.17	1131
2324.22	9079
2342.05	1623
2351.19	3259
2366.15	1281
2410.03	1594
2432.12	1556
2446.15	1156
2563.13	2421

2679.22	842
2717.14	1660

Table 107: Peak list of sample 8FGH used for Mascot search.

m/z value	Intensity
1049.45	2616
1206.55	6288
1306.59	2636
1331.58	4079
1335.60	2251
1348.60	3477
1429.65	2495
1461.66	1672
1480.57	3733
1526.63	4270
1548.68	2858
1598.76	1248
1681.61	1367
1700.77	6145
1822.82	1771
1867.83	1879
1884.84	33833
1910.85	3394
1926.84	1776
1941.85	1394
1967.87	1251
2026.92	747
2052.03	3220
2071.86	816
2238.06	1111
2253.07	1360
2303.12	1000
2324.15	14245
2342.00	968
2351.13	1262
2366.14	1032
2381.17	1104
2432.06	891
2447.07	1182
2563.07	1398
2679.12	813
2893.04	689

Table 108: Peak list of sample 8A used for Mascot search.*

m/z value	Intensity
550.13	11101
573.37	8396
768.50	6770
786.51	11198
881.28	56146
1330.63	4099
1708.83	10156
1749.80	2653
1785.82	3371
1848.82	2733
1884.86	4841
1910.88	3026
2231.94	2276
2253.09	2140
2303.17	1815
2342.03	4082
2432.09	1386
2440.99	2281
2679.16	1190
2717.09	1740
2774.12	1089

Table 109: Peak list of sample 8BC used for Mascot search.*

m/z value	Intensity
1317.64	23743
1605.81	15587
1749.80	13061
1926.88	25040
1941.89	17912
1954.90	12502
1983.90	12254
2342.05	10066
2432.10	5040
2864.43	6504