

DIPLOMARBEIT

**Analysis and identification of extracellular proteins
of *Hypocrea atroviridis* through gel electrophoresis
and mass spectrometry**

ausgeführt am

Institut für Chemische Technologien und Analytik
der Technischen Universität Wien

unter der Anleitung von

Univ. Prof. Dr. Günter Allmaier

durch

Reingard Schandl

Matznergasse 40/16, 1140 Wien

06.05.2007

Reingard Schandl

Danksagung

Ich möchte mich an dieser Stelle bei den folgenden Personen bedanken:

Bei Univ.Prof. Dr. Günter Allmaier und Univ.Prof. Dr. Christian Kubicek, die mir ein interessantes Diplomarbeitsthema und die Mitarbeit am EU-Forschungsprojekt TrichoEST (QLK3-CT-2002-02032), über das ich auch finanziert wurde, ermöglicht haben.

Bei meinen betreuenden Assistentinnen Dr. Martina Marchetti und Dr. Verena Seidl für ihre Geduld und Unterstützung bei meiner Arbeit. Weiters möchte ich mich bei Dr. Verena Seidl für die Bereitstellung der 2D-Gele bedanken auf die diese Arbeit aufbaut.

Bei meinen Arbeitsgruppenkollegen und -kolleginnen Mag. Roswitha Braunrath, Mag. Tom Grunert, Mag. Jasmin Hirschmann, Dr. Christian Laschober, Dr. Corina Mayrhofer, Dr. Roland Müller, Dr. Ernst Pittenauer, Dr. Gerald Stübiger, Dr. Wolfgang Winkler und Dr. Martin Zehl für die gute Zusammenarbeit und die zahlreichen praktischen Tipps.

Bei meinen Eltern und Geschwistern für die Unterstützung während meines gesamten Studiums, die mir speziell in schwierigen Phasen geholfen hat und mir erlaubt hat meine Interessen und Vorstellungen zu verwirklichen.

Bei meinem Freund Christian Dörflinger für die Ermutigung während meiner Diplomarbeit ebenso wie für seine Geduld und sein Verständnis.

<u>1</u>	<u>INTRODUCTION</u>	<u>1</u>
1.1	BIOLOGICAL BACKGROUND	1
1.1.1	The organism used in this study	1
1.1.2	Biocontrol	2
1.1.3	Biocontrol mechanisms of <i>Hypocrea/Trichoderma</i>	2
1.1.4	The TrichoEST project	4
1.2	PROTEOMICS	5
1.2.1	The nature of proteins	5
1.2.2	Definition and aim of proteomics	7
1.3	APPLIED METHODS OF PROTEIN ANALYSIS	9
1.3.1	Sample preparation techniques	9
1.3.2	Enzymatic protein cleavage	11
1.3.3	Gel electrophoresis	13
1.3.4	MALDI-TOF-MS	23
1.3.5	Protein Database Search	31
1.4	AIMS OF THIS STUDY	36
<u>2</u>	<u>METHODS</u>	<u>37</u>
2.1	USED CHEMICALS	37
2.2	SAMPLE PREPARATION AND 2D GEL ELECTROPHORESIS	38
2.3	TRYPTIC DIGEST	39
2.3.1	Applied solutions	39
2.3.2	Procedure	39
2.4	PNGASE F DIGEST	41
2.4.1	Applied solutions	41
2.4.2	Procedure	41
2.5	ZIPTIP PURIFICATION	42
2.6	MALDI-TOF-MS	43
2.6.1	Instrumentation	43
2.6.2	Sample preparation	43
2.6.3	Calibration	43
2.7	DATABASE SEARCH AND BIOINFORMATICS	44

3	RESULTS AND DISCUSSION	46
3.1	EVALUATION AND COMPARISON OF THE 2D GELS	46
3.1.1	2D gels with glucose as carbon source	47
3.1.2	2D gels with host cell walls as carbon source	47
3.2	IDENTIFICATION OF PROTEIN SPOTS FROM 2D GELS WITH GLUCOSE AS CARBON SOURCE	50
3.2.1	Spot G1 & G2 – Epl1	50
3.2.2	Spot G3	55
3.3	IDENTIFICATION OF PROTEIN SPOTS FROM 2D GELS WITH CELL WALLS OF POTENTIAL HOSTS AS CARBON SOURCE	63
3.3.1	Spots R1, B1, P1, R2, B2 & P2	63
3.3.2	Spots R3, B3 & P3	71
3.3.3	Spots R4 & B4 – <i>R. solani</i> & <i>B. cinerea</i> cell walls	74
3.3.4	Spot P4 – Exoglucanase	82
3.3.5	Spots R5 & B5	85
3.3.6	Spot P5	90
3.3.7	Spot R6 - Alginate lyase	95
3.3.8	Spots R7 & B7	99
3.3.9	Spots R8, B8 & P8 – Epl1	101
3.3.10	Spots R9, B9 & P9	103
3.4	EXPERIMENTS WITH PNGASE F	105
4	CONCLUSION	107
5	REFERENCES	109
6	APPENDIX	114

1 Introduction

1.1 *Biological background*

1.1.1 The organism used in this study

Hypocrea/Trichoderma species are filamentous fungi belonging to the division Ascomycota, order Hypocreales. Initially mainly the *Trichoderma* anamorphs, the asexual reproducing stages, were described. During the last years more and more of them were linked to the sexual reproducing *Hypocrea* teleomorphs [1]. Additionally new phylogenetic studies led to many revisions of the taxonomy [2].

Hypocrea/Trichoderma species are found commonly in soils, decaying wood or plant organic matter and are capable of metabolizing a wide range of different carbohydrates. Their geographical extension reaches from the polar circle to the equator, covering therefore diverse climates. They are very competitive in their habitats due to the production of a large number of spores and a rapid growth rate. Additionally they are capable of mycoparasitism, which means they can attack and destroy other fungi, especially plant pathogens, by degrading their cell walls and feeding on the dead cell contents [3, 4].

These properties make a variety of *Hypocrea/Trichoderma* species very interesting for biocontrol and other industrial purposes. For example formulations containing among others *H.lixii* (anamorph: *T. harzianum*), *H. atroviridis* (*T. atroviride*) and *T. asperellum* are commercially available as biocontrol agents as well as plant growth promoters and soil conditioners [5, 6]. A different application is the employment of *H. jecorina* (*T. reesei*) for the production of cellulases and hemicellulases and its application to express heterologous proteins [7].

H. atroviridis (*T. atroviride*) P1, the strain used in this study, was first believed to belong to the species *T. harzianum*, but phylogenetic studies led to renaming in *T. atroviride* [8]. Only in 2003 its teleomorph was discovered [9], as mostly its asexual form is found in nature, which proliferates through green conidia. It shows the properties already described for the genus, especially also mycoparasitism and biocontrol ability, so it is a biotechnologically interesting organism.

1.1.2 Biocontrol

A definition of the term biocontrol was proposed by Cook and Baker in 1983 [10]: “*Biological control is the reduction of the amount of inoculum or disease-producing activity of a pathogen accomplished by or through one or more organisms other than man.*” According to this definition the use of more resistant cultivars of the host is a way of biocontrol as well as the application of antagonistic organisms which reduce the disease rate. The biological processes involved can be found throughout evolution but only a few are yet subjected to research of their industrial application, e.g. in agriculture.

In the farming industry the substitution or reduction of chemical pesticides is a major aim. They are at the moment the most common method to prevent crop disease, but the number of resistant pathogens is increasing. Additionally consumers are focusing more and more on ecological and health consequences of the excessive use of chemical plant protectants.

Pest management through biocontrol shows some advantages compared to conventional methods, for example the possibility of larger intervals between applications due to self perpetuation and a higher host specificity. Nevertheless the results of this organic approach are less predictable as biological organisms are stronger influenced by e.g. the weather conditions and the amount of disease producing microorganisms as chemical agents.

Information about the specific properties of a biocontrol species helps to lessen this disadvantage by reducing the unknown parameters. The focus of investigation lies therefore especially on organisms which can reduce important diseases, for example *Hypocrea/Trichoderma* species showing biocontrol activity against plant pathogenic fungi. The importance of controlling these pests is pointed out by the fact that fungi and fungus-like Oomycota (especially species belonging to the genera *Rhizoctonia*, *Botrytis*, *Phytophthora*, *Pythium*, *Sclerotinia* and *Fusarium*) are responsible for approximately 70% of all major crop diseases [11].

1.1.3 Biocontrol mechanisms of *Hypocrea/Trichoderma*

Already in the early 1930s the biocontrol activity of *Hypocrea/Trichoderma* spp. against *Rhizoctonia* and *Sclerotinia* was discovered by Weindling [12]. Since then, *Hypocrea/Trichoderma* species have been subjected to various experiments, during which the different mechanisms of biocontrol were discovered. The exact processes may show variations between the different species, but the general concepts are alike for all *Hypocrea/Trichoderma* spp. exhibiting biocontrol, among them *H. lixii* (*T. harzianum*), *H. atroviridis* (*T. atroviride*), *H. virens* (*T. virens*), *T. viride* and *T. apserellum*.

Biocontrol by *Hypocrea/Trichoderma* spp. is based on different mechanisms: competition, plant beneficial processes, rhizosphere modification, antibiosis and mycoparasitism. They can occur in parallel in the soil and have synergistic effects.

1.1.3.1 Competition

Organisms growing in the same habitat are always competing for nutrients and living space. *Hypocrea/Trichoderma* species are able to mobilize and take up soil nutrients more efficiently than many other microorganisms and are capable of degrading carbohydrate polymers such as cellulose, glucan and chitin. Together with the longevity of their spores and their rapid growth rate this results in an advantage against other microorganisms in their natural habitat [5].

1.1.3.2 Plant beneficial processes

The presence of *Hypocrea/Trichoderma* species in the rhizosphere of plants enhances root growth and crop productivity. This effect is not only caused by the reduction of pathogenic organisms in the surroundings but also observed without disease pressure. An increase in the uptake of various nutrients in the roots was demonstrated independently of the presence of other microorganisms in the soil. Interaction between *Hypocrea/Trichoderma* and plants is based on the colonization of the plant roots by penetration into the epidermis. Once in direct contact with the plant *Hypocrea/Trichoderma* not only withstands its toxic metabolites but also induce further defence mechanisms against pathogenic organisms. These reactions include the production of low molecular weight compounds like phytoalexins, which have antimicrobial properties. Some of them act not only locally but even induce systemic resistance, which means disease inhibition distant from the part of the plant that is in contact with *Hypocrea/Trichoderma* [13, 14].

1.1.3.3 Rhizosphere modification and antibiosis

The chemical composition of the soil influences germination and growth of plants as well as microorganisms. Especially the pH affects the survival of many species and the stability of any secreted protein or metabolite. *Hypocrea/Trichoderma* species are best adapted to acidic soil but they can prosper over a wide pH range. They can change their metabolism according to the surrounding conditions by expressing a set of proteins optimized for the respective external pH. This can even lead to a modification of the environmental pH. Additionally most *Hypocrea/Trichoderma* strains produce toxic compounds to impede the spore germination (fungistasis) or survival (antibiosis) of pathogenic microorganisms.

1.1.3.4 Mycoparasitism

Many *Hypocrea/Trichoderma* spp. (for example *H.lixii* (*T. harzianum*), *H. atroviridis* (*T. atroviride*) and *T. apserellum*) directly attack other fungi in a complex process called mycoparasitism, whose exact mechanism may vary slightly between the species. It starts with the release of low molecular weight compounds that are cleaved from the host cell wall by constitutively secreted, hydrolytic *Hypocrea/Trichoderma* enzymes. These low-molecular weight compounds, presumably oligosaccharides, can in turn induce the production of further cell wall degrading enzymes such as chitinases and therefore trigger the mycoparasitic response [15]. *Hypocrea/Trichoderma* then grows towards the other fungus and once it reaches its host, it attaches and coils around its hyphae, forming appressoria-like structures on them [16]. Production of further hydrolytic enzymes like chitinases, glucanases and proteases together with fungitoxic compounds like peptaibols lead to lysis of the host cell wall [17, 18] and *Trichoderma* feeds on the dead cell contents. The regulation of cell-wall-degrading enzymes of *Hypocrea/Trichoderma* spp. is a major field of research and has received a lot of attention in the past few years. A variety of different chitinases, glucanases, proteases and other hydrolytic enzymes are already known from various species. It has been shown that some of them are expressed constitutively, others are upregulated during mycoparasitism [19-21].

1.1.4 The TrichoEST project

The TrichoEST project was a research project funded by the European Union which was carried out at nine different groups including academic research facilities and industrial partners. The main goals of the project were the isolation, identification and characterisation of genes and gene products from *Hypocrea/Trichoderma* species. As one of the first steps an EST library containing genes expressed under various conditions by different strains were built [22, 23]. The investigated proteins and the developed techniques are of high interest for agricultural and environmental sciences as well as for the corresponding industries.


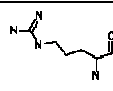




















1.2 Proteomics

1.2.1 The nature of proteins

The term “protein” was first introduced in 1838 by the Swedish chemist Berzelius [24], but until the middle of the 20th century their structure and their functions in living organisms remained undiscovered. Only in 1955 Sanger discovered the amino acid sequence of insulin which earned him a Nobel Prize in 1958 [25]. Today it is common knowledge that proteins consist of amino acids. In fact, all organic molecules carrying an amine attached to a carbon atom featuring a carboxyl group could be called amino acids, but only 22 α amino acids – those shown in table 1.2-1 – are encoded by the DNA [26] and only 20 of them are commonly found in proteins. They form linear chains connected through amide bonds, which are called peptide bonds, to build up the macromolecules. The sequence in which the amino acids are arranged is called the primary structure of a protein, but just the correct folding into regularly repeating local (secondary) and a protein specific global (tertiary) structure, in many cases additionally the interaction of more than one subunit (quaternary structure), guarantees the biological activity. Only then certain amino acids in the active center(s) of enzymes come spatially into contact which enables, for example, the catalysis of a reaction. Structure analysis is therefore an important research area, especially because this information can help to understand substrate – protein interactions and reaction pathways.

If the structure of a protein is destroyed by heating or treatment with certain chemicals the process is referred to as denaturation. Denatured proteins are biologically inactive and in most cases they are not able to refold themselves after the denaturing agent is removed. The activity of a protein also depends on the environmental pH value as it varies with the protein's thereby provoked charge, which is determined by the functional groups of the amino acids. Every protein has a pH value at which it has no net charge, the isoelectric point (pI). As this varies greatly between different proteins it is one of the attributes used in separation techniques. Another feature exhibiting large diversity is the size of proteins. It ranges from small polypeptides as for example Insulin with 51 amino acids and a molecular weight of 5.7 kDa to giant multifunctional proteins, such as Apolipoprotein B with over 4600 amino acids and more than 500 kDa molecular weight.

Table 1.2-1 List of the 22 encoded amino acids

Name	Formula	3-letter code	1-letter code	Mono-isotopic mass	Average mass	Structure
Alanine	C ₃ H ₅ NO	Ala	A	71.03712	71.08	
Arginine	C ₆ H ₁₂ N ₄ O	Arg	R	156.10112	156.19	
Asparagine	C ₄ H ₆ N ₂ O ₂	Asn	N	114.04293	114.10	
Aspartic acid	C ₄ H ₅ NO ₃	Asp	D	115.02695	115.09	
Cysteine	C ₃ H ₅ NOS	Cys	C	103.00919	103.14	
Glutamic acid	C ₅ H ₇ NO ₃	Glu	E	129.04260	129.12	
Glutamine	C ₅ H ₈ N ₂ O ₂	Gln	Q	128.05858	128.13	
Glycine	C ₂ H ₃ NO	Gly	G	57.02147	57.05	
Histidine	C ₆ H ₇ N ₃ O	His	H	137.05891	137.14	
Isoleucine	C ₆ H ₁₁ NO	Ile	I	113.08407	113.16	
Leucine	C ₆ H ₁₁ NO	Leu	L	113.08407	113.16	
Lysine	C ₆ H ₁₂ N ₂ O	Lys	K	128.09497	128.17	
Methionine	C ₅ H ₉ NOS	Met	M	131.04049	131.19	
Phenylalanine	C ₉ H ₉ NO	Phe	F	147.06842	147.18	
Proline	C ₅ H ₇ NO	Pro	P	97.05277	97.12	
Pyrrolysine	C ₁₂ H ₁₉ N ₃ O ₂	Pyl		237.1477	237.31	
Serine	C ₃ H ₅ NO ₂	Ser	S	87.03203	87.08	
Threonine	C ₄ H ₇ NO ₂	Thr	T	101.04768	101.10	
Selenocysteine	C ₃ H ₅ NOSe	SeC	U	150.95364	150.03	
Tryptophan	C ₁₁ H ₁₀ N ₂ O	Trp	W	186.07932	186.21	
Tyrosine	C ₉ H ₉ NO ₂	Tyr	Y	163.06333	163.18	
Valine	C ₅ H ₉ NO	Val	V	99.06842	99.13	

Some proteins show a variety of isoforms, which are all encoded by the same gene. The variations can either be introduced during the translation process if the mRNA sequence is spliced at different locations, or afterwards. In the second case the isoforms are the result of posttranslational modifications, for example glycosylations, phosphorylations, sulphatations or acetylations. As the modifications lead to slight mass and pI changes, a protein present in different isoforms often forms a series of spots next to each other called “train” in 2D gel electrophoresis.

During the ripening process leading from the translated amino acid sequence to the fully functional protein not only posttranslational modifications are added to the protein, but also signal peptides and other targeting signals are cleaved off from the immature protein. The role of these short amino acid sequences is to target a protein within a cell, but they are not necessary for its final function once it is located in the right cell compartment.

1.2.2 Definition and aim of proteomics

The term “proteome” is relatively young and was introduced by Marc Wilkins in 1995 in analogy to the word “genome” [27]. The genome represents the entire genetic information of a cell, the proteome describes all proteins present in an examined biological system. The genome is the same in every cell of an organism, whereas the proteome varies not only between different tissues or organisms but also depends on the metabolic situation, the environmental conditions and many other factors. “Proteomics” is studying the proteome and such investigations not only include the quantitative and qualitative analysis of the complete protein pattern of a biological system, for example a cell, an organism or a body fluid, but also the comparison of this proteome with others measured under different conditions, for example different climate, different tissue or different nutrient access. A typical question that is addressed in proteomics studies is the identification of up- and down regulated proteins as a reaction to a specific elicitor. This information enables conclusions about their potential biological functions. .

The high dynamic range of proteome concentrations can complicate the analysis and make it difficult to obtain reproducible results. To minimize this effect, further research in this field is carried out to optimize the experimental methods. The development of proteomics is very important as it is not possible to deduce the precise expression pattern of a protein in a cell solely from the DNA sequence of the corresponding gene. Not even the abundance of the mRNA transcript necessarily correlates tightly with the respective protein level. Various factors such as mRNA stability and posttranscriptional and posttranslational

processes can influence the amount of protein that is produced at a certain stage in the life cycle of a cell.

Information about the proteome is a key to understand the complex interactions within cells as proteins are responsible for almost all activities. Additionally to those functioning as enzymes, proteins are also relevant as structural building blocks like keratin, the basic material of mammal hair and skin, as parts of the transport system, like haemoglobin which transports oxygen in the blood, or for transducing signals that make it possible for the cell to react on external stimuli and interact with its environment. The information about the primary structure is saved in the DNA, but proteins are more intricate than just the translation of the genetic code into a series of amino acids. Especially the processes leading to different protein isoforms add complexity to the proteome and cannot be fully predicted at the present time [28]. With the aid of proteomics efforts are made to discriminate these variants. Sufficient knowledge of the genome is nevertheless necessary, as otherwise the identification of proteins is by far more difficult. Therefore the rapidly increasing number of fully sequenced genomes will enforce the importance of proteomic in the next years [29].

1.3 Applied methods of protein analysis

The main steps in proteomics are protein separation, e.g. by electrophoresis, the analysis of the resulting sample fractions, e.g. by MALDI-MS, and protein identification, e.g. by database search. Nevertheless also the handling of the proteins before and between these steps is very important but will only be discussed briefly in the corresponding chapters.

1.3.1 Sample preparation techniques

Sample preparation is a very critical step in proteome analysis as it involves the isolation of all proteins present in the biological system of interest which normally do not form a homogeneous group. The raw sample often consists of a complex protein mixture contaminated by different undesired components like nucleic acids, membrane components and various other impurities. Therefore the main objectives in initial protein preparation are an increase of the concentration of the interesting analytes and the removal of the contaminations.

This requires sample manipulation steps that all harbour the risk of protein loss and proteomic pattern change. Especially for low concentrated and/or poorly soluble components such as membrane proteins it can be very challenging to find the optimal purification procedure.

Commonly used methods include precipitation, ultrafiltration, dialysis and chromatography.

- Precipitation can be carried out with different agents, for example salts, e.g. ammonium sulfate, organic solvents, e.g. acetone or ethanol, or trichloroacetic acid. Depending on the agent, the proteins can denature, making a later resolubilisation difficult. It is possible to carry out fractionation with e.g. ammonium sulfate, because not all proteins are precipitated at the same salt concentration. Compared to the other described processes, precipitation needs the least instrumental effort. A disadvantage is that precipitation with the substances mentioned above cannot be carried out at very low protein concentrations.
- During ultrafiltration, components bigger than a defined molecular weight are retained by a filter membrane, e.g. modified cellulose or PVDF, whereas water and small substances may pass the barrier. The selection of a suitable pore size allows

the removal of small proteins if they are not in the focus of interest. Depending on the membrane and the analyte, adsorption onto the membrane and therefore protein loss can occur.

- Dialysis is carried out as a desalting or buffer exchange step. It works with a semipermeable membrane, e.g. cellulose acetate or nitrocellulose, which is placed between the sample solution and the water or buffer with which the solvent should be replaced. Due to the osmotic pressure, the different salts diffuse through the membrane, whereas proteins are retained on one side. Also in this case adsorption onto the membrane might lead to protein loss.
- Chromatographic methods are based on the distribution of the analyte between a mobile and a stationary phase. High performance liquid chromatography (HPLC) and gas chromatography (GC) are examples for this technique but require instrumental effort and are not typically used for sample preparation. Some procedures requiring far less instrumental complexity work with the same principle. Only examples of these methods are discussed, as they are applied for sample preparation. To separate proteins from impurities, one of the two groups has to be retained on the stationary phase material while the other one stays in the mobile phase. Proteins bind for example onto silica particles modified with alkyl residues, a typical reversed phase chromatography material, under aqueous conditions and can be removed from it through the addition of a solvent with an organic fraction. While they are bound, impurities can be washed away. This principle is used for Millipore ZipTip[®] technology, a desalting and concentration method where the chromatography material is placed into the end of a pipette tip [30]. An example where the proteins stay in the liquid phase is size exclusion chromatography. As big molecules they cannot pass into the small long channels of the column material, e.g. cross-linked dextran, and therefore elute before the impurities which penetrate into these holes.

Often more than one method is applied on a sample to get satisfactory results, even though every additional step carries the risk of a significant protein loss. Finally a solution ready for further use is obtained.

1.3.2 Enzymatic protein cleavage

Native proteins show a high molecular weight and complexity, so the analysis of their structure is difficult. Reducing the intricacy by cutting them into smaller pieces or removing posttranslational modifications allows easier detection of their primary structure. Of course the locations of the cuts have to be predictable to allow conclusions regarding the native protein. Various enzymes from different organisms exist which cleave peptide or other relevant bonds, e.g. the connection of a posttranslational modification to the protein, at specific sites. Their reaction parameters are already well known which allows their application for analytical and preparative purposes.

Some enzymes allow a digest as well in solution as directly in the electrophoresis gel matrix. It is recommended to denature the protein and reduce and alkylate its cysteins before the cleavage procedure to enable the applied enzyme to reach all its target bonds without steric hindrance. This also avoids that peptides stay connected to each other through disulfide bonds. Further preparative steps include the removal of any interfering components like the stain of the gel or alkylation reagents and adding digestion buffer to the dry protein or the gel matrix. Then the enzyme is applied and the cleavage carried out under the required conditions. Peptides of in-gel digested proteins can be washed out of the matrix after a peptide bond cleavage. The resulting solutions contain the applied enzyme as well as the buffer salts, which have to be removed if they disturb the further use of the proteins or peptides.

In this work, two different enzymatic digestions were carried out. Trypsin was used to cleave the proteins into peptides and PNGase F was used to remove N-glycanes:

- Trypsin is the most commonly used enzyme to cleave proteins into defined peptides for analysis. It is an endoprotease, which means it hydrolyses peptide bonds in the middle of a protein, whereas exoproteases remove amino acids from the end of a protein. Trypsin cleaves specifically peptide bonds C-terminal of lysine and arginine. The serine protease is found in the digestive system of many different species, including humans, but mostly bovine trypsin is used for biochemical applications. It has its highest activity at a pH from 7.5 to 9 and a temperature of 37 °C. As the natural form is prone to autolysis, a modified type with increased stability through cross linking to a hydrophilic polymer is available from different commercial sources.

- PNGase F is the short form for Peptid-N⁴-(Acetyl-β-Glucosaminy) Asparagin Amidase, an enzyme originally found in *Flavobacterium meningosepticum*. It is used to cleave asparagine bound N-glycans whose oligosaccharide is longer than the chitobiose core unit. The resulting products of this reaction are the intact carbohydrate, ammonia and an aspartic acid instead of asparagin in the protein chain. Releasing the sugar residues before MALDI-MS simplifies the analysis and allows identifying apparently different proteins as isoforms with different glycosilation. However, PNGase F only cleaves off N-glycans not O-glycans and its function can be hindered through steric reasons. The enzyme shows its highest activity at a temperature of 37 °C and a pH from 7 to 9, but it is also functional at a pH as low as 5.

1.3.3 Gel electrophoresis

1.3.3.1 Introduction

History of electrophoresis goes back to the year 1930, when the Swedish scientist Arne Tiselius used it to separate human serum proteins. His research gained him a Nobel Prize in 1948 [31]. 11 years later, in 1959, the first use of polyacrylamide gels signified an important innovation. Raymond and Weintraub used this inert transparent matrix to separate DNA and protein mixtures by their size [32]. Development of electrophoretic media was continued by Hjerten, who applied Agarose gels to analyze DNA fragments in 1961 [33]. Another 5 years later, in 1966, the introduction of carrier ampholytes to form a stable pH gradient, led to the possibility to separate the analytes by a different characteristic – their pI value [34]. The new technique was called isoelectric focusing (IEF) and formed the basis for the second very important step in the development of electrophoresis. Nevertheless the innovation took place only in 1975, when O'Farrell and Klose combined IEF with sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE) to a new method called two dimensional (2D) gel electrophoresis [35, 36]. This breakthrough enabled the separation of total cell lysates and therefore to analyze proteomes. The method was once more refined in 1988 by the invention of immobilized pH gradients [37]. Today it is an important analytical tool due to its resolution, high reproducibility and high loading capacity.[38]

1.3.3.2 General concepts of electrophoresis

The migration of charged particles in an electric field constitutes the basic principle of electrophoresis. Substances differing in charge and/or size feature unequal electrophoretic mobilities, which leads to a separation in discrete zones during migration.

Depending on the media applied, the influence of the substance's size on the separation differs. In practice, either a solution, or a stabilizing matrix, like a gel or a membrane, is used. In the case of membranes size nearly has no influence and separation is mainly based upon charge differences, whereas for gels, size and charge affect the separation. For both the applied high voltage is a critical issue as it causes heating effects due to Joule heat. Many proteins are heat sensitive and can denature if subjected to high temperatures. Additionally heat causes convection, resulting in peak dispersion and a decrease of resolution. This effect is less pronounced if a matrix – an anticonvective medium – is used than if electrophoresis is carried out in solution [39].

The commonly applied electrophoretic techniques can be divided into three main types, namely isotachopheresis, isoelectrical focusing and zone electrophoresis.

- In isotachopheresis, a discontinuous buffer system is applied to concentrate samples between two buffer zones where they have highly different electrophoretic mobilities.
- The separation principle of isoelectrical focusing is the pI value of the analyzed molecules which move in a pH gradient.
- Zone electrophoresis uses a continuous buffer system with a constant pH value. The analytes are separated according to their different electrophoretic mobilities.

The high resolving power of especially gel electrophoresis is one of the main reasons for its many analytical and preparative applications in biochemistry, molecular biology and medicine. In 2D gel electrophoresis both isoelectrical focusing and zone electrophoresis are used as complementary techniques to give a maximum of resolution, as described in chapter 1.3.3.4 and 1.3.3.5.

To separate charged molecules in an electric field, their migration properties must vary. Accelerating (F_e) and frictional (F_{fr}) forces are the most influencing physical properties. (q - charge of the particle, E - electric field strength, f_c - friction coefficient, v - velocity)

$$F_e = q \cdot E \quad \text{Equation 1.3-1}$$

$$F_{fr} = f_c \cdot v \quad \text{Equation 1.3-2}$$

Every substance moves at a specific velocity defined through the balance of these two forces, as shown in equation 1.3-3. The proportional factor between the electric field strength, which is constant for a given system, and the velocity is defined as the electrophoretic mobility u.

$$\begin{aligned} q \cdot E &= f_c \cdot v & F_e &= F_{fr} \\ v &= \frac{q}{f_c} \cdot E \\ v &= u \cdot E \end{aligned} \quad \text{Equation 1.3-3}$$

For small spherical particles the mobility can be described by the Stokes law. (z - valence, e - elementary charge, r - Stokes radius, η - viscosity of the solution)

$$u = \frac{z \cdot e}{6\pi \cdot \eta \cdot r} \quad \text{Equation 1.3-4}$$

Proteins or peptides are normally non-spherical particles and therefore they cannot be described by equation 1.3-4 but by the empirical correlation in equation 1.3-5. (MW - molecular weight)

$$u \propto \frac{z \cdot e}{(MW)^{2/3}} \quad \text{Equation 1.3-5}$$

As already shown in equation 1.3-3 the velocity increases with the electric field strength. Therefore also the variation in the velocities of different substances is amplified and the resolution should improve. However, a 10-fold higher voltage leads to a 100-fold increase of Joule heat as shown in equation 1.3-6. (W - generated joule heat per volume, c - concentration of the electrolyte, F - Faraday constant)

$$W = E^2 \cdot u \cdot z \cdot F \cdot c \quad \text{Equation 1.3-6}$$

1.3.3.3 Sample requirements

Samples for electrophoretic experiments have to fulfill a set of requirements to guarantee reproducible results. They should not contain salt or buffer concentrations higher than 50 mM, therefore often desalting steps during the sample preparation are necessary. Nevertheless the proteins should stay completely in solution as small solid particles as lipid droplets or insoluble proteins can disturb the separation by blocking the pores of the gel. Additionally denaturing and reducing steps are required for most techniques to ensure that every protein is present in only one conformation without aggregation or interaction with other proteins.

Therefore various components which do not interfere with electrophoresis are used. The composition of the sample solution varies depending on the kind of electrophoresis performed. In this work the focus lies on 2D gel electrophoresis, therefore a relevant

sample preparation is carried out before isoelectric focussing. A sample solution used for this purpose commonly contains the following components [40]:

- Neutral chaotropes, e.g. Urea: destruction of non-covalent and ionic interactions
- Non-ionic and zwitterionic detergents, e.g. Triton X-100, CHAPS: prevention of protein aggregation through hydrophobic interaction
- Reducing Agents, e.g. Dithiothreitol: breaking of disulfide bonds
- Carrier ampholytes: enhance solubilisation, minimize protein aggregation
- Buffers or bases, e.g. Tris: very low concentrations, enhance solubilisation

1.3.3.4 Isoelectric focusing (IEF)

The separation principle of isoelectric focusing is the difference in the isoelectric point (pI) of proteins. As they do not have a net charge at this pH value, their movement in the electric field stops at exactly this point. Beside the variation in amino acid composition of different proteins or splice variants of the same protein, post translational modifications influence the pI of a protein so it is possible to separate different proteins and isoforms of a protein with this technique. To reach the separation a pH gradient is used where the different proteins migrate to the pH corresponding to their pI.

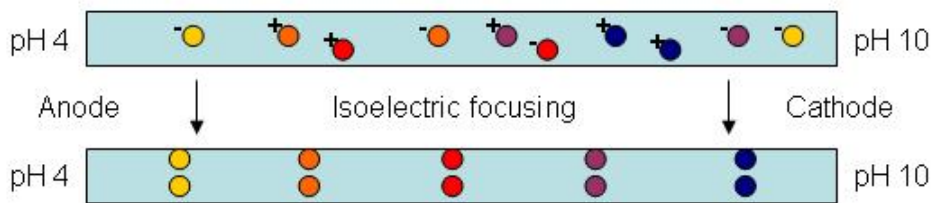


Figure 1.3-1 Principle of isoelectric focusing

The pH gradient should be stable, have a constant conductivity and a constant buffer capacity and commonly either carrier ampholytes or immobilized pH gradients are used.

Carrier ampholytes

Various small aliphatic zwitterions with different pI values are used together as carrier ampholytes. They guarantee high buffer capacities, high solubility of proteins and constant conductivities near the pI. Commonly gels used for this technique contain 2 % carrier

ampholytes. When the gel is cast, the zwitterionic molecules are spread equally over the whole gel, leading to an average pH. If an electrical field is applied, the charged ampholytes migrate towards their corresponding electrodes, due to their different mobilities they generate a pH gradient for the subsequent separation of the proteins.

The production of pH gradients with the aid of carrier ampholytes in polyacrylamide and agarose gels is therefore rather easy. Additionally these components have the advantage of supporting protein solubility. Their disadvantages are the low reproducibility and stability of the created gradient, their heat sensitivity and the tendency to drift towards the cathode.

Immobilized pH gradient (IPG)

Immobilized pH gradients are produced during casting of the gel. Derivatives of acrylamide with weak acidic or basic properties, called immobilines, are copolymerized together with the standard components acrylamide and bisacrylamide (see chapter 1.3.3.6). Their content is continuously varied along one of the axis of the gel, creating the desired pH gradient.

In general, IPG gels are available as dried strips and can be purchased for various pH ranges (e.g. pH range 4.0-5.0 or 3.5-9.5) with a plastic backbone. They have to be rehydrated with a solution containing the components mentioned in chapter 1.3.3.3. It is possible to directly have the sample that is to be analyzed dissolved in the rehydration solution and load it during the rehydration step.

As the pH gradient is a part of the polymer matrix, its reproducibility is very high and it is stable throughout the separation. Additionally, IPGs have the advantage of high loading capacities up to 500 μg of protein for a strip with 11 cm length. Their disadvantage is the complex production together with the long separation time.

Three different types of loading the sample onto the IPG strip are commonly used:

- Rehydration loading: Especially IPG strips are dried and have to be rehydrated before use. The sample can be a component of the rehydration solution, which allows the loading of comparatively large sample quantities and minimizes the risk of protein precipitation during the loading process. Some rehydration equipments allow active loading, which means the application of a low voltage (e.g. 50 V) to improve the migration of especially high molecular weight proteins into the strip. Loading without an applied voltage is called passive loading.

- Cup loading: In this case sample cups are put in contact with the gel. They contain the sample and where they are placed onto pH gradient depends on the nature of the proteins of interest. Gentle handling is necessary as the polyacrylamid layer of the IPG strip is very fragile. Additionally this method only allows lower protein loads (maximum 100 μg of protein) and carries the risk of protein precipitation.
- Paper-bridge loading: The sample solution is soaked up with a paper-bridge, which is then placed onto the gel.

Once the sample is loaded, the gel is placed in an electric field. In general the voltage is slowly increased to the final electric field strength where the focusing is carried out for several hours in case of IPG strips. An example for an IEF program for an 11 cm IPG strip is given in table 1.3-1. The amperage per strip is limited with 50 μA to avoid overheating and burning of the strips.

Table 1.3-1 IEF program for an 11 cm IPG strip

Step	Duration (h:mm)	Voltage (V)	Volt-hours (kVh)
1	1:00	500	0.5
2	1:00	1000	1.0
3	1:50	8000	12.5
Total	3:50		14.0

1.3.3.5 SDS –PAGE

SDS-PAGE is applied to separate proteins according to their molecular weights. Following the principles of electrophoresis, the charges of the analytes influence their mobilities in an electric field. To allow neglecting this effect, all substances need to have the same charge to mass ratio. Therefore an anionic detergent like SDS is used, which attaches at a very constant ration of 1.4 g SDS per 1 g protein to these macromolecules. As long as enough SDS is added to the sample buffer as well as to the running buffer, it builds SDS-protein-micelles with a negatively charged surface and covers the charge of the native protein. Furthermore this inhibits the formation of intermolecular interactions or aggregations. To break the disulfide bonds of the proteins often reducing agents, e.g. Dithiothreitol, are added. Only the reduction together with an additional heat denaturation step guarantees that all proteins unfold completely so that their shape does not influence the separation and their molecular weight can be determined.

SDS-PAGE can be applied either alone or in combination with IEF as 2D gel electrophoresis. In the first case normally sample solutions are pipetted into pockets on the cathode side of the gel, whereas in the second case a focused IEF strip is inserted. To allow molecular weight determination, a marker consisting of proteins with known sizes is applied next to the samples.

1.3.3.6 Gel types for gel electrophoretical experiments

Two different matrices are commonly used in gel electrophoresis: polyacrylamide, mainly for proteins, and agarose, mainly for DNA but also for large proteins.

Polyacrylamide

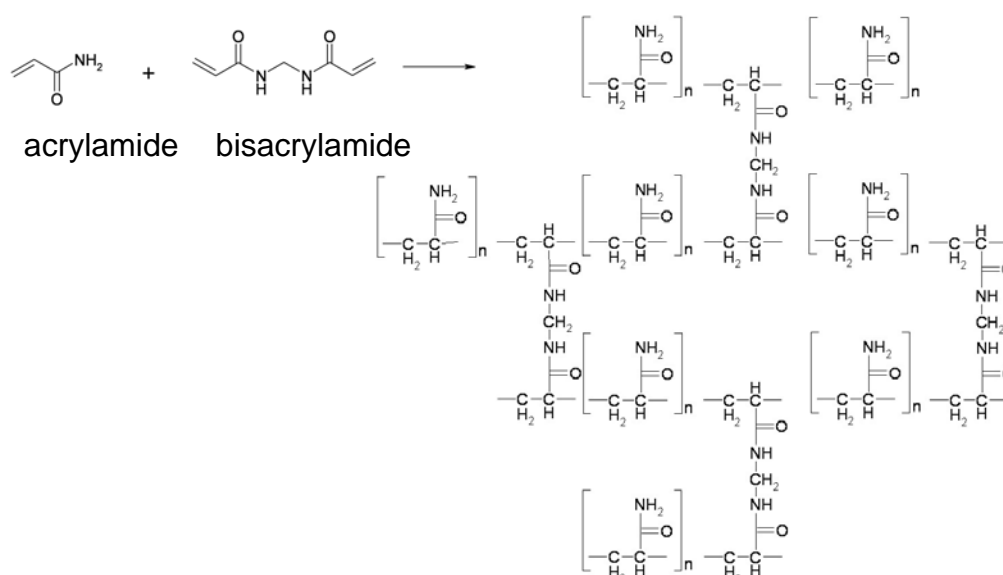


Figure 1.3-2 Chemical reaction to build polyacrylamide

Polyacrylamide gels are inert, transparent and their pore size can be varied to provide an optimized molecular weight range for the proteins that are analysed. The pore size is defined by the ratio between acrylamide and bisacrylamide monomers. Whereas acrylamide polymerizes to long chains, bisacrylamide is responsible for the cross linking of these chains. Specifications of a gel state the total concentration of acrylamide T and the degree of cross linking C. Typically gels for protein analysis have 3-12 % T and 3-5 % C and a pore size in the range of a few nm. A higher total acrylamide concentration leads to smaller pores as well as a higher cross linking degree. In general gels used for IEF have a lower total acrylamide concentration (e.g. 5 % T) than gels for molecular weight

separation. It is also possible to produce gels with a gradient, for example from 4 to 12 % total acrylamide concentration.

Advantages of polyacrylamide gels include good separation quality over a wide mass range, low electroosmosis, high stability, good storability, clarity and the suitability for many different staining methods. On the other hand the monomers are toxic, they have a limited pore size which does not allow separating proteins bigger than 800 kDa and alkaline gels can only be stored for a short time.

Agarose

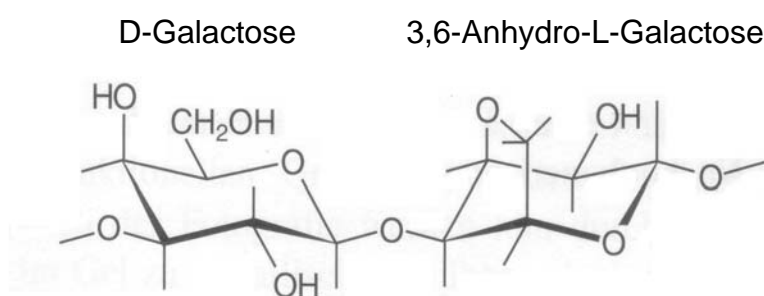


Figure 1.3-3 Chemical structure of the disaccharide which is the base of agarose

Agarose is a polysaccharide obtained from red seaweed. It is dissolved in nearly boiling water at a concentration in the range of 1 % to cast gels which solidify during cooling. The resulting gelatinized agarose is less transparent than polyacrylamide and has pore sizes in the range of a few hundred nm. Therefore it shows a poor separation of proteins below 100 kDa. It is also susceptible to electroosmosis and tends to a strong background staining especially for sensitive staining methods. Nevertheless its easy production, the fact that it is not toxic and the possibility to carry out specific detection methods for example with antibodies within the gel facilitate its use for large proteins.

1.3.3.7 Protein visualization

Proteins can be visualized with many different methods including staining with Coomassie dyes, silver staining, fluorescent dyes, autoradiography and fluorography. This listing ranges from lower to higher sensitivity, reaching a limit of detection in the fmol range for the last two methods. Whereas Coomassie and silver staining are traditionally very important, the use of fluorescent dyes as SYPRO[®] Ruby get more popular as they show a higher linear detection range [41]. As Coomassie was used in this work, the dye is described closer as well as the more sensitive silver staining.

Independent of the staining method, the gels are normally scanned to digitalize the information. Modern software tools can be applied for quantization and comparison of the scanned gels.

Coomassie dyes

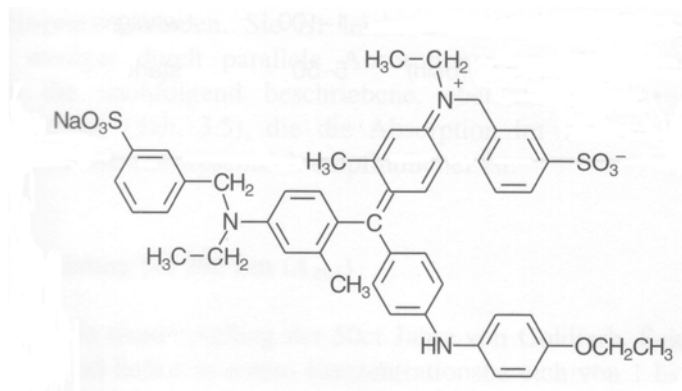


Figure 1.3-4 Coomassie brilliant blue G-250

Most commonly Coomassie G and R (figure 1.3-4), both anionic triphenylmethane dyes showing a high affinity to proteins and high extinction coefficients at a wavelength of 610 nm, are used [42]. Protein visualization with these dyes is normally carried out in two steps. First the whole gel is saturated with the blue dye and then destaining is performed. The matrix has a lower affinity to the dye than the proteins, whose hydrophobic amino acid side chains can interact with the organic molecules. Therefore the gel destains first leaving visible blue protein bands or spots with a typical protein content of 100 ng to 1 μ g. By leaving the gel longer in the destaining solution it is possible to completely remove the non covalent staining, which is a great advantage compared to other methods. On the other hand, proteins featuring an extreme pI or glycosilation can present a problem for this staining technique.

Silver staining

Silver staining is mostly used if Coomassie dyes are too insensitive as its detection limit goes down to less than 1 ng [43]. It is also more time consuming to apply than Coomassie staining. The treatment with silver ions leads to the formation of reduced silver seed crystals at the amide bonds of the proteins [44]. This process happens before the addition of sodium carbonate. As soon as this component is added, the remaining silver ions are reduced to elemental silver. The already built silver seed crystals accelerate the reaction

and therefore lead to a stronger staining at the site of the proteins. When the optimum staining difference between the protein spots and the background is reached, the reduction is stopped with a weak acid. The gel finally shows brown to black protein spots on an often slightly yellow background. It is possible to destain it again with a combination of potassium ferricyanide and sodiumthiosulphate which complex the elemental silver. Extensive washing is necessary if the gel is used for tryptic digest.

1.3.4 MALDI-TOF-MS

1.3.4.1 Introduction

Mass spectrometry is used for molecular weight determination of ionized analytes in a high vacuum (e.g. less than 10^{-4} Pa for a time-of-flight instrument). Mass spectrometers consist of an ion source to convert the analytes into gaseous ions, a mass analyzer to separate them according to their mass/charge ratio and a detector to detect and quantify them. For ion generation and mass separation various techniques are applied depending on the analytes and the required accuracy, m/z range as well as sensitivity. Matrix-assisted laser desorption / ionization (MALDI) and electrospray ionization (ESI) are commonly used for biochemical applications because they allow gentle desorption/ionization of macromolecules nearly without fragmentation. Therefore molecular weights of e.g. proteins can be determined with these techniques. Different ion separation techniques could be applied with both methods. Usually MALDI instruments are equipped with a time of flight (TOF) analyzer, whereas this technique was not used for quadrupole or ion traps. ESI is preferably coupled with quadrupoles or ion traps.

In this study a MALDI-TOF instrument was used. The MALDI technique was developed from laser desorption and ionisation (LDI) mass spectrometry. The first experiments using a laser as primary beam source for mass spectrometry of organic molecules were already carried out in the 1970s [39]. Thin layers of sample on a metallic surface were irradiated with pulsed or cw laser beams leading to a direct desorption/ionization of excited molecules, but also to a high degree of fragmentation. Therefore the peak intensities were low and substances with masses over 2 kDa could not be detected intact. These characteristics limited the applications of laser desorption/ionization (LDI) mass spectrometry in biochemical analysis.

In 1987 the invention of a new sample application method by Karas, Hillenkamp and Tanaka resulted in a breakthrough [45-47]. The new technique was called matrix-assisted laser desorption/ionization (MALDI) because the analytes were dissolved in a solid/liquid matrix before desorption/ionization. By means of small organic molecules or nanoparticles with high absorption coefficients at the specific laser wavelength it was possible to desorb and ionize macromolecules nearly without fragmentation. Tanaka's research focused on MALDI-MS gained him the Nobel Prize in Chemistry in 2002. Today MALDI-TOF-MS is frequently used in biochemical analysis because of its very high m/z range, which is

theoretically unlimited (practically up to 1 MDa), mass accuracy (below 0.1% error) and sensitivity (detection limit in the attomol range).

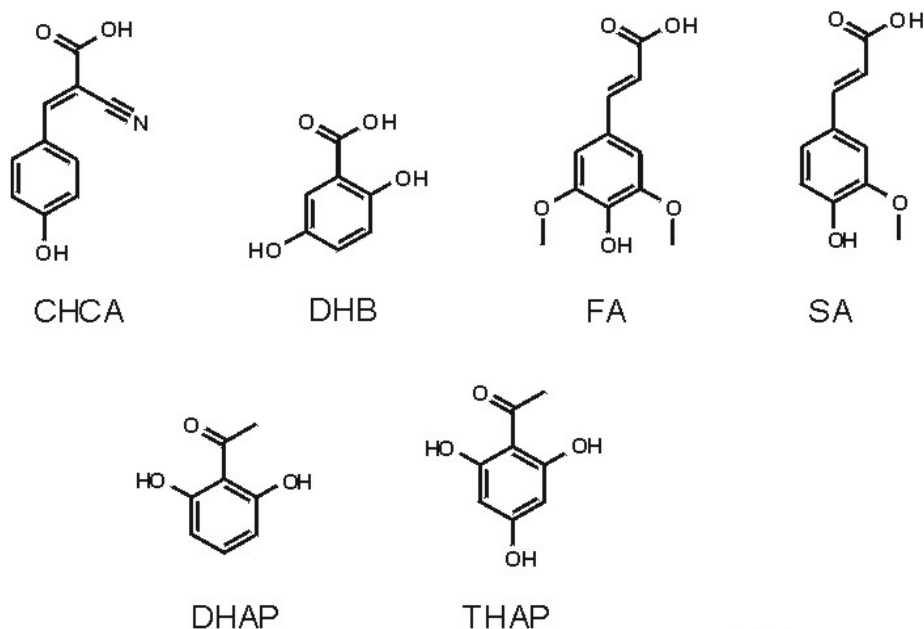


Figure 1.3-5 Chemical structures of some commonly used MALDI matrices:

CHCA: α -Cyano-4-hydroxy-cinnamic acid, DHB: 2,5-dihydroxy benzoic acid, FA: ferulic acid, SA: sinapinic acid, DHAP: 2,6-dihydroxyacetophenone, THAP: 2,4,6-trihydroxyacetophenone

1.3.4.2 Sample preparation for MALDI-TOF-MS

For matrix assisted laser desorption/ionization the sample has to be mixed with a matrix. This matrix is usually one of the small organic molecules shown in figure 1.3-5, but also other substances can be employed, as long as they have a high absorption coefficient at the applied laser wavelength and are able to transfer the received energy partly to the sample. The best suitable matrix for a special problem has to be discovered empirically, even though some recommendations are available for special classes of molecules. Transfer of laser energy is the most important function of a matrix, but it also isolates the analyte molecules from each other, inhibiting interactions, and cares for their ionization through e.g. proton transfer.

To ensure an ideal desorption/ionization process, matrix and sample have to form a crystalline layer on the target, which is usually a metal plate with a defined surface, e.g.

polished stainless steel. In the last years new target surfaces were developed to improve the sensitivity, for example AnchorChip™ targets featuring a hydrophobic Teflon surface with hydrophilic gold anchors on which the matrix analyte mixture concentrates [48] or DIOS-targets (desorption/ionization on silicon), which have a porous silicone surface and do not require a matrix [49].

Both sample and matrix are normally applied as solutions in water and/or organic solvents. Thereby highest purity is required, as salts, buffers, detergents and other contaminations can interfere with the co-crystallization. The deposition of the analyte with a typically 1 000 to 10 000-fold molar excess of matrix can be performed by four different methods [50].

- Volume technique: the two solutions are mixed in an Eppendorf tube and an aliquot is then applied on the target. The miscibility and solubility of all components in the mixture has to be ensured.
- Dried droplet method: the mixing of the solutions is carried out directly on the target. The same requirements as for volume technique are valid.
- Thin layer technique: matrix solution is left to dry on the target and on the thin homogeneous solid layer the sample is applied. The normally aqueous sample solution dissolves a part of the matrix and forms the co-crystals on the surface of the preparation.
- Sandwich method: matrix and sample are applied analogous to the thin layer technique, dried and additionally covered with a further matrix layer.

The final amount of sample needed lies usually in the range of femtomol.

1.3.4.3 Ion generation

In a MALDI mass spectrometer, ion generation occurs in a high vacuum surrounding at less than 10^{-2} Pa where a pulsed laser beam is focused over an optical system onto the sample / matrix preparation. The wavelength, pulse duration and intensity depend on the applied instrument. Generally UV- or IR-lasers can be used, whose laser beam spot diameters lie in the range of 30-150 μm . Pulse duration of a few ns for UV- and up to 150 ns for IR-lasers and a laser intensity of $10^6 - 10^7$ W/cm² are commonly applied. Examples for lasers (and their wavelengths) are N₂-lasers (337 nm), Nd:YAG-lasers (355 or 266 nm), Er:YAG-lasers (2.94 μm) and CO₂-lasers (10.6 μm).

When the laser beam hits the matrix-sample crystals a cascade of events starts, leading finally to the production of charged analyte molecules [51]. Not all mechanisms are completely understood, but the following steps are presumed:

The matrix takes up the energy which leads to the desorption of matrix molecules and cluster ions into the gaseous phase. Further the laser energy causes excitation states in the matrix molecules provoking ionization processes. While the clusters and aggregates disintegrate, the charge is transferred from the ionized and / or radical matrix molecules to the analyte.

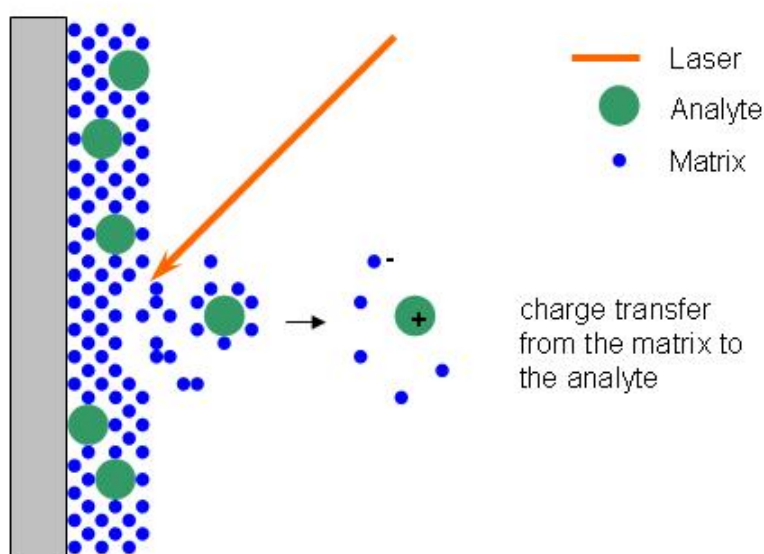


Figure 1.3-6 Principle of MALDI ion generation

Most often a proton is added or removed, but also adducts of cations or anions can result in charged analytes. Additionally desorption of preformed ions and radical ion generation by the loss of an electron are potential sources of gaseous ions [52].

Depending on the applied mode, either the positively or the negatively charged gaseous molecules are accelerated in the direction of the mass analyzer by a potential difference of 3 - 40 kV between the sample slide and a grid or lense (distance 3 - 15 mm). In the first years of MALDI MS this electric field was always applied constantly but since the invention of time-delayed extraction (or pulsed extraction) in 1995 [53], many instruments have the possibility to switch it time-displaced to the laser pulse. With this method the variation of the kinetic energy of produced ions can be reduced as ions with higher initial velocity moving further into the direction of the mass analyzer are accelerated less than

their slower counterparts when the field is switched on. The delay time has to be optimized for a certain mass range and enhances the resolution for this region.

1.3.4.4 Mass analyzer

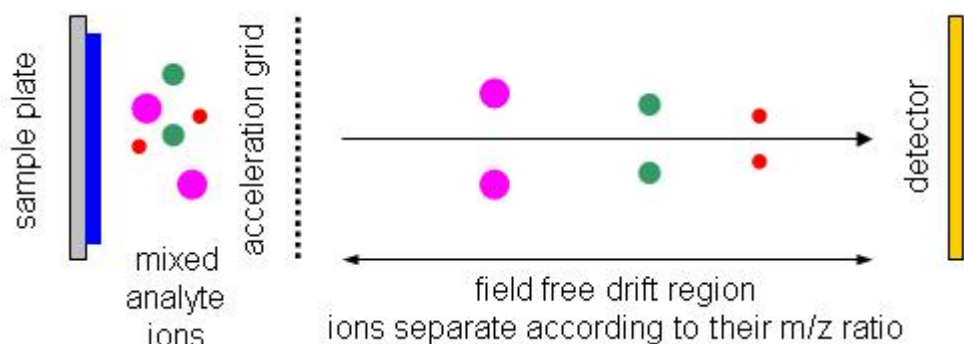


Figure 1.3-7 Principle of a linear TOF analyzer

MALDI instruments are often equipped with a time-of-flight (TOF) mass analyzer. The principle of this technique is the fact that ions accelerated with a constant potential difference show different velocities according to their mass/charge ratio and therefore also need different times to pass the same distance. The starting point for time measurement is given by the laser pulse. Ion formation as discussed in the previous chapter takes only a few ns and the ions are then accelerated towards the detector. High mass ions with low charge gain less velocity than low mass ions or those with higher charge. Once in the field free drift region which is usually between 0.1 and 4 m long, they separate according to their mass/charge ratio. The relation between the mass/charge ratio and the time when the ions reach the detector can be calculated with equation 1.3-7, which is valid as long as no friction forces influence the velocity. (m - mass, v - velocity, U - applied potential, z - number of elementary charges of the ion, e_0 - elementary electric charge, E_{kin} - kinetic energy, s - length, t - time)

$$\frac{m \cdot v^2}{2} = U \cdot z \cdot e_0 = E_{kin}$$

$$\frac{m}{z} = \frac{2U \cdot e_0}{v^2} \quad v = \frac{s}{t}$$

$$\frac{m}{z} = \frac{2U \cdot e_0}{s^2} \cdot t^2$$

Equation 1.3-7

The condition is considered as fulfilled because the drift region is under a vacuum with less than 10^{-4} Pa. During an experiment the applied potential and the length of the drift free region are constant. Therefore it is possible to calculate the mass/charge ratio of an ion by measuring the time it takes before it reaches the detector. Usually the correlation is not computed but determined by an empirical calibration function.

Even though theoretically all ions with the same mass/charge ratio arrive at the detector at the same time, in reality they show a kinetic energy as well as a time and place distribution before the acceleration which influences their flight time. To reduce these effects the “reflectron” TOF was introduced in 1973 [54] and advanced to a “curved field reflectron” TOF in 1993 [55].

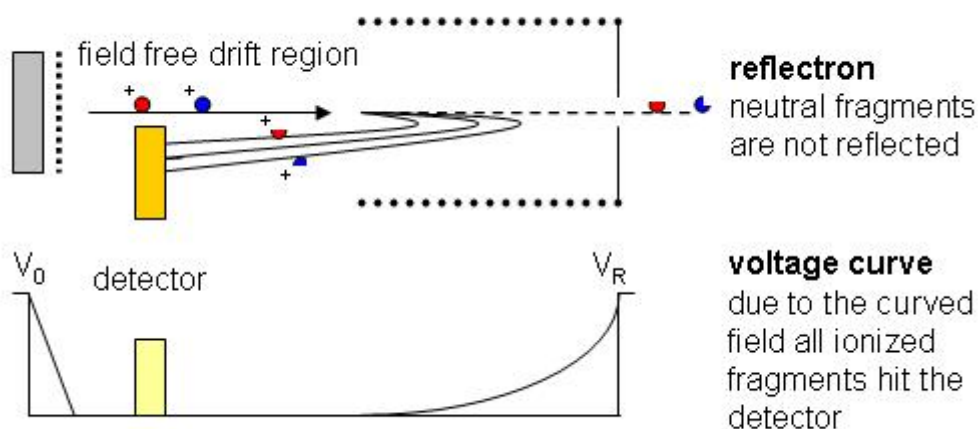


Figure 1.3-8 Principle of a curved field reflectron TOF

A reflectron TOF possesses an ion reflector, which is an electric potential gradient, at the end of the linear drift region where the ions turn around to finally reach a detector close to the ion source. Ions with higher kinetic energy penetrate deeper into the reflectron field as those with lower kinetic energy and therefore need to pass a slightly longer way until – in case the geometry and voltage are optimized – both finally reach the detector at the same time. This focusing effect and the longer drift region (e.g. 2 m) are responsible for the increased resolution of reflectron TOF spectra.

In every mass analysis, a part of the analyte ions fragment during the field free drift in a process called post source decay (PSD). In linear TOF instruments the fragments have the same velocity as their parent ion and therefore are detected at the same mass/charge ratio (m/z). However, their kinetic energy is not the same, so they are reflected differently in a reflectron TOF, reaching the detector at a time that corresponds neither to the mass of the

parent ion nor their own. In PSD experiments therefore an additional calibration has to be applied to obtain the correct mass values. Of course only charged fragments are reflected, whereas neutral fragments are lost in the reflectron. Because of these effects a reflector TOF is less sensitive than a linear instrument, but at the same time the fragmentation pattern provides further information about the analytes.

With the original reflectron instruments, most of the fragment ions were lost as they did not reflect onto the detector due to their lower kinetic energy. Therefore curved field reflectron instruments were developed by altering the reflectron field and the entrance point of the ion current. In this way the focusing of the ions onto the detector was improved and the measurement of nearly all charged ion fragments generated during the first drifting time was possible. At the same time the newer instruments are equipped with an ion gate. This makes it possible to systematically select a parent ion and study its fragmentation pattern while suppressing all other primary analyte ions.

1.3.4.5 MS/MS experiments of peptides

MS/MS or MS² means that an analyte ion is selected in the first step MS experiment, subjected to fragmentation and the fragments are again analyzed by MS. For example a digested protein is analyzed with mass spectrometry, leading to a set of peptide ions. One of these peptide ions is chosen, fragmented and the resulting pattern is again analyzed with MS. Depending on the mass analyzing technique this step can sometimes even be repeated again by fragmenting one of the generated fragments to get MS³ or through further repetition MSⁿ information.

TOF/RTOF instruments are only capable of MS² experiments. The fragmentation of the primary analyte ions can happen in two different ways. Either the ions decay by themselves during the first drift region, leading the previously described PSD, or they are collided with an inert gas like helium in a collision gas cell and fragment. This method is called collision induced dissociation (CID). If the precursor ion collision energies are in excess of 1 kV (e.g. 20 kV) it is called “high-energy” CID, whereas energies in the range of 100 V characterise “low-energy” CID [56].

If peptides are subjected to PSD or CID experiments, they show fragmentation patterns depending on their specific amino acid sequence. The preferred locations of fragmentation, mostly along the peptide backbone, are known [56]. Therefore it is possible to deduce sequence information from a MS/MS spectrum. Peptide fragments are conventionally labeled as shown in figure 1.3-9 [57, 58]. In PSD experiments predominantly b and y ions

can be found, as well as losses of neutral molecules like water or ammonia. On the other hand high energy CID experiments can lead to different fragment ion types shown in figure 1.3-9 [56]. Especially immonium ions (figure 1.3-9 b) and fragments from observed side chain losses (figure 1.3-9 c) can be additionally found in such spectra.

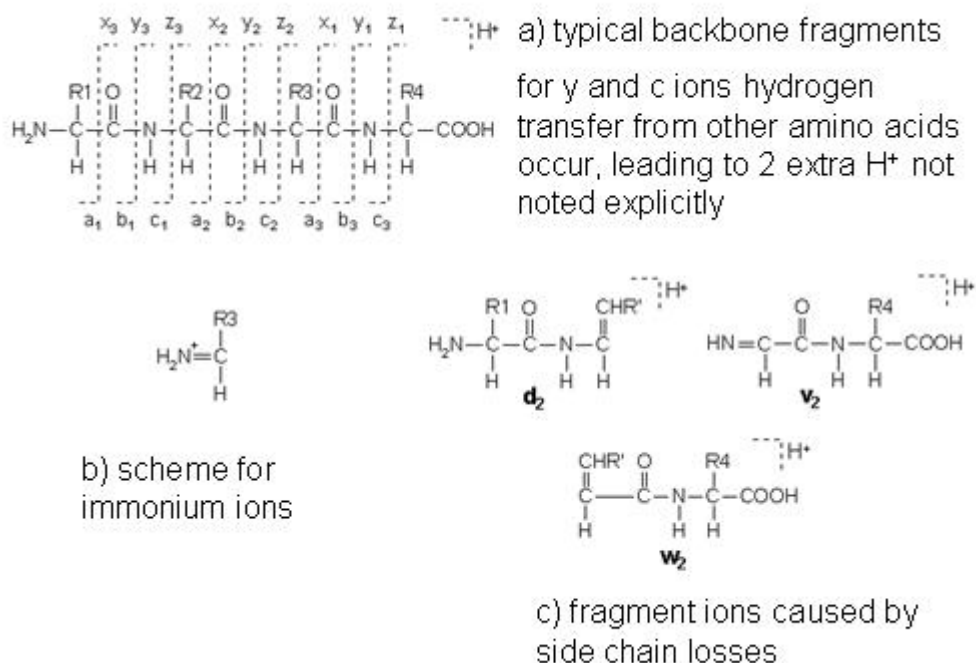


Figure 1.3-9 Fragmentation pattern of peptides

As the exact fragmentation mechanisms are not fully understood and are further depending on the neighbouring amino acids, it is difficult to predict which of a peptide's potential fragments will be generated and at what abundance. Only if an ion sequence of the same fragment type, like consecutive y ions, is present in a spectrum, a manual interpretation is possible. Additionally complexity is added by the fact that some amino acids have exactly the same mass (I and L) and can only be distinguished if other characteristic ions resulting from side chain cleavages (d, v, w ions) are present in the spectrum.

1.3.5 Protein Database Search

1.3.5.1 Introduction

In the last years, the amount of data resulting from mass spectrometry experiments of proteins has increased dramatically and it would be impossible to interpret it without the aid of bioinformatics. Protein identification with the aid of mass spectrometry got popular after the invention of MALDI and ESI as gentle ionization techniques with low detection limits. The idea that the peptide mass set derived from the digestion of a protein with a specific enzyme is characteristic for it was already born in 1989 [59], but the use of these “peptide mass fingerprints” for protein identification through database search was simultaneously developed by five research groups in 1993 [60-64]. In 1994 Mann and coworkers showed that also “sequence tags”, the combination of a peptide mass and a short part of its sequence along with the fragment ion masses denoting the beginning and end, can be applied for protein identification [65]. Additionally Eng and colleagues proved the possibility to directly use MS/MS data of peptides for the same goal [66].

With the progression of computer development and internet use, the amount of data available increased and the first search engines accessible through the World Wide Web evolved, among them Protein Prospector (UCSF) [67], ProFound (Rockefeller University) [68] and Mascot (Matrix Science) [69]. Not only protein databases were used as information source, but also DNA databases were adapted to the search by translating them into proteins in all six possible frames [70, 71]. Different scoring algorithms were developed to facilitate the differentiation of true results from insignificant matches. All in all the methods improved constantly and the amount of identified proteins presented each year would be impossible without the progress in the development of bioinformatic tools.

1.3.5.2 General aspects of database searches

Databases are huge information pools where a single search criterion most often fits a great amount of entries, like there will be uncountable proteins with a pI of 5. Only the combination of different properties reduces the matching results until in the best case just one correct match is left. Therefore it might seem best to choose the most restrictive search parameters. However, if the restriction is too tight, the correct match might be discarded erroneously. For example if the mass tolerance is set too low, peptides which in reality belong to the protein are excluded leading to a lower significance of the result. Some other restrictions are not determined by the data but depend on the content of the database, e.g. it

can be sensible to set the taxonomy to a single species if its whole genome is in the database but it is better to include related species as well if this is not the case. For every search, or at least every set of searches, the parameters have to be considered carefully to get an optimal result.

Not only the search parameters but also the quality of the database has to be kept in mind. It is possible that it contains redundant or erroneous data, especially as many protein entries are derived from DNA sequencing experiments which are prone to sequencing errors.

1.3.5.3 Peptide mass fingerprint search

Peptide mass fingerprints (PMFs) were the first mass data used for protein identification. They have the advantage that they can be measured on every MALDI or ESI mass spectrometer. They consist simply of the m/z set derived from mass spectrometry of a digested protein. These data are compared with the theoretical peptide m/z sets that should result from a theoretical digestion of the proteins in the searched database. A match can be found if a sufficient amount of peptide masses can only be assigned to a single protein. The quality of the result is best if great parts of the protein's sequence can be found as peptides (i.e. a high sequence coverage) and few peptides cannot be explained as results of a digest of the protein.

Every modification causing that the mass of a peptide changes (e.g. alkylation of the cysteins, posttranslational modifications or a different amino acid at one position) would prevent the assignment, therefore search engines allow to specify fixed and variable modifications in the search criteria. A list of potential modifications and the resulting mass differences can be found in Unimod (www.unimod.org, [72]). Amino acid exchanges can not be defined as modifications and therefore reduce the amount of correctly assigned peptides. This is especially important if the proteome of the examined species is not in the database but only the one of a related one. Unless the protein sequence is not highly similar, no match will be found.

It is also nearly impossible to get a good match from a database whose entries do not correspond to entire proteins with this method as its statistics rely on the fact that a single defined sequence is the origin of the peptides. Translations of DNA databases whose entries do not correspond to open reading frames but to protein fragments (e.g. expressed sequence tag data) or more than one protein (e.g. untreated genomic data) are not the ideal choice for peptide mass fingerprint search.

It is also difficult to get good results if the sample represents not only one protein but a mixture. In some cases it is possible to assign the peptides to two or three components, but the significance is often low.

Enhanced identification can be performed with MS/MS data of the peptides and the application of one of the other search algorithms.

1.3.5.4 MS/MS ion search

MS/MS ion search can be performed directly with uninterpreted MS/MS data and is therefore also suitable for automated search. The algorithm tries to fit the MS/MS fragments to the sequence of the peptides in the database by applying the fragmentation patterns discussed in chapter 1.3.4.5. Some algorithms also interpret the intensity of the individual peaks [69].

To identify a peptide, its exact amino acid sequence has to be in the database. Any modification will prevent its identification in standard search. However, as a protein digest contains more than one peptide, it is probable that at least one of them fits to the database entry. To find modified or unspecifically cleaved peptides within an identified protein, some search engines allow a second pass search with modified search criteria in a limited number of protein results. For example Mascot search engine abandons enzyme specificity and tests the complete list of modifications serially as well as the set of amino acid substitutions that can arise from single base substitutions or – in case of nucleic acid sequences - insertions or deletions. Nevertheless peptides carrying more than one modification cannot be identified through this method.

Of course there are peptides that can be found in more than one protein. Then the combination of MS/MS data from more than one peptide belonging to the same protein is necessary. In case a digested protein mixture is examined, it is sometimes difficult to decide if just one of the potential proteins is present or if more of them are part of the mix.

1.3.5.5 Sequence query

For a sequence query, at least a part of one peptide sequence has to be available. It is often possible to obtain such information of an MS/MS spectrum even if not all ions of one fragment type are visible. Sequences of for example three amino acids are not long enough for sequence homology searches, but if this information is combined with the information of the peptide mass and the ion masses marking their beginning and end, they can be used for protein identification. A query of this kind is called sequence tag and the specified

sequence is used as filter during the search, therefore the search is rapid compared to MS/MS ion search. It is even possible to allow error tolerant search by relaxing the peptide mass constraint so that mass differences either before or after the tag are ignored. With this method it is possible to identify peptides that contain an amino acid change in comparison to the database. Using this method usually requires manual interpretation of the MS/MS data and is only suggested if the peptide mass fingerprint and the MS/MS ion search failed.

1.3.5.6 The Mascot search engine

The Mascot search engine was developed in 1998 [69] as a successor of MOWSE [63], which was one of the first programs to identify peptide mass fingerprints. It includes peptide mass fingerprint search, sequence query and MS/MS ion search possibilities and is freely available on the internet (www.matrixscience.com) as well as an in-house server version. It uses a probability based scoring system whose principal algorithm was already developed for its antecessor MOWSE.

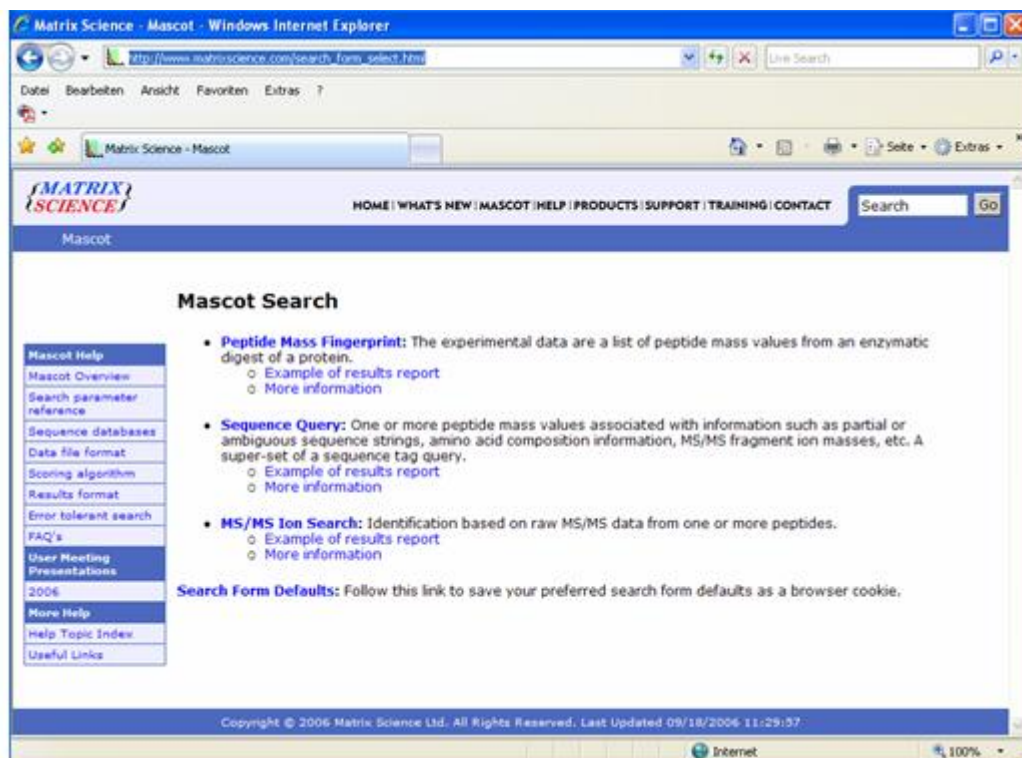


Figure 1.3-10 Mascot internet search portal

In the public online version of Mascot the databases MSDB, NCBIInr, SWISS-PROT and dbEST can be searched as well as a randomized database to control the significance of results. Using in-house server versions allows the usage of earmarked databases. In this

work, a tailor-made bioinformatic facility was used which contained the genome sequence information from the fungi *Aspergillus nidulans*, *Giberella zeae*, *Magnaporthe grisea*, *Neurospora crassa*, *Ustilago maydis* and *H. jecorina* and the TrichoEST database, which includes 26 different cDNA libraries derived the following *Hypocrea/Trichoderma* strains: *H. atroviridis* (*T. atroviride*) B11 and P1, *H. lixii* (*T. harzianum*) A6, CECT2413, T3, T22 and Th2, *H. virens* (*T.virens*) T59, *T. asperellum* T53, *T. longibrachiatum* T52, *T. stromaticum* and *T.viride* T78 [22, 23].

1.4 Aims of this study

This study was focused on the extracellular proteome of *H. atroviridis*. As described in chapter 1.1, many species of the *Hypocrea/Trichoderma* genus are important biocontrol organisms and a range of extracellular enzymes are involved in biocontrol related mechanisms. Revealing the identity of these secreted enzymes was the main goal of this work. A better understanding of the biocontrol mechanisms is important to evaluate and improve the use of *Hypocrea/Trichoderma* as biological fungicides.

As a first step the secretome of *H. atroviridis* from cultures grown on the simple carbon source glucose, was studied to investigate which proteins are constitutively secreted by *H. atroviridis*.

As a second step the proteins secreted by *H. atroviridis* when grown with chitin or antagonist cell walls (*Rhizoctonia solani*, *Botrytis cinerea*, *Pythium ultimum*) as carbon source were studied. A large increase in the variety and concentration of extracellular proteins, especially hydrolytic enzymes, was expected. Identification the proteins was part of the work as well as the qualitative comparison of the secretomes.

MALDI-MS and -MS/MS experiments were applied for the protein identification. Available genomic and EST databases of *H. atroviridis* and its relatives were used for searching. Additionally searches in the protein databases MSDB, NCBI nr and SWISS-PROT were employed as well.

2 Methods

2.1 Used chemicals

Acetone, p.a. (Merck, Darmstadt, Germany)

Acetonitrile, ACN, p.a. (Merck, Darmstadt, Germany)

Adrenocorticotrophic hormone fragment 18–39, ACTH (Standard peptide; Sigma, St. Lois, MO, USA)

Ammonia solution, 25% (Merck, Darmstadt, Germany)

Ammonium hydrogencarbonate, NH_4HCO_3 , >99,5% (Fluka, Buchs, Switzerland)

α -Cyano-4-hydroxy cinnamic acid, CHCA (Sigma, St. Lois, MO, USA)

Bradykinin fragment 1–7 (Standard peptide; Sigma, St. Lois, MO, USA)

DL-Dithiothreitol, DTT, >99% (Sigma, St. Lois, MO, USA)

Ethanol, p.a. (Merck, Darmstadt, Germany)

Formic acid, HCOOH , p.a., ~98% (Fluka, Buchs, Switzerland)

Human angiotensin II (Standard peptide, Sigma, St. Lois, MO, USA)

Invertase, proteomics grade (Sigma, St. Lois, MO, USA)

Iodoacetamide, IA (Sigma, St. Lois, MO, USA)

Methanol, p.a. (Merck, Darmstadt, Germany)

P₁₄R (Standard peptide, Sigma, St. Lois, MO, USA)

PNGase F, proteomics grade (Sigma, St. Lois, MO, USA)

Trypsin, modified, sequencing grade (Roche, Basel, Switzerland)

Water, ultra high quality, UHQ

2.2 Sample preparation and 2D gel electrophoresis

The starting material for further analysis carried out during the presented diploma thesis were 2-D gels of extracellular proteins prepared by Dr. Verena Seidl [73, 74]. As the methods applied for the generation of the gels are important for the interpretation of the results, the major steps are outlined in this chapter in accordance with the mentioned literature [73, 74].

Hypocrea atroviridis P1 (ATCC74058) was grown in submersed cultures (48h, 25 °C, 150 r.p.m., medium composition: 0.68 g/L KH₂PO₄, 0.87 g/L K₂HPO₄, 1.7 g/L (NH₄)₂SO₄, 0.2 /L KCl, 0.2 g/L CaCl₂, 0.2 g/L MgSO₄·7H₂O, 2 /L FeSO₄·7H₂O, 2 mg/L, MnSO₄·7H₂O, 2 mg/L ZnSO₄·7H₂O) on glucose (1% w/v), colloidal chitin (1% w/v) and *Rhizoctonia solani*, *Botrytis cinerea* and *Pythium ultimum* cell walls. Culture supernatants were isolated by filtration and purified and concentrated by ultrafiltration, dialysis and trichloroacetic acid / acetone precipitation. The pellets were resolubilized in 2D sample buffer (9 M urea, 2% CHAPS, 1% DTT, 0.5% carrier ampholytes, 0.1% (v / v) protease inhibitor cocktail (Sigma, St. Louis, MO, USA)) and stored at -80°C.

For the first dimension of the 2D gel electrophoresis, pH 4-7, 17 cm immobilized pH gradient (IPG) strips (Bio-Rad) were rehydrated overnight with 300 µg extracellular proteins dissolved in 300 µL 2D sample buffer. The focusing of the IPG strips was then carried out in the IEF cell (Bio-Rad) by running through the following program: a linear ramp from 0 to 300 V over 1 h, a linear ramp from 300 to 1000 over 1h, a linear ramp from 1000 to 10 000 V over 2 h, and 60 000 Volthours at 10 000 V_{max} with a limit of 50 µA per IPG strip.

To prepare the IPG strips for the second dimension, the strips were equilibrated in equilibration buffer (6 M urea, 2 % SDS, 0.05 M Tris/HCl, pH 8.8, 20 % glycerol) containing 2% dithiothreitol and equilibration buffer containing 2.5 % iodoacetamide for 15 minutes each. Afterwards they were mounted on 12 % SDS polyacrylamide gels. On the left side of the strips, PageRuler Prestained Molecular Weight Marker (Fermentas, Burlington, Canada) was loaded to allow molecular mass determination. For SDS-PAGE, 25 mA per gel for the stacking gel and 35 mA per gel for the separating gel were applied.

Four to five replicates of gels were made for each sample. The proteins were visualized by staining them with Simply Blue (Invitrogen, Paisley, UK). The stained gels were stored in a solution containing 20 % NaCl and 0.05 % NaN₃ (to prevent microbial growth on gels).

2.3 Tryptic Digest

2.3.1 Applied solutions

NH₄HCO₃ buffer

50 mM NH₄HCO₃

pH 8.5 adjusted with ammonia solution

Dithiotreitol solution

10 mM DTT in 50 mM NH₄HCO₃

Iodoacetamide solution

55 mM IA in 50 mM NH₄HCO₃

Trypsin solution

12.5 ng Trypsin / μ L in 50 mM NH₄HCO₃

Extraction Solution

5% HCOOH / ACN (1/1, v/v)

2.3.2 Procedure

The tryptic digest was carried out as described by Shevchenko et al [44]. Protein spots and a gel-blank, an equally sized piece without detectable proteins, were cut out of the gel with a stainless steel scalpel. In case of intensive spots bigger than 7.5 mm³ aliquots of approximately that size were taken.

The gel pieces were cut into cubes of approximately 1 mm³ and transferred into a microcentrifuge tube. They were washed two times with 30 μ L UHQ and then two times with 30 μ L of a mixture UHQ/ACN 1:1 (v/v) for ten minutes in each case. Afterwards 30 μ L of pure ACN were added to the gel pieces until they shrunk due to dehydration. Then they were rehydrated with 30 μ L NH₄HCO₃ buffer to which the same amount of ACN was added after 5 minutes and they were incubated for further 10 minutes. After the liquid was removed the gel pieces were dried in a vacuum centrifuge.

To ensure that all disulfide bridges were reduced, the gel pieces were incubated in 30 μ L of dithiotreitol solution for 45 minutes at 56 °C on a thermoblock. To avoid the formation of

new disulfide bridges the cysteins were derivatized by incubation with 30 μL iodoacetamide solution for 30 minutes in the dark. Subsequently the solution was replaced with 30 μL NH_4HCO_3 buffer to which the same amount of ACN was added after 5 minutes. 10 minutes later the whole liquid was removed and the gel pieces were again dried in a vacuum centrifuge.

Then an aliquot of trypsin solution was added to cover the gel pieces and they were fully rehydrated at 4 $^\circ\text{C}$ for 15 minutes. In case of excessive solution the liquid was removed and 20 μL NH_4HCO_3 buffer were added to all tubes. The tryptic digestion was then carried out in two different ways: either at 37 $^\circ\text{C}$ over night or in a domestic microwave oven at 170 W for 10 minutes. Afterwards the microcentrifuge tubes were centrifuged and cooled to room temperature before 20 μL ACN were added.

A tube only containing trypsin solution was subjected to digestion with every batch as a trypsin blank. In this case of course the liquid was not removed before the addition of the NH_4HCO_3 buffer. In all further steps it was treated identically to the samples.

The digestion solution was extracted from the microcentrifuge tube containing the gel particles after 10 minutes and collected in a fresh tubes. 20 μL extraction solution were added to the gel particles and incubated for 10 minutes. This step was repeated two times, all extracts were pooled and dried in a vacuum centrifuge. Two times 10 μL 0.1% TFA were added to the residue and dried again to remove the carbonic acid. Then the lyophilized peptides were stored at -20 $^\circ\text{C}$.

2.4 PNGase F Digest

2.4.1 Applied solutions

NH₄HCO₃ buffer

50 mM NH₄HCO₃

pH 8.5 adjusted with ammonia solution

PNGase F stock solution

1 U PNGase / μ L

Invertase standard solution

5 mg Invertase / μ L in 50 mM NH₄HCO₃

2.4.2 Procedure

A few chosen protein spots were subjected to PNGase F digest. In this case the following steps were carried out additionally to the in-gel digest procedure before the rehydration of the gel pieces with trypsin solution.

3 μ L PNGase F stock solution and 17 μ L NH₄HCO₃ buffer were added to the dried gel pieces and left 15 minutes at 4 °C for rehydration. The digestion itself was then carried out in two different ways: either over night at 37 °C or in the microwave oven at 170 W for 10 minutes. In each case the microcentrifuge tubes were centrifuged afterwards and cooled to room temperature. The liquid, which contained the released N-glycanes, was collected in a new microcentrifuge tube and stored at -20 °C in case it was needed for further analysis. The gel pieces were dried in a vacuum centrifuge and subjected to tryptic digest in a microwave oven as described above.

With every batch a blank that contained only PNGase solution and buffer and a standard consisting of 1 μ L invertase standard solution treated as the gel pieces were digested in parallel. In these cases the whole liquid was lyophilized without removing the the separated N-glycanes.

2.5 ZipTip purification

Peptides derived from in-gel digest were purified with C-18 (in case of the proteins secreted from *Trichoderma atroviride* when grown on glucose) or μ C-18 (all other proteins) ZipTips[®] (Millipore, Billerica, MA, USA). In both cases the procedure was the same, only the amount of solution aspirated in every step except of the final eluting was 10 μ L for C-18 and 5 μ L for μ C-18 ZipTips[®].

The lyophilized peptides were dissolved in 15 μ L 0.1 % TFA for the purification step.

First the tips were wetted three times with 0.1 % TFA/ACN (1/1, v/v), dispensing the solution to waste. Then they were equilibrated three times with 0.1 % TFA, dispensing again the solution to waste. Afterwards the sample was aspirated and dispensed at least five times to bind the peptides to the reversed phase material. They were then washed three times with 0.1 % TFA and finally eluted with 5 μ L (C-18 ZipTips[®]) or 1 μ L (μ C-18 Zip Tips[®]) 0.1 % TFA/ACN (1/1, v/v) either into a new Eppendorf tube (C-18 ZipTips[®]) or directly on the MALDI-target (μ C-18 ZipTips[®]).

2.6 MALDI-TOF-MS

2.6.1 Instrumentation

The instrument, on which all mass spectrometry experiments were carried out, was an Axima TOF² from Shimadzu Biotech Kratos Analytical (Manchester, UK). It is equipped with a pulsed nitrogen laser (wavelength 337 nm, pulse width 4 ns), a curved field reflector and a collision gas cell (collision gas: helium). All peptide mass fingerprints were recorded in positive ion reflectron mode applying 20 kV acceleration voltage. The setting for delayed extraction was varied between the optimums for ions with m/z 1500 and m/z 2500 to get the best results for every peptide mass finger print.

Mass spectra of peptide mass fingerprints were acquired by accumulating 200 to 1000 single unselected laser shots whereas for PSD and CID MS/MS experiments of characteristic tryptic peptides 1000 to 5000 laser shots were averaged.

2.6.2 Sample preparation

Sample preparation was carried out on stainless steel MALDI target plates using thin-layer preparation technique. The matrix, CHCA, was first dissolved in acetone to a final concentration of 12 mg/mL. 0.5 μ L of this solution were deposited on the MALDI target. After the evaporation of the acetone a thin layer of matrix crystals remained. On this 1 μ L of ZipTip[®] purified sample solution was placed and dried gently at room temperature.

2.6.3 Calibration

Table 2.6-1 List of applied peptides for external calibration

Peptide	Formula	[M+H] ⁺	
		monoisotopic	average
Bradykinin fragment 1-7	C ₃₅ H ₅₃ N ₁₀ O ₉	757.40 Da	757.87 Da
Human angiotensin II	C ₅₀ H ₇₂ N ₁₃ O ₁₂	1046.54 Da	1047.21 Da
P ₁₄ R	C ₇₆ H ₁₁₃ N ₁₈ O ₁₆	1533.86 Da	1534.85 Da
ACTH fragment 18-39	C ₁₁₂ H ₁₆₆ N ₂₇ O ₃₆	2465.20 Da	2466.72 Da

For external calibration before all measurements 0.5 μ L of an aqueous solution containing 2 pmol/ μ L of each of peptide shown in table 2.6-1 were placed on a thin layer of pre-spotted matrix.

2.7 Database search and Bioinformatics

All MS and MS/MS spectra were evaluated with the program Shimadzu Biotech Launchpad 2.7 to get lists of monoisotopic or average m/z-values of the detected peptides or fragment ions respectively. To obtain the figures shown in chapter 3, the PSD and CID spectra were baseline subtracted and smoothed with a Savitsky-Golay algorithm (smoothing filter width 20 channels), whereas spectra of peptide mass fingerprints were neither baseline subtracted nor smoothed. The m/z data sets were submitted to the Mascot search engine, either for peptide mass fingerprint searches or for MS/MS ion searches. In case of MS/MS ion search not only the data from one single peptide were used but also the combined information of all fragmented tryptic peptides belonging presumably to the same protein were submitted to the search engine.

Searches were run against all fungal proteins and DNA sequence information from the public databases MSDB, NCBIInr and SWISS-PROT, the predicted translation frames from *H. jecorina* genome database called TreeseiV12predtran (www.trichoderma.org) and the TrichoEST database (www.trichoderma.org, [22, 23]), which contains cDNA libraries derived from the following *Hypocrea/Trichoderma* strains: *H. atroviridis* (*T. atroviride*) B11 and P1, *H. lixii* (*T. harzianum*) A6, CECT2413, T3, T22 and Th2, *H. virens* (*T. virens*) T59, *T. asperellum* T53, *T. longibrachiatum* T52, *T. stromaticum* and *T. viride* T78.

As the Mascot search engine needs some restrictions for the search in the databases, the following settings were provided if not stated otherwise:

- *Peptide mass fingerprint search*: enzyme trypsin (maximum one missed cleavage), fixed modification carbamidomethyl (C), variable modification oxidation (M), monoisotopic $[M+H]^+$ ions and peptide tolerance ± 0.7 Da.
- *MS/MS ion search*: enzyme trypsin (maximum one missed cleavage) or semitrypsin or no enzyme, fixed modification carbamidomethyl (C), variable modification oxidation (M), average masses, precursor ion charged 1+ and tolerance ± 0.7 Da, product ion tolerance ± 1.0 Da.

The three different enzyme options for MS/MS search were each applied for all peptides and in all searched databases. Only proteins which gave a significant hit in terms of the

probability based MOWSE Score (significance threshold $p < 0.05$) as result of the MS/MS search were treated as identified.

In all other cases the five results with the highest scoring of each database were regarded closer to identify potential series of fragment ions. All series with four or more amino acids were transformed into sequence tags listed among the results.

Additionally the MS/MS spectra were evaluated manually. Therefore the differences of the m/z values of all peaks were compared with the known masses of all amino acids. In some cases sequence tags could be deduced with this method. If these tags were longer than three amino acids they were submitted to the Mascot sequence query as error tolerant tags. Experience showed that most error tolerant tags gave a significant scoring for all resulting protein hits, even when the amount of matching sequences was very high. Therefore the reliability of the scoring is not granted and only single protein hits were regarded closer.

3 Results and Discussion

3.1 Evaluation and comparison of the 2D gels

All 2D gels covered a pI range of 4 – 7 and a molecular weight range from 15 to 200 kDa. Consistent with the expectations *Hypocrea atroviridis* produced more different extracellular proteins when cultivated on host cell walls than during growth on glucose. While the gels of the proteomes resulting from *Rhizoctonia solani* and *Botrytis cinerea* cell walls as carbon sources were very similar each other, the proteins gained from experiments with *Pythium ultimum* seemed to stretch over a narrower pI range. As the distribution pattern showed nevertheless some similarities to the other two gels, all “cell wall” spots were analyzed in parallel experiments.

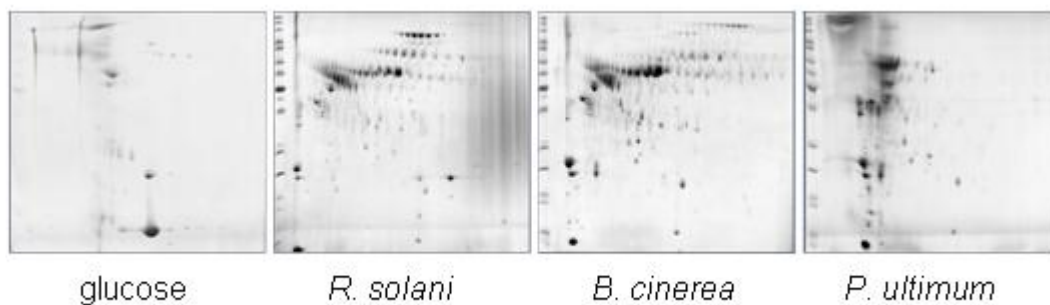


Figure 3.1-1 Overview of all gels

3.1.1 2D gels with glucose as carbon source

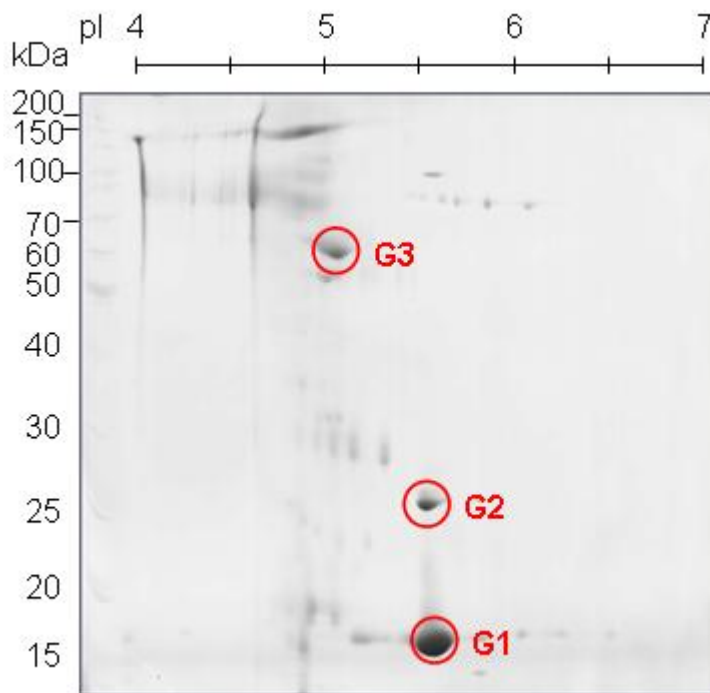


Figure 3.1-2 Extracellular proteome of *H. atroviridis* growing on glucose

The extracellular proteome of *H. atroviridis* grown on glucose is shown in figure 3.1-2. Three dominant spots, labeled G1 to G3 are visible next to some far less intensive spots. These three proteins were digested and then subjected to MS and MS/MS analysis.

3.1.2 2D gels with host cell walls as carbon source

During growth on different host cell walls, *H. atroviridis* secreted the proteomes shown in figures 3.1-3 to 3.1-5. The spots picked from these gels were labelled with the first letter of the host's name (*R. solani*, *B. cinerea*, *P. ultimum*) and a number. Spots carrying the same number are most likely corresponding proteins. In a first attempt the spots with the numbers 1 to 3 were selected for analysis, as they are single spots with a low to medium molecular weight (20 – 30 kDa), do not belong to trains and are visible under all conditions at the same molecular weight and in the same pI region. In a next step, the spots R4 to R6 were picked from the proteome secreted during growth on *R. solani* cell walls as they were further intensive spots under these conditions. Then most likely corresponding proteins on the other gels were searched, leading to the spots B4, B5, P4 and P5. The proteome secreted during the use of *B. cinerea* cell walls as carbon source was used to pick spots B7

to B9. Spot B7 was selected because it belonged to a train which possibly interfered with spot B4. Knowing which peptides belong to B7 might enable the identification of peptides wrongly associated with B4. Spot B8 seemed to correspond to spot G1 and spot B9 was selected for completeness of single spots with high intensity even though its low molecular weight made identification less likely. Also in this case it was tried to match the spots with corresponding proteins from the other gels, leading to the selection of spots P8, P9 and R7 to R9.

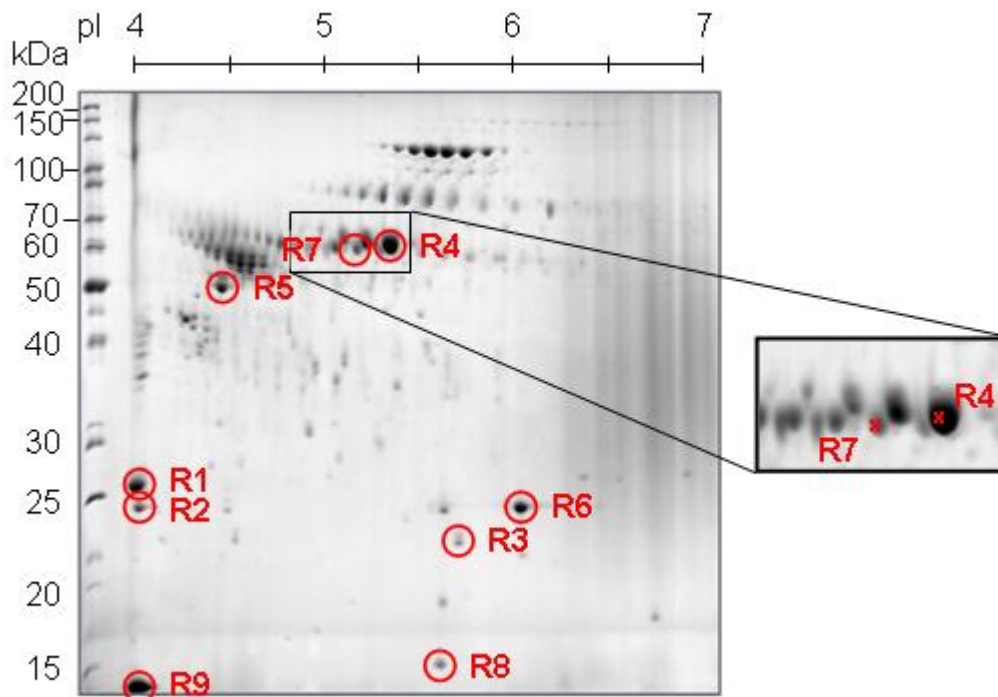


Figure 3.1-3 Extracellular proteome of *H. atroviridis* growing on *R. solani* cell walls

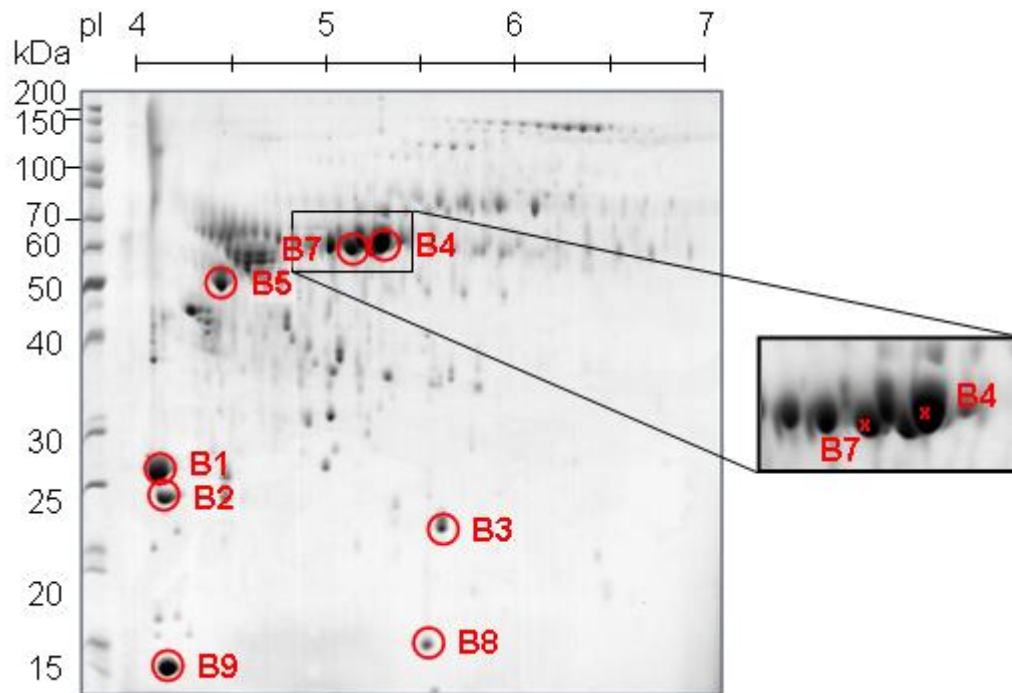


Figure 3.1-4 Extracellular proteome of *H. atroviridis* growing on *B. cinerea* cell walls

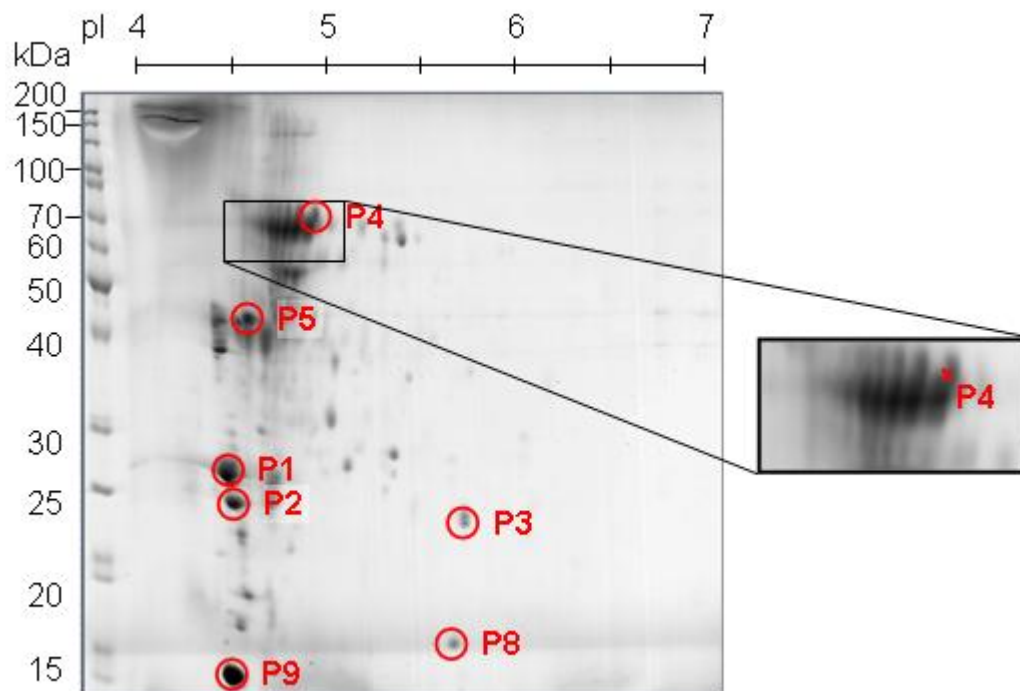


Figure 3.1-5 Extracellular proteome of *H. atroviridis* growing on *P. ultimum* cell walls

3.2 Identification of protein spots from 2D gels with glucose as carbon source

3.2.1 Spot G1 & G2 – Epl1

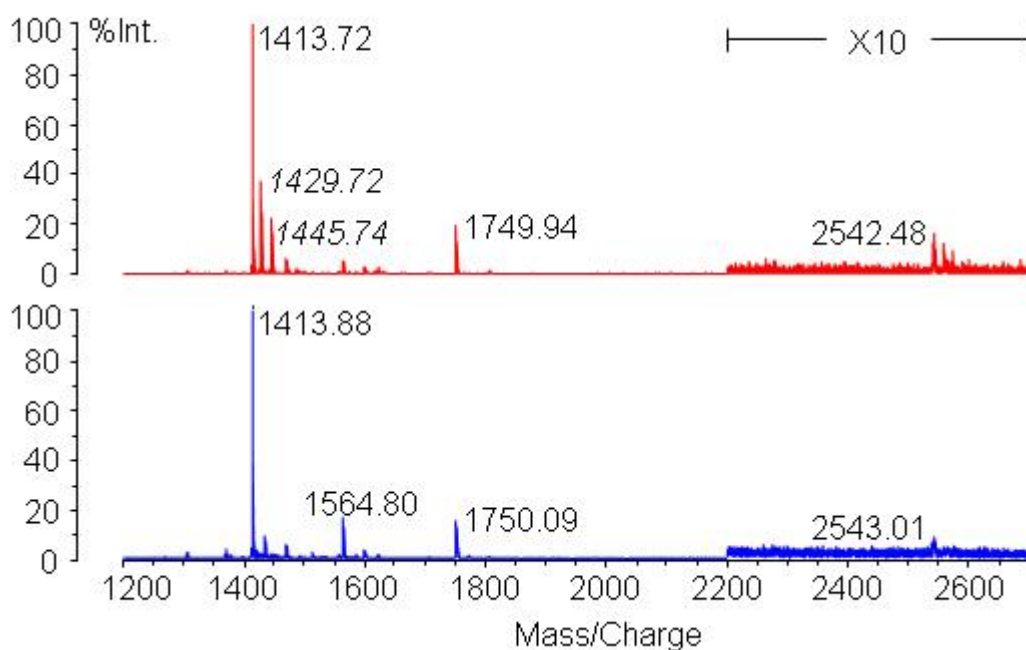


Figure 3.2-1 Peptide mass fingerprints of G1 (above) and G2 (below)

Spot G1 is a protein of approximately 16 kDa with a pI of 5.5 to 5.7, whereas spot G2 has a molecular weight of approximately 27 kDa but a similar pI. MS analysis revealed that their PMFs were identical except for 4 additional peptides observed in G1 but not G2, as can be seen in figure 3.2-1 and table 3.2-1. None of the two PMFs resulted in a significant protein hit in any of the searched databases, so PSD and CID MS/MS experiments were carried out. They showed identical fragment ions for the corresponding peptides a, b and c of G1 and G2, indicating that they represent the same protein. The PSD spectra of the peptides a, b, c and d can be found in figures 3.2-2 and 3.2-3

Table 3.2-1 PMF – monoisotopic m/z values of G1 and G2 peptides (bold = belong to the identified protein)

Peptide	G1	G2	Sequence identified by MS/MS	Mascot ion score
	1305.67	1305.85		
	1369.55	1369.80		
a	1413.72	1413.88	YHWQTQGQIPR	34
a ₁	1429.72		YHWQTQGQIPR + 1 Ox (HW)	49
a ₂	1445.74		YHWQTQGQIPR + 2 Ox (HW)	49
	1470.76	1470.93		
	1491.72	1491.87		
b	1564.67	1564.80	DTVSYDTGYDDASR	124
	1600.79	1600.94		
c	1749.94	1750.09	SLTVVSCSDGANGLITR	50
d	2542.48	2543.01	FPYIGGVQAVAGWNSPSCGTCWK	-
	2558.46			
	2574.42			

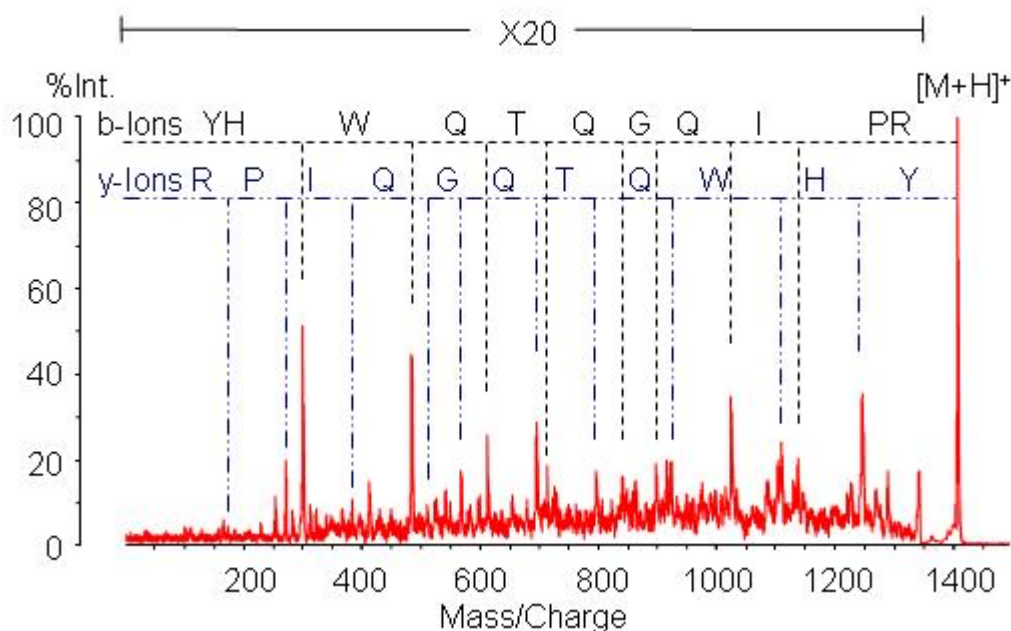


Figure 3.2-2 PSD spectrum of peptide a (m/z 1413.72)

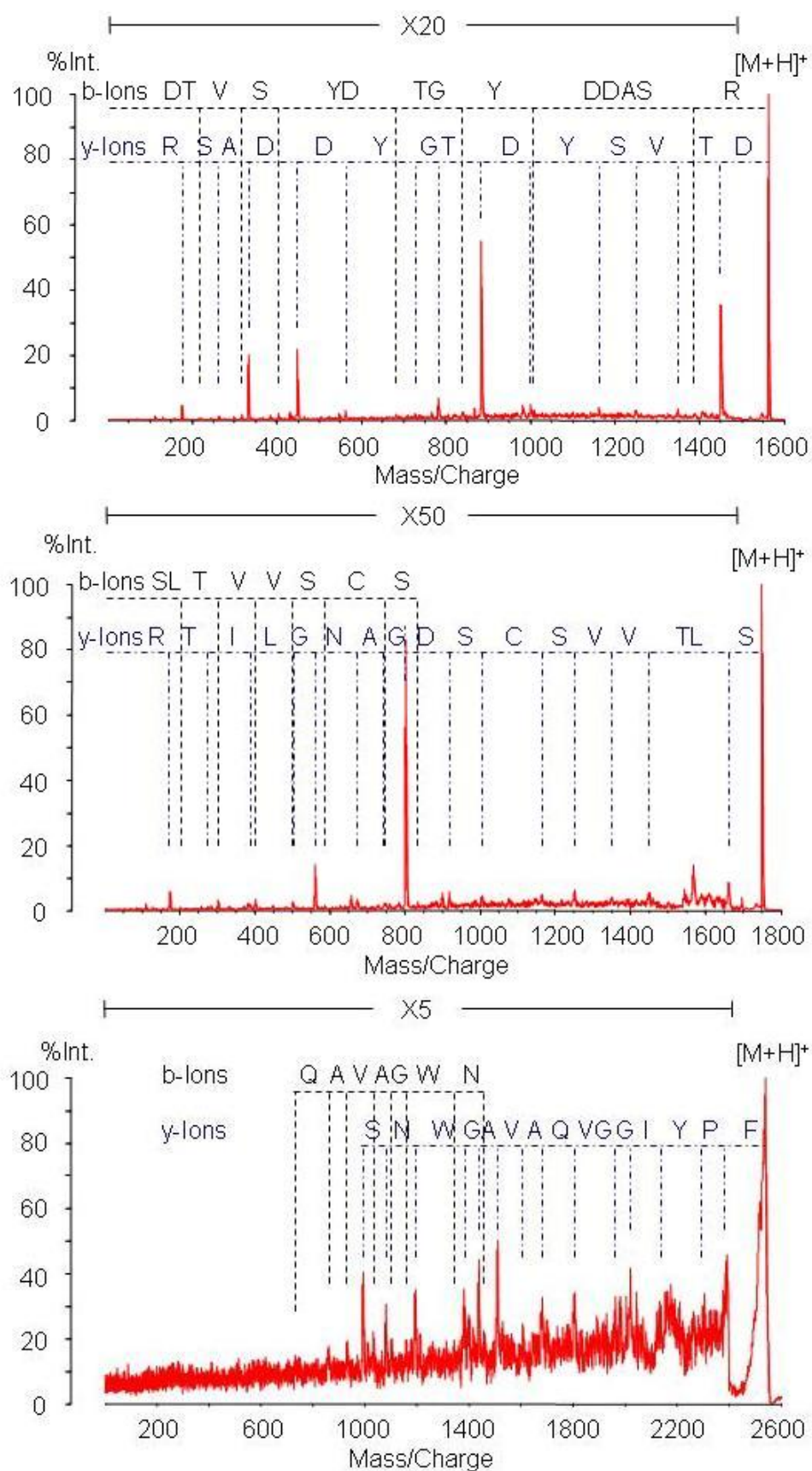


Figure 3.2-3 PSD spectra of peptides b (above, m/z 1564.8), c (middle, m/z 1750.0) and d (below 1542.5)

Mass lists combined with intensity information (see appendix-table 1) from each peptide separately and from all of them together were submitted to the Mascot search engine. With the specification of Trypsin as enzyme and oxidation of methionine as variable modification, only peptide a could be identified in the TrichoEST database. Allowing oxidation of tryptophan and/or histidine as further variable modification and abandoning enzyme specification led to a significant protein hit and sequence matches for the peptides a, a₁, a₂, and b. Peptide b is not carrying an oxidation in contrast to a₁ and a₂, but its N-terminus is not cleaved specifically. This is due to the fact that the database entry of the protein includes a signal peptide and peptide b forms the N-terminus of the mature protein. A second pass search of the result showed that also peptide c belonged to the protein but had a base substitution of T → A compared to the sequences found in the TrichoEST database. All resulting sequences can be found in table 3.2-1. The identified proteins were the expressed sequence tags (ESTs) L12T11P105R09908 (DDBJ/EMBL/GenBank accession number AJ901879) from *H. atroviridis* 11 or L14T53P106R00046 (DDBJ/EMBL/GenBank accession number AJ902344) of *T. asperellum*, both with a total Mascot score of 131. As both of these ESTs do not belong to *H. atroviridis* P1, the employed species, this result was obtained through cross-species identification due to close relation.

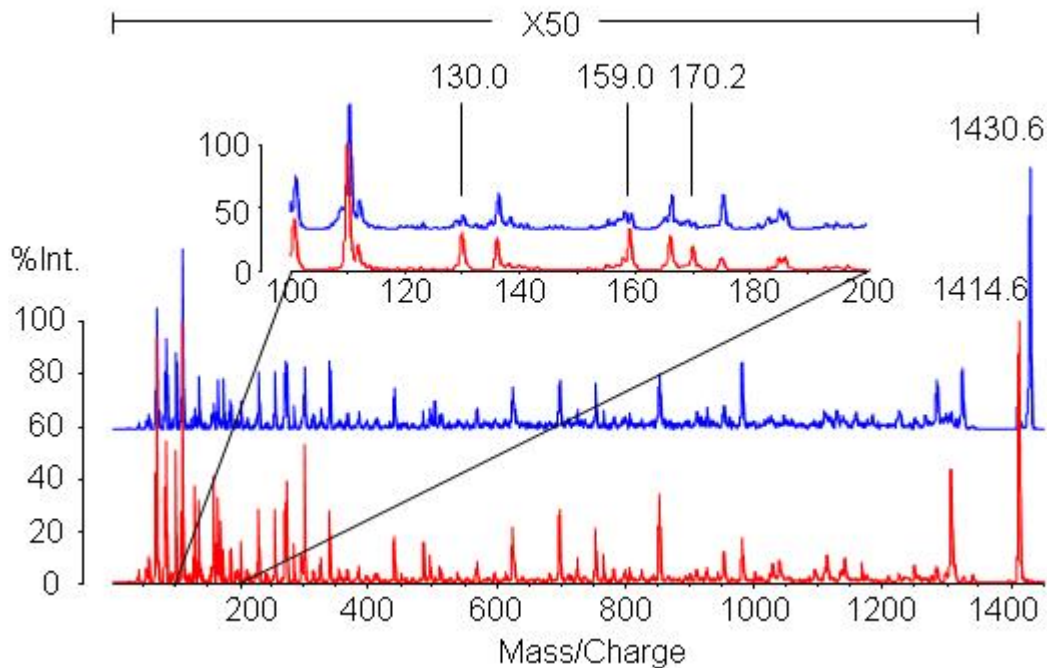


Figure 3.2-4 CID spectra of peptide a (m/z 1413.7, red) and a₁ (m/z 1430.7, blue) labelled peaks in the enlarged area: immonium ions and related ions associated with tryptophane

Further analysis of the MS/MS spectra was carried out manually. It allowed clear localization of the oxidation of peptide a₁ at the tryptophan applying data obtained through high energy CID experiments, which showed the disappearance of the immonium ion and the related ions associated with tryptophane in the oxidised peptide (figure 3.2-4). Nevertheless it was not possible to locate the second oxidation in peptide a₂. The double oxidation of tryptophane was more likely than an additional mono-oxidation of the less reactive histidine [75]. During the manual revision also peptide d was found to belong to the identified protein as the partial sequence obtained from the spectrum fitted. Mascot search was not able to match it, as the corresponding protein of *H. atroviridis* P1 (see figure 3.2-5) features two amino acid changes in comparison to the found results. Therefore a sequence coverage of 66.6 % by tryptic peptides and 54.2 % by sequencing experiments could be reached.

MQFSNLF**K**LA LFTAAVSADT VSYDTGYDDA **S**RSLTWSCS DGANGLITRY
 HWQTQGQIP**R** FPYIGGVQAV AGWNSPSCGT **CW**KLTYS**G**KT IYVLAVDHSA
 AGFNIGLDAM NALTNGNAVQ YGR**V**DATASQ VAVSN**C**GL

Figure 3.2-5 Sequence of Epl1 – blue: signal peptide, red: identified sequence, bold: cleavage sites; the C-terminal peptide was identified by PMF

Further bioinformatic analysis as described in Seidl et al [74] showed that this protein was a novel member of the cerato-platanin family (IPR010829). Fungal proteins belonging to this group have a low molecular weight, contain 4 conserved cysteine residues and have been shown to be involved in phytotoxicity and the triggering of plant defence responses [76, 77]. The *H. atroviridis* protein was consequently named Epl1 (eliciting plant response-like protein 1). Transcript analysis was used to identify which conditions led to its expression and under all tested conditions a transcript, albeit with varying abundance, could be detected. These findings were in good agreement with results from the analysis of the extracellular proteomes of *H. atroviridis* during growth on different host cell walls (see chapter 3.3).

3.2.2 Spot G3

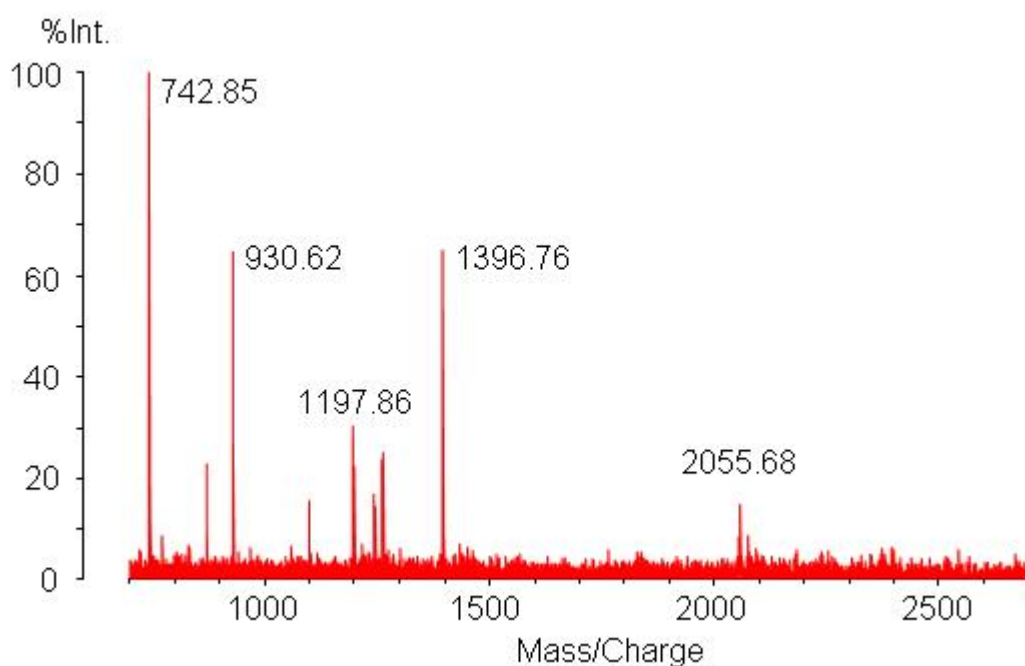


Figure 3.2-6 Peptide mass fingerprint of G3

Table 3.2-2 PMF – monoisotopic m/z values of G3 peptides

Peptide	Mass	Sequence identified by MS/MS
a	742.85	[I L][I L]E[I L]VR
	870.79	
	930.62	
	1099.89	
b	1197.86	sequence tags: chapter 3.2.2.2
c	1245.95	sequence tags: chapter 3.2.2.3
	1261.85	
d	1396.76	sequence tags: chapter 3.2.2.4
e	2055.68	sequence tags: chapter 3.2.2.5

Spot G3 is a protein of approximately 62 kDa with a pI of 5.0 to 5.2. Its peptide mass fingerprint is shown in figure 3.2-6 and table 3.2-2. Submitted to the Mascot search engine it gave no significant protein hit in any of the searched databases, so MS/MS experiments were carried out. The resulting fragment mass lists for peptides a to e can be found in appendix-table 2 and appendix-table 3.

Searches with these data gave no significant protein result either, so the MS/MS spectra of the peptides were regarded closer manually. Only in case of peptide a a full sequence could be deduced from the MS/MS fragment ions, but due to the small size of this peptide this did not allow the identification of the protein.

3.2.2.1 Peptide a – m/z 742.85

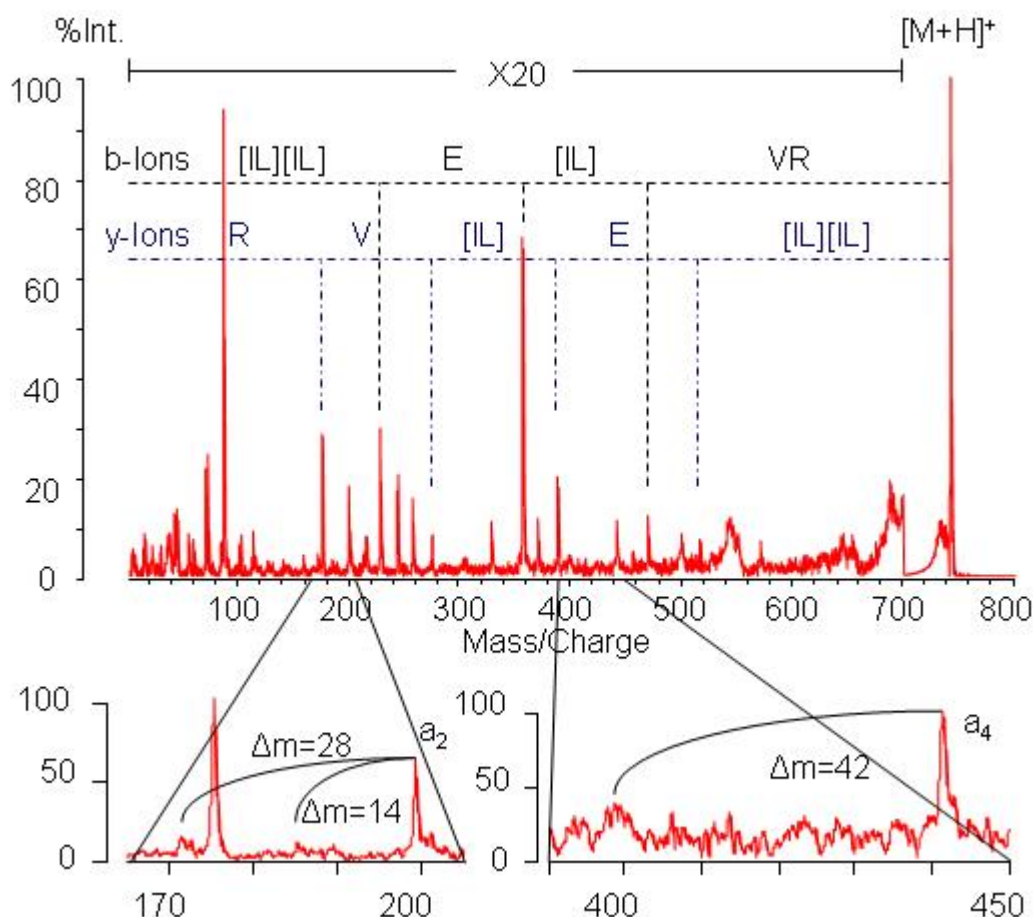


Figure 3.2-7 CID spectrum of peptide a (m/z 742.85), insets to show the differentiation between leucine and isoleucine

Mascot search of CID fragment masses of peptide a did not lead to significant peptide or protein identification. Therefore the spectra were subjected to manual interpretation. It was possible to completely elucidate a potential sequence of the peptide as shown in figure 3.2-7 drawing the following conclusions:

Due to the strong peaks of the corresponding immonium ion, the presence of leucine or isoleucine in the sequence was deduced. Further a high probability of a terminal arginine was signaled by the immonium region and the peak at 175.4, the y_1 ion. Beginning with

this the other y-ions were annotated. The preliminary sequence was then tested if it could explain the other peaks apparent in the spectrum. All of them belonged to the peptide, among them not only a- and b-, but also internal fragment ions.

These conclusions led to the sequence shown in figure 3.2-7, but the ambiguity between leucines and isoleucines still had to be resolved. Differentiation between these two amino acids is only possible if significant d or w ions are present. In case of peptide a the corresponding peaks have a very low intensity, so they were not considered during the first identification step, but could be located in a second screening. They indicate the sequence [I/L]IELVR. Nevertheless also the ambiguous sequence was used for database search.

3.2.2.2 Peptide b – m/z 1197.86

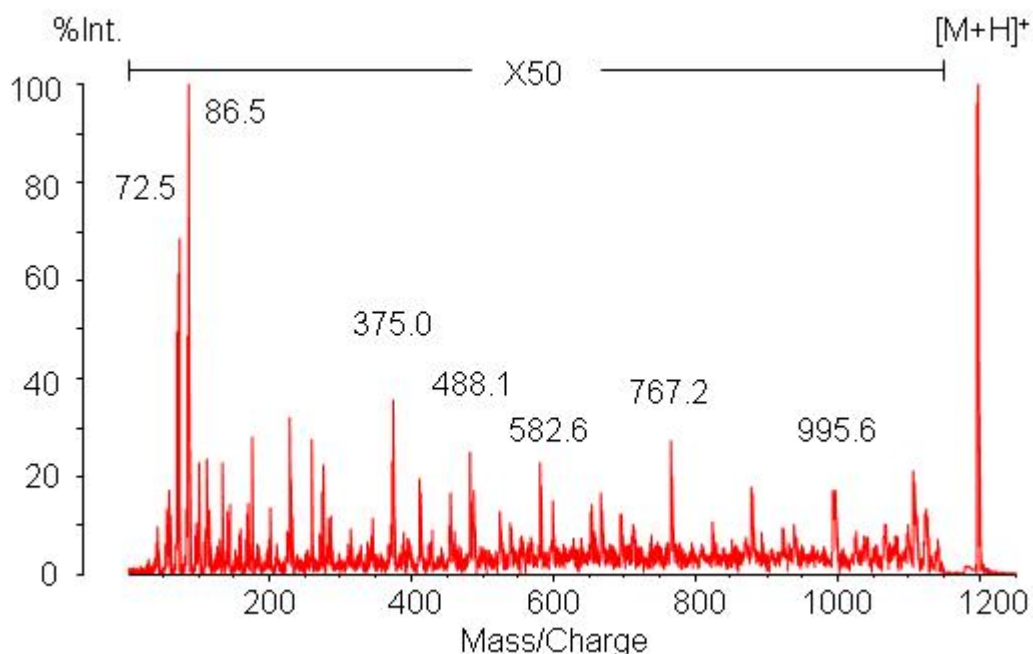


Figure 3.2-8 CID spectrum of peptide b (m/z 1187.86)

Mascot searches of the mass lists obtained by MS/MS experiments of peptide b did not lead to sequence identification. Manual interpretation could not be performed successfully either. However, the following conclusions can be drawn:

- The C-terminal amino acid of the peptide is probably an arginine, as the enzyme used for digestion was trypsin and the corresponding immonium and related ions are present as well as a peak at m/z 175.6, which would be the y_1 -Ion in this case. On the other hand a C-terminal terminal alanine would fit with the strong peak at

m/z 1108.5 and is therefore common among the Mascot search results as long as no enzyme is specified.

- From the immonium ions it can be predicted that either leucine or isoleucine is probably present in the peptide, whereas it is not likely to contain histidine, phenylalanine or tyrosine.
- Analysis of the Mascot search results based on the high energy CID MS/MS experiments led to the following potential sequence tags: 412.3-[I|L][Q|K][I|L][I|L]D-995.6; 412.3-AT[I|L]AG-825.8; 412.3-[I|L]G[I|L]A-767.2; 454.8-A[Q|K][I|L]ND[I|L]A-1198.7 (y-ions); 525.8-[Q|K][I|L][I|L]D[I|L]A-1198.7 (y-ions); 525.8-[Q|K]N[I|L]DR-1198.7 (a-ions); 525.8-[Q|K][I|L][I|L]N-995.58; 525.8-G[I|L]A[I|L]-880.5; 582.6-SV[I|L]DR-1198.7 (a-ions); 654.5-[I|L][I|L]DR-1198.7 (a-ions); 668.1-V[I|L]DR-1198.7 (a-ions); 695.8-A[I|L]DR-1198.7 (a-ions); 767.2-[I|L]D[I|L]A-1198.7 (y-ions)
- Manual analysis of the PSD-spectrum yielded the following additional sequence tags: 0.0-R[I|L]GNVV-658.1 (y-ions); 411.8-[I|L]GNN[I|L]-922.8

3.2.2.3 Peptide c – m/z 1245.95

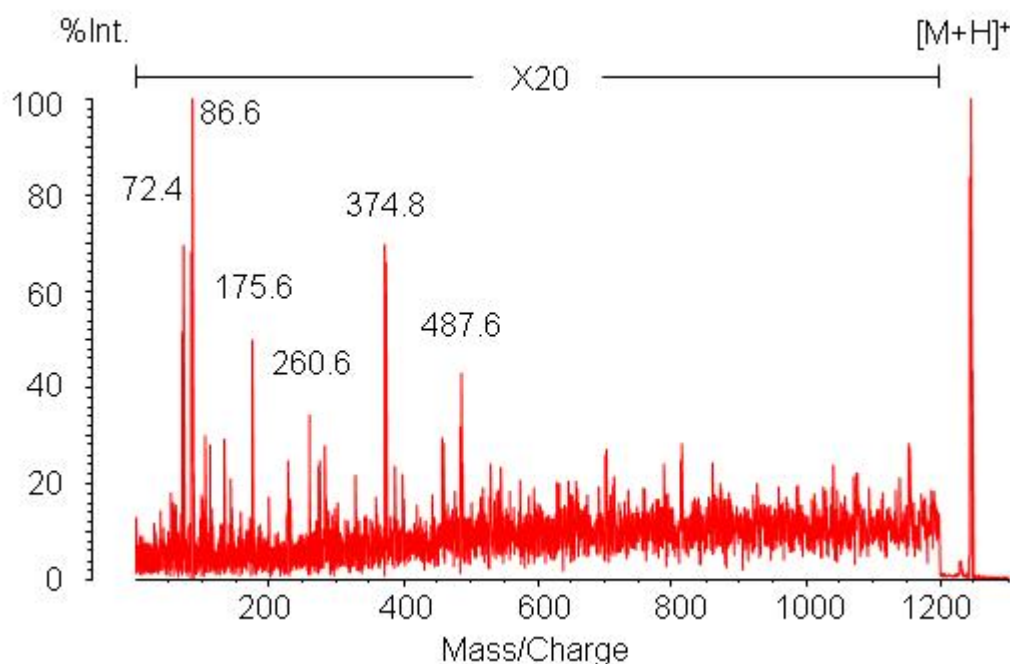


Figure 3.2-9 CID spectrum of peptide c (m/z 1245.95)

In case of peptide c, Mascot search of MS/MS mass information did not lead to significant peptide or protein information. Neither could the full sequence be assigned by manual interpretation. Only the following characterizing statements can be made:

- The MS/MS spectrum of the mass region up to approximately m/z 500 resembles greatly the spectrum of peptide b, whereas there are only low intensity peaks at different m/z values in the region of higher masses. Therefore a modification leading to the mass difference of 48 Da might be located in the middle of one of the peptides. According to Unimod, potential sources are different amino acid exchanges, a triple oxidation or the side chain loss of oxidized methionine.
- As the immonium ion region is nearly identical and the spectrum contains a peak at 1156.0 – at a difference of approximately 48 to the corresponding peak in the MS/MS spectrum of peptide b – the conclusions drawn in paragraph one and two of peptide b – C-terminal arginine or alanine, the presence of leucine or isoleucine and absence of histidine, phenylalanine or tyrosine – are equally valid for peptide c.

- Analysis of the Mascot search results led to the following potential sequence tags: 0.0-VC-camN[I|L]-487.6 (b-ions); 0.0-NVPAS-487.6 (y-ions); 0.0-N[I|L][QK]M-459.6 (a-ions); 0.0-RGRA-459.6 (y-ions); 0.0-[I|L]GSV-329.0 (a-ions)
- Manual analysis of the spectrum showed one further possibility for a sequence tag: 0.0-RVV[I|L]-487.6

3.2.2.4 Peptide d – m/z 1396.76

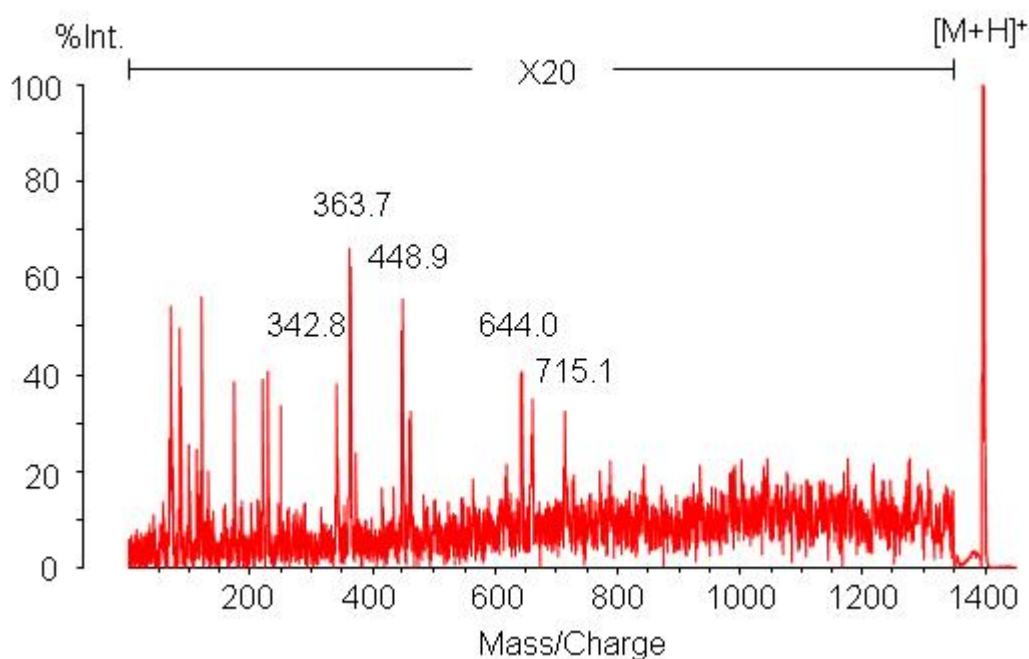


Figure 3.2-10 CID spectrum of peptide d (m/z 1396.76)

Peptide d could not be identified by Mascot MS/MS search. Manual interpretation of the spectrum was difficult because the peak intensity decreases strongly for masses higher than m/z 700. Nevertheless it was possible to draw the following conclusions:

- The related peaks in the immonium region and at m/z 175.4 suggest a C-terminal arginine. Further the immonium ions indicate the presence of glutamine and phenylalanine, whereas the sequence is not likely to contain histidine, isoleucine, leucine, tyrosine and tryptophan.
- Analysis of the Mascot search results yielded the following sequence tags: 0.0-[F|Mox]ENV-462.6 (a-ions); 0.0-[I|L]Y[N|D]V-462.6 (a-ions)

- Manual calculation of peak differences led to further sequence tags as long as the low intensity peak at 564.0 was included into the mass list: 249.6-NVTP[Q|K]-788.5; 448.9-DP[Q|K]-788.5

3.2.2.5 Peptide e – m/z 2055.68

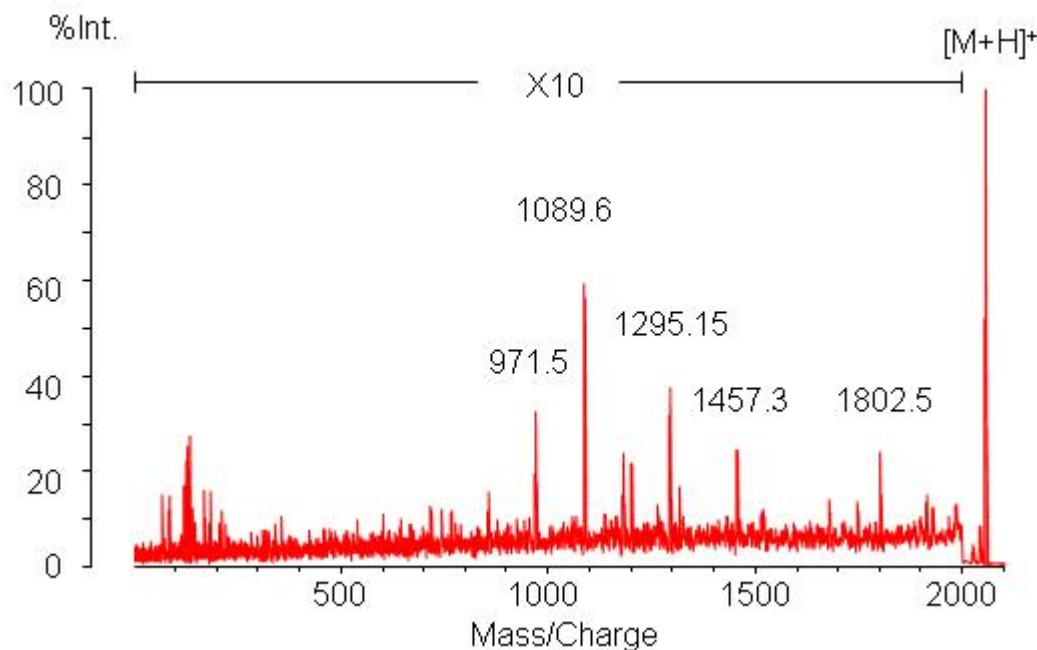


Figure 3.2-11 CID spectrum of peptide e (m/z 2055.68)

Mascot MS/MS search did not lead to significant results when applied to mass lists gained from CID or PSD experiments with peptide e. Manual interpretation did not allow the assignment of a sequence either. Nevertheless the following characteristics of the peptide could be deduced from the spectra:

- The immonium ion region hints the presence of isoleucine or leucine, lysine, phenylalanine and tryptophane, whereas the peptide is not likely to contain histidine.
- Analysis of the Mascot search results based on the CID MS/MS experiments led to the following sequence tags: 0.0-[I|L]P[Q|K]NDGAP-766.0 (a-ions); 0.0-PVNG[F|Mox][I|L]-601.7 (a-ions); 0.0-PAAAAA-426.0 (a-ions); 0.0-PAAV-311.3 (a-ions); 340.3-[F|Mox][I|L]PC-857.2; 340.3-PS[I|L]G-743.9; 340.3-S[I|L]G[F|Mox]-743.9; 487.5-[I|L]D[I|L]P-926.3; 926.3-Y[I|L][I|L]D-1430.6

- Manual peak analysis of a PSD MS/MS spectrum resulted in further sequence tags:
602.6-Y[I|L]C-camDY-1316.9; 743.7-NNP[I|L][I|L]YGYA-1748.84

3.2.2.6 Analysis of the sequence tags

All manually acquired sequence tags including the identified sequence of peptide a were submitted together to error tolerant Mascot search. In none of the databases any protein was found to contain more than one of the tags, not even the combination of peptide a and a sequence tag. Therefore the identification of spot 3 was not possible.

3.3 Identification of protein spots from 2D gels with cell walls of potential hosts as carbon source

The extracellular proteomes of *H. atroviridis*, when grown on cell walls of different potential hosts as carbon source, were analyzed in parallel experiments (2D gels see p 47 – 49). Spots sharing the same number were supposed to possibly contain the same protein. In some cases – for example in case of spots R8, B8 and P8 – this assumption was confirmed by the fact that the PMFs contained the same m/z values and MS/MS fragmentation patterns of corresponding peptides were identical. Then the spots were treated in a single identification procedure. Also if the identity of protein spots sharing the same number could neither be confirmed nor denied by MS/MS data, they were treated in one chapter. If already the PMF signalled discrepancies, the spots of the different gels were analyzed and described separately.

3.3.1 Spots R1, B1, P1, R2, B2 & P2

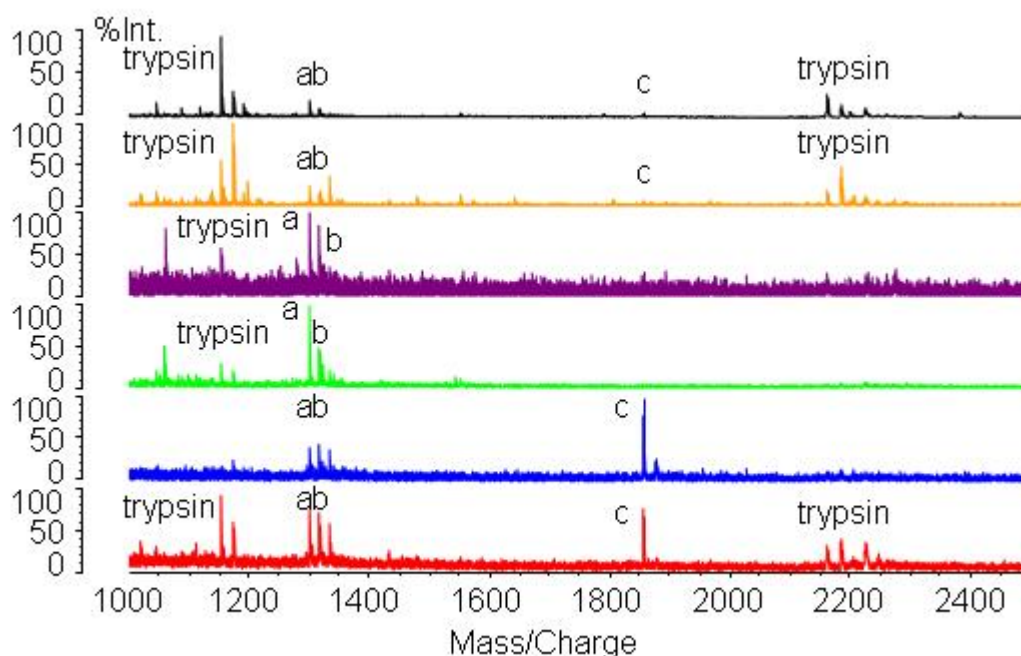


Figure 3.3-1 Peptide mass fingerprints of R1 (black), R2 (orange), B1 (violet), B2 (green), P1 (blue) and P2 (red)

All spots with number 1 have a molecular weight of approximately 26 kDa, whereas those with the number 2 have a molecular weight of approximately 24 kDa. Spots R1, B1, R2

and B2 all show a pI close to the border of the gel at 4.0, but R2 and B2 are about 0.1 pH-units more alkaline than R1 and B1. The pI difference between P2 and P1 is 0.1 as well, but they are located around a pH of 4.5.

Table 3.3-1 PMFs – monoisotopic m/z values of R1, R2, B1, B2, P1 and P2 peptides

Peptide	R1	R2	B1	B2	P1	P2
a	1301.64	1301.52	1301.92	1301.47	1301.70	1301.53
b	1317.63	1317.56	1317.91	1317.43	1317.60	1317.60
	1333.72	1333.58		1333.48	1333.66	1333.57
	1349.64	1349.54				
				1543.48		
	1791.58					
c	1855.79	1855.74			1855.85	1855.62
	2383.74					

As shown in figure 3.3-1 and table 3.3-1, a significant part of the m/z values can be found in all PMFs. This suggests that they might all contain the same protein and the mass divergence is probably caused by posttranslational modifications or modifications introduced during the various sample preparation steps. This theory is supported by the slight pI variations, which are often found between differently modified proteins. However the nature of the modification cannot be determined as there are no significant PMF differences between spots with number 1 and spots with number 2.

Interestingly the pI also varies depending on the carbon source, which is either due to the presence of distinct protein modifications under different conditions or an artificial discrepancy acquired during preparation.

The theory of the six spots being the same protein is enforced by the fact that part of the differences observed in the mass spectra between the different gel spots can be explained by the impossibility to obtain spectra of the same quality for all six protein samples. For example peptide c (m/z 1855.8) is also present in the PMFs of spots B1 and B2, but its intensity is too low, so no monoisotopic m/z value could be obtained for this peak.

The six peptide mass fingerprints were submitted to Mascot search, but no significant protein hits were obtained in any of the searched databases. Therefore the peptides a, b and c were subjected to MS/MS experiments. Not all of them yielded good quality fragment patterns and the peak intensity was very low in some cases. Nevertheless the visible

fragment ions in the corresponding MS/MS spectra were identical as shown for peptide a in figure 3.3-2, so the conclusion that the six spots might be the same protein is corroborated.

The mass lists resulting from these experiments (see appendix-table 4) were submitted to Mascot search for each of the six spots separately. None of them gave a significant peptide or protein hit, so manual interpretation was applied. Therefore always the spectrum with the highest quality for each peptide was picked independent from the fact that they were originally obtained from different spots. Nevertheless no complete sequence could be found for any of the three peptides. Potential sequence tags and other characteristics are specified in the following chapters.

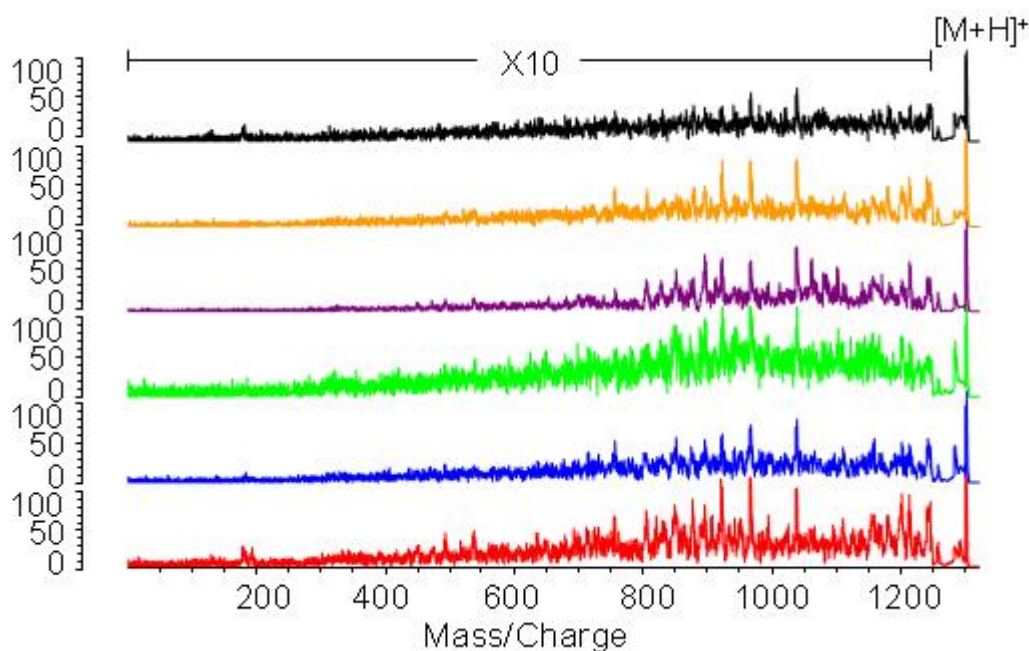


Figure 3.3-2 PSD spectra of peptide a (m/z 1301.6) of R1 (black), R2 (orange), B1 (violet), B2 (green), P1 (blue) and P2 (red)

3.3.1.1 Peptide a – m/z 1301.6

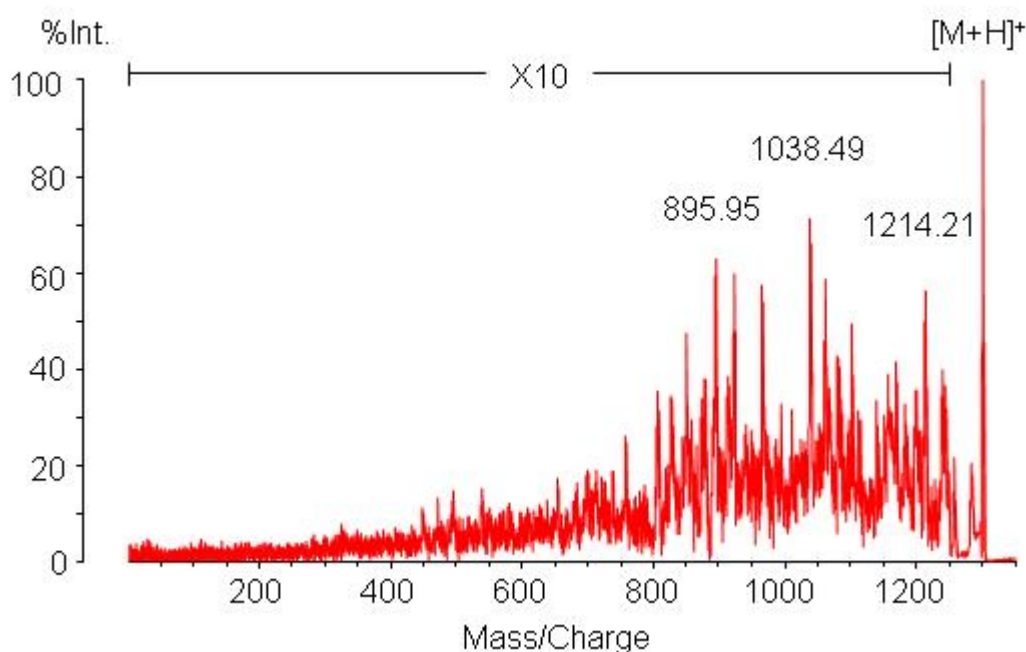


Figure 3.3-3 PSD spectrum of peptide a (m/z 1301.6) of spot B1

Mascot searches of the MS/MS mass lists did not lead to sequence identification, so manual interpretation was carried out, using the spectrum resulting from peptide a of spot B1. It led to the following conclusions:

- The terminal amino acids could be predicted by calculating characteristic mass differences to the precursor ion. As only PSD spectra were recorded, no immonium ions were available to assist this finding. Trypsin as applied enzyme cuts C-terminal of arginine or lysine, so the C-terminal amino acid has a high probability to be one of these two and the peak at 1156.59 hints the lysine. The first amino acid might be a serine because of the peak at 1214.21 or a threonine because of the peak at 1200.56.
- A comparison with a spectrum of peptide b showed some peaks with a difference of 16 Da, the same as between the precursors. So all sequence information based on this kind of peaks can be applied for both spectra. The mass difference suggests the presence of an oxidation in peptide b.
- Analysis of the Mascot search results led to the following potential sequence tags: 678.28-A[I|L]Y[Q|K][F|Mox]-1302.38 (y-ions); 700.17-[Q|K][I|L]AAVT-

- 1302.38 (b-ions); 700.17-[Q|K][I|L]ARN-1302.38 (b-ions); 700.17-GA[I|L]A-1012.27; 713.45-DSPAMS-1302.38 (y-ions); 750.48-GGYKK-1302.38 (b-ions); 757.92-A[I|L]AAEA-1302.38 (b-ions); 804.30-[I|L]YAYT-1302.38 (y-ions); 808.70-AS[F|Mox]TS-1302.38 (y-ions); 834.64-M[F|Mox]-ST-1302.38 (y-ions); 850.40-TSD[F|Mox]-1302.38 (y-ions); 852.38-NAYT-1302.38 (y-ions); 866.38-TA[Q|K]A-1302.38 (y-ions); 895.49-A[F|Mox]ST-1302.38 (y-ions)
- Additionally, manual interpretation led to the following sequence tags: 851.58-DAY-1200.56; 895.95-AAY-1200.56

3.3.1.2 Peptide b – m/z 1317.6

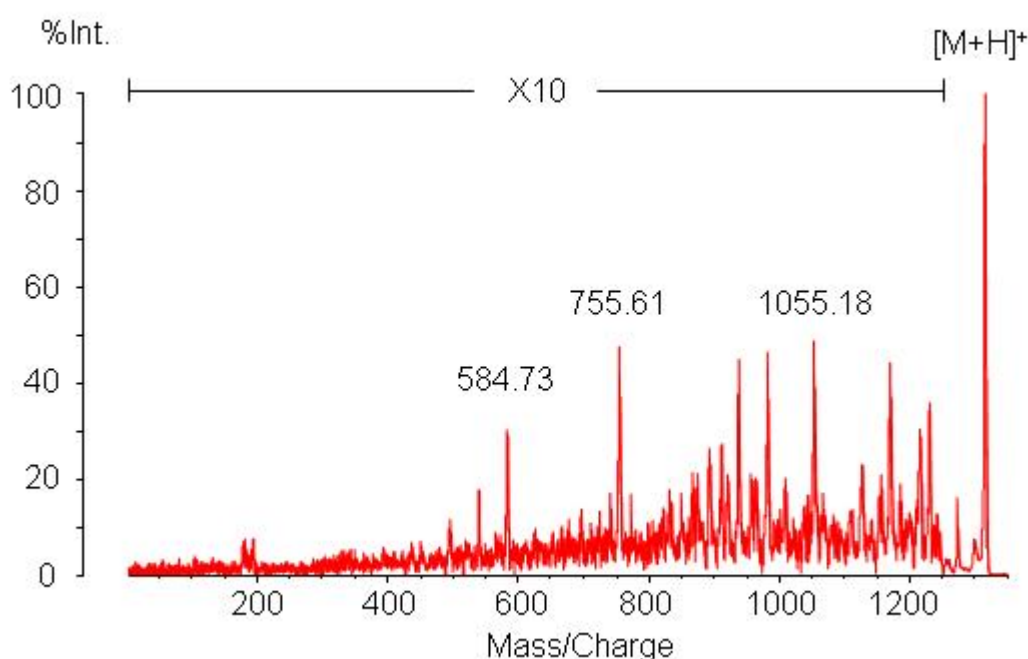


Figure 3.3-4 PSD spectrum of peptide b (m/z 1317.6) of spot P2

The mass lists obtained by MS/MS experiments of peptide b were submitted to Mascot search, but did not give any significant result. The subsequent manual interpretation allowed characterizing the peptide as follows:

- The difference of 16 Da to peptide a suggests the presence of an oxidation, but the typical mass loss of CH_3SOH which normally occurs for oxidised methionine is not observed.

- Drawing the conclusions analogue to peptide a, the C-terminal amino acid is probably a lysine because of the peak at 1171.31, the N-terminal one either a serine because of the peak at 1230.96 or a threonine because of the peak at 1217.23.
- The two strong peaks in the low mass area, 179.68 and 193.17, cannot be explained as a combination of two amino acids to form a b_2 -ion or an internal fragment. In case one of them is a y-ion there would be no arginine or lysine involved, but the C-terminal amino acids of the peptide would be either a cysteine (carbamidomethylated y_1 ion at 179.68) or two serines (y_2 ion at 193.17).
- Analysis of the Mascot search results led to the following potential sequence tags: 539.68-ITLAKSY-1318.29 (y-ions); 584.93-[I|L]M[Q|K][Q|K]W-1318.39 (a-ions); 584.93-NM[Q|K][Q|K]-1086.97; 585.26-NG[I|L]DAGC-cam-1318.29 (a-ions); 585.26-[I|L]G[I|L]N[F|Mox]-1129.77; 699.09-GRATAGS-1318.29 (b-ions); 755.84-PSYDT-1318.29 (y-ions); 756.13-LDQTS-1318.29 (b-ions); 798.52-DATSK-1318.29 (b-ions); 807.88-RYAT-1318.39 (y-ions); 851.27-NH[Q|K]S-1318.39 (y-ions); 851.51-SDYT-1318.29 (y-ions); 867.97-DAVY-1318.29 (y-ions); 868.33-DAGTS-1318.29 (b-ions); 868.33-DTS[Q|K]-1318.29; 895.15-AH[Q|K]S-1318.29 (y-ions); 895.96-SETS-1318.29 (b-ions); 913.27-AADE-1318.29 (b-ions); 913.27-VFAS-1318.29 (y-ions); 939.54-A[F|Mox]AA-1318.29 (b-ions)
- The following sequence tags were acquired through manual interpretation: 867.97-ADY-1217.23; 867.97-DAY-1217.23

3.3.1.3 Peptide c – m/z 1855.8

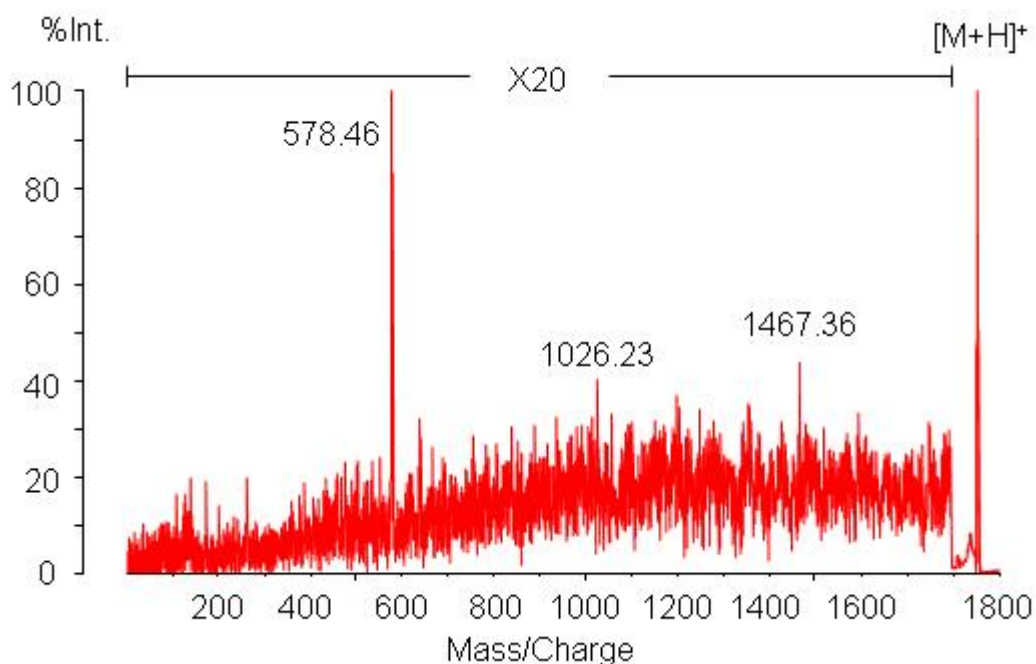


Figure 3.3-5 PSD spectrum of peptide c (m/z 1855.8) of spot P2

The fragmentation pattern of peptide c showed one peak with a very high intensity among many others, which are nearly undetectable due to a very low signal to noise ratio. Nevertheless the mass lists obtained were subjected to Mascot search. No significant peptide or protein match could be achieved, so manual interpretation was carried out. Only up to a mass range of approximately m/z 900 the peaks were clearly visible and allowed the assignment of sequence tags. On the whole, the following conclusions could be drawn:

- The C-terminal amino acid is probably an arginine, as the corresponding y_1 peak at m/z 174.58 is strongly visible, whereas no statements about the N-terminus can be made.
- The analysis of the Mascot search results led to a single potential sequence tag: 0.0-RS[Q|K]SG-534.70 (y -ions). This tag was also discovered by manual interpretation together with the following other potential tags: 0.0-RSNTTD-694.07 (y -ions); 0.0-RSNTY-639.71 (y -ions); 0.0-RS[Q|K]SY-639.71 (y -ions)

3.3.1.4 Analysis of the sequence tags

All manually acquired sequence tags were submitted individually and together as error tolerant sequence tags to Mascot search. Individual searches gave either no or multiple results. The combined search in all databases gave no significant result which contained more than one of the tags, except of the two tags of peptide a and b which have the same sequence only shifted by 16 Da. These two of course can be matched to the same peptide in error tolerant sequence tag search. The identification of spots R1, B1, P1, R2, B2 and P2 was therefore not possible.

3.3.2 Spots R3, B3 & P3

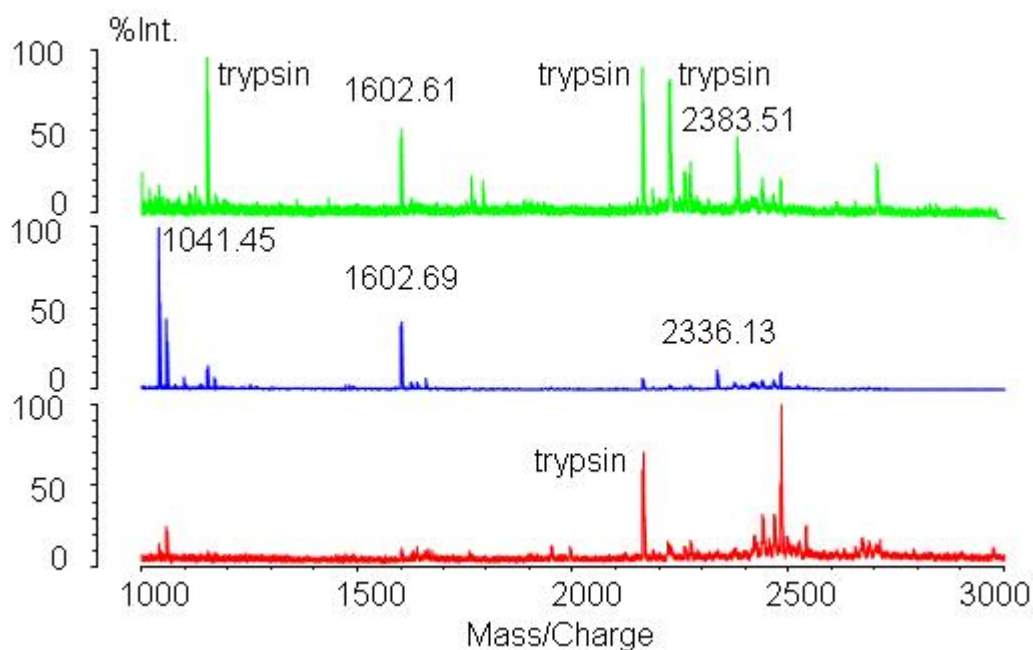


Figure 3.3-6 peptide mass fingerprints of R3 (green), B3 (blue) and P3 (red)

Table 3.3-2 PMF – monoisotopic m/z values of R3, B3 and P3 peptides

Peptide	<i>R. solani</i>	<i>B. cinerea</i>	<i>P. ultimum</i>
	1001.52		
	1041.33	1041.45	1041.75
		1253.63	
		1475.74	1476.30
		1484.75	
R3a	1602.61	1602.69	1603.31
		1659.82	
		1759.48	
	1791.49		
		1949.88	1950.59
		2336.13	
	2383.51		

The proteins analyzed as spots R3, B3 and P3 have a molecular weight of approximately 23 kDa and their pI value is 5.6 under all three different growth conditions. Nevertheless it was not possible to assure the identity of the three proteins as their peptide mass

fingerprints showed only a few common m/z values and no MS/MS spectra from corresponding peptides could be obtained. Only the *R. solani* peptide m/z 1602.61 fragmented well enough to result in an analyzable PSD spectrum. The peptide mass fingerprint data of all three proteins as shown in figure 3.3-6 and table 3.3-2 were submitted to Mascot search, but no significant protein hit was obtained. The MS/MS data of the fragmented peptide (see appendix-table 5) were submitted to Mascot search as well and resulted in a peptide sequence with significant homology as described in the following chapter.

3.3.2.1 Peptide R3a – m/z 1602.61

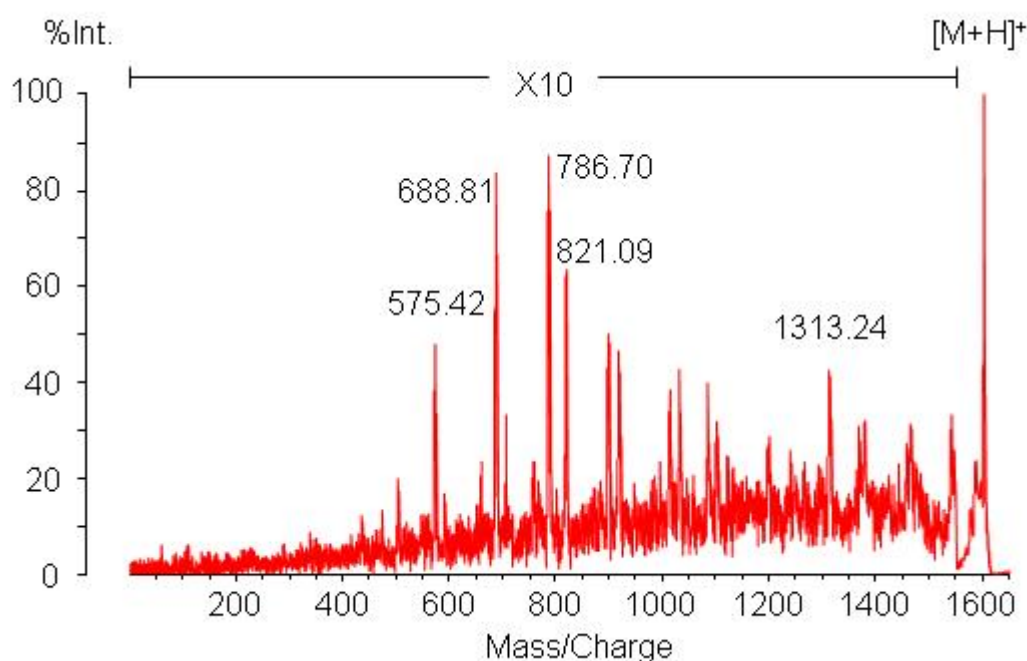


Figure 3.3-7 PSD spectrum of peptide R3a (m/z 1602.01)

Mascot search of the m/z lists obtained from peptide R3a resulted in a peptide hit with a Mascot score of 30 in the TrichoEST database. The significant homology threshold lay at a score of 26 and the identity threshold at a score of 32. The assigned peptide sequence is HSNYALVLNANDGSK and belongs to the EST L22T11P131R11775. NCBI BLASTP search [78] of the sequence showed its homology to various arabinofuranosidases from different fungi, including *H. jecorina* (*T. reesei*) (Q92455) and *H. ceramica* (P48792) with the top E values of 10^{-28} .

The *H. jecorina* (*T. reesei*) Alpha-N-arabinofuranosidase precursor is a protein with 51 kDa including a 21 amino acid signal peptide (potential). It belongs to the glycosyl

hydrolase 54 family and its function lies in the plant cell wall polysaccharide L-arabinan degradation pathway. This gene was cloned from *H. jecorina* and the protein overexpressed in *Saccharomyces cerevisiae* in 1996 by Margolles-Clark et al. [79].

The R3 protein seems to resemble this arabinofuranosidase, but the identification was not significant. No further peptide masses fit the peptides derived from the EST, but it has only 17 kDa and is therefore far shorter than the homologous proteins.

3.3.3 Spots R4 & B4 – *R. solani* & *B. cinerea* cell walls

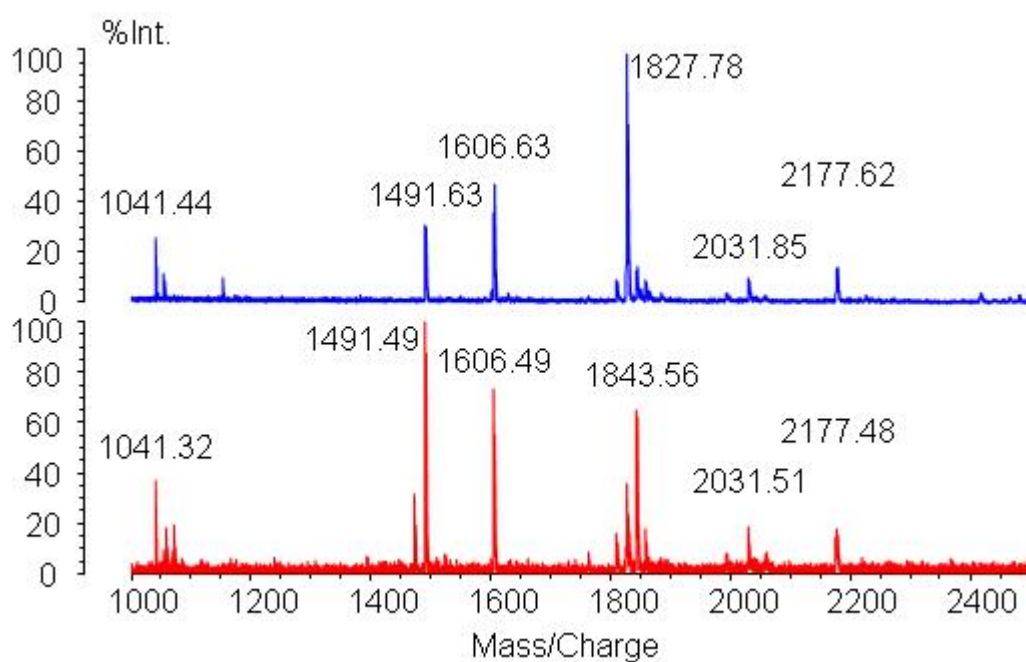


Figure 3.3-8 Peptide mass fingerprints of R4 (above) and B4 (below)

Table 3.3-3 PMF – monoisotopic m/z values of R4 and B4 peptides

Peptide	<i>R. solani</i>	<i>B. cinerea</i>
	1041.44	1041.32
a	1055.57	1055.49
	1071.52	1071.39
		1395.57
		1473.48
b	1491.63	1491.49
c	1606.63	1606.49
	1809.71	1809.48
d	1827.78	1827.52
e	1843.81	1843.56
	1859.83	1859.60
	1884.85	
	2031.85	2031.51
f	2177.62	2177.48

Spots R4 and B4 are proteins of approximately 61 kDa with a pI of 5.3 to 5.4. Their PMFs consist of nearly the same m/z values and also the MS/MS fragmentation of corresponding peptides was identical, therefore they are very probably the same protein. On the other hand spot P4 showed a very different peptide mass fingerprint and is therefore treated on its own in chapter 3.3.4.

The two peptide mass fingerprints were submitted to Mascot search, but no significant protein result was obtained. Therefore MS/MS experiments were carried out, leading to fragment mass lists for six different peptides as shown in appendix-table 6. Mascot search was carried out for each of the peptides separately and for all peptides belonging to one spot together. None of the searches allowed the identification of a peptide or protein. Manual interpretation followed, based always on the spectrum with the highest quality for each peptide, independent if the spot of origin was R4 or B4.

3.3.3.1 Peptide a – m/z 1055.5

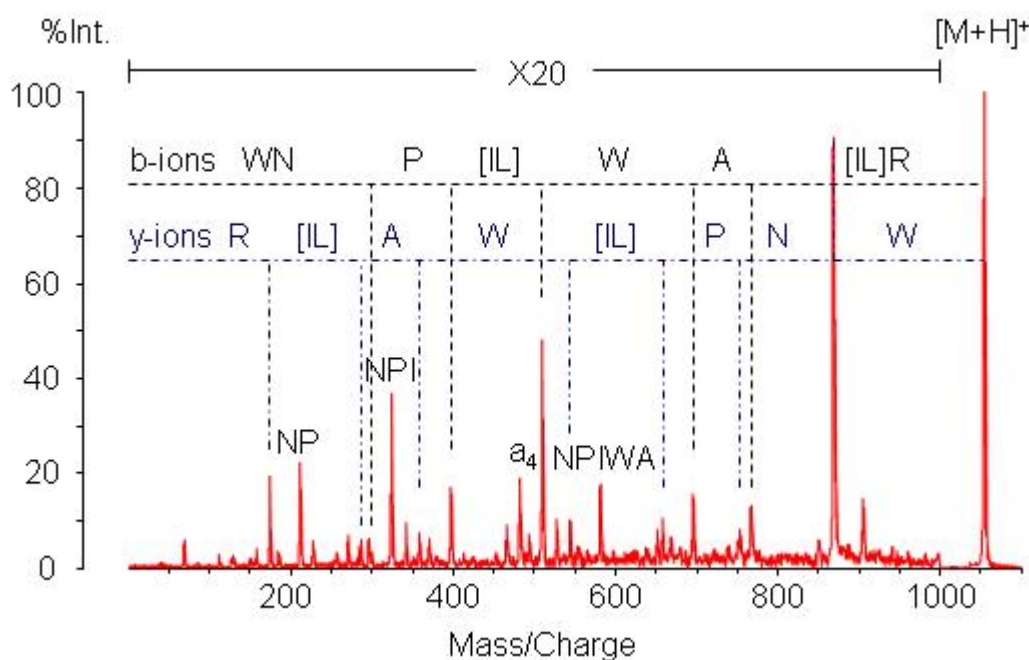


Figure 3.3-9 PSD spectrum of peptide a (m/z 1055.5)

The Mascot search of MS/MS data obtained by fragmentation of peptide a did not lead to significant peptide or protein results. Manual interpretation led to the sequence as shown in figure 3.3-9.

Even though the spectrum was generated by post source decay, some strong ions were observed in the immonium region at 70.19, 112.32 and 159.26 Da. They suggested the

presence of tryptophan and arginine, the latter being the C-terminus as the y_1 -ion at 175.35 Da confirmed. The mass difference between the strong peak at 869.84 and the precursor led to tryptophan as the N-terminus. Between the two identified peaks a y -ion sequence could be assigned, which was then confirmed by the corresponding b -ions. Finally all further strong peaks could be explained as internal fragments, a -ions and neutral losses. The ambiguity between leucine and isoleucine could not be resolved as no CID spectra of the peptide were recorded.

The database search of the peptide WNP[I|L]WA[I|L]R showed that this sequence did not belong to any protein in any of the searched databases.

3.3.3.2 Peptide b – m/z 1491.6

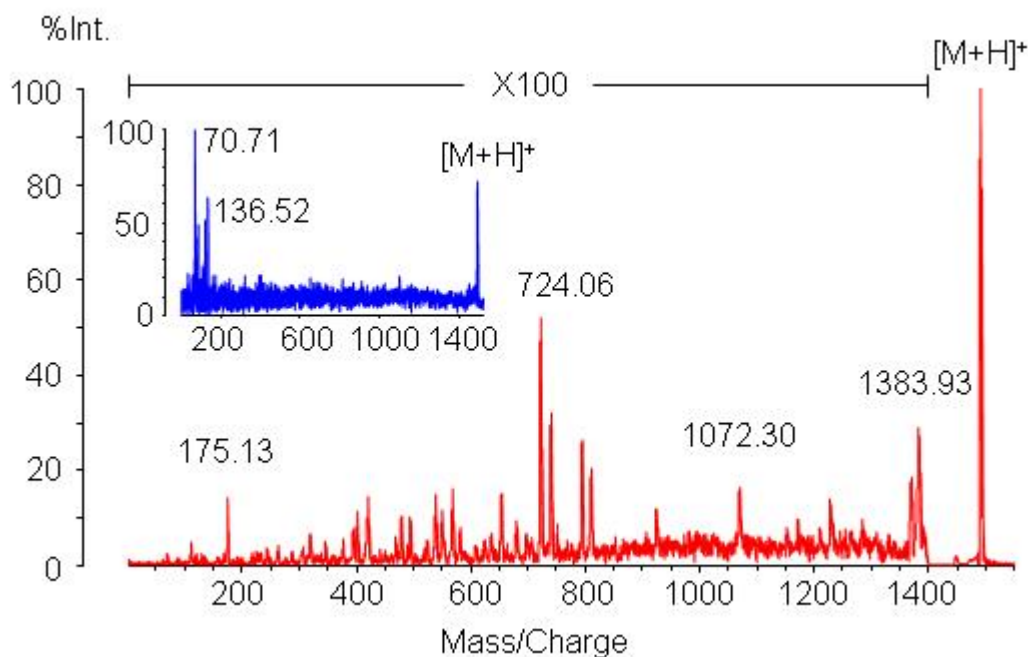


Figure 3.3-10 PSD (red) and CID (blue) spectrum of peptide b (m/z 1491.6)

Peptide b was subjected to PSD and CID experiments, but only the post source decay fragmentation resulted in high quality spectra. Therefore this mass list was submitted to Mascot search, but no protein or peptide hits were obtained. Subsequent manual interpretation allowed the following conclusions:

- The immonium region of the CID spectra showed peaks at 70, 72, 86, 87, 112, 120 and 136 Da, therefore the presence of arginine, leucine or isoleucine, phenylalanine and tyrosine in the peptide can be deduced. The peak at 175.13 Da together with

the information of the digesting enzyme results in a high probability for arginine being the C-terminal amino acid.

- Analysis of the Mascot search results led to the following potential sequence tags: 0.0-RNDH[I/L]-654.40 (y-ions); 0.0-R[I/L]DH[I/L]SAG-868.07 (y-ions); 0.0-RMD[F|Mox]RS-812.67 (y-ions); 0.0-RLD[F|Mox]-551.07 (y-ions); 175.13-MA[F|Mox]E-654.40; 289.15-NS[F|Mox]SS[I/L][F|Mox]; 306.71-D[F|Mox]RS[I/L]-996.26; 306.71-SPYS-741.26; 322.55-VMI[F|Mox][I/L][F|Mox]-1073.83; 421.75-EEDEA-996.26; 347.05-GH[I/L]SA[I/L]-925.86; 421.75EF[I/L][I/L]G-983.16; 741.26-A[I/L][F|Mox]T[I/L]C-cam-1492.4 (a-ions)
- Manual interpretation of the spectrum added two further potential sequence tags: 321.09-GY[I/L]SA[I/L][F|Mox]-1072.30 and 346.84-GH[I/L]SA[IL][F|Mox]-1072.30. Both tags gave no results in any of the databases

3.3.3.3 Peptide c – m/z 1606.6

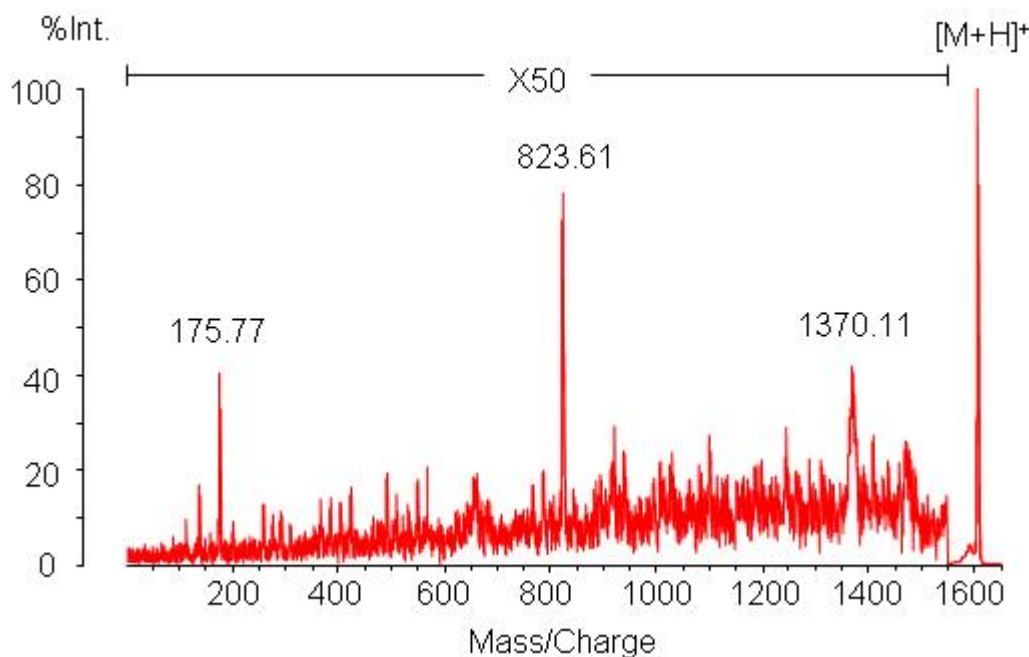


Figure 3.3-11 PSD spectrum of peptide c (m/z 1606.6)

MS/MS experiments with peptide c produced among others the spectrum presented in figure 3.3-11. The corresponding mass lists were submitted to Mascot search, but no result was obtained. Therefore manual interpretation was carried out, leading to the following statements about the peptide:

- CID experiments generated no high quality spectrum, but contained the following ions in the immonium region: 59, 70, 72, 86 and 136 Da. Therefore the presence of arginine, leucine or isoleucine and tyrosine in the sequence was very probable. The peak at 175.77 implied a C-terminal arginine and the mass difference of the peak at 1470.46 to the precursor made an N-terminal histidine plausible.
- Analysis of the Mascot search results led to the following potential sequence tags: 0.0-RD[N|D][I|L]-517.91; 201.37-GYGS-565.54
- Manual interpretation of the spectrum added further potential sequence tags: 0.0-RV[Q|K]Y-566.67 (y-ions); 0.0-R[F|Mox][Q|K][Q|K]-532.52 (y-ions); 0.0-R[F|Mox][Q|K]S-491.14 (y-ions); 201.63-GY[Q|K]-549.65. All four tags gave either no or more than a single result in all databases.

3.3.3.4 Peptide d – m/z 1827.7

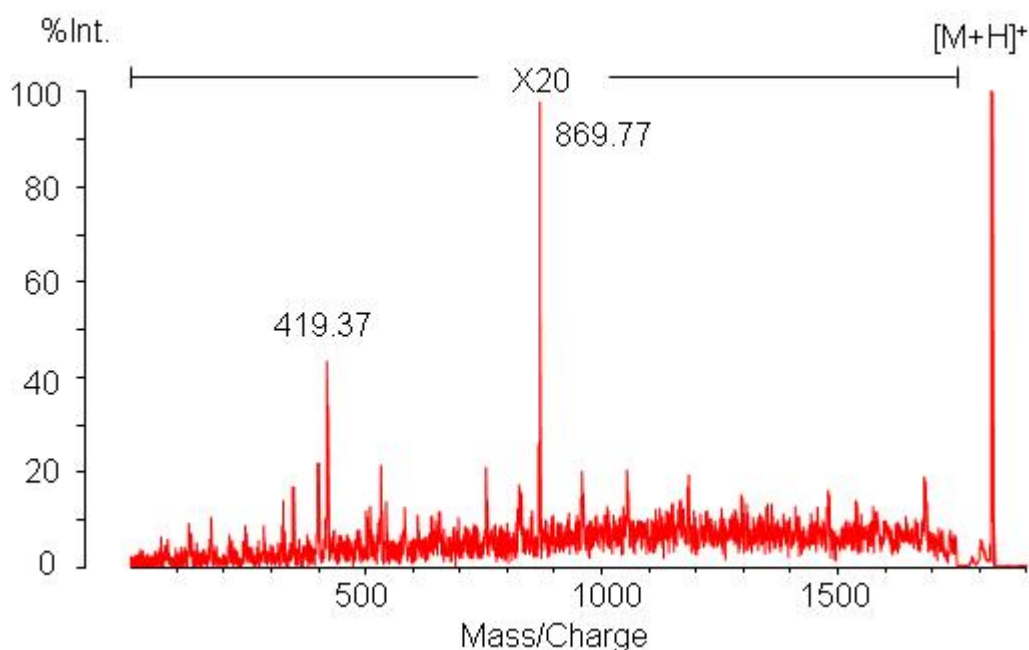


Figure 3.3-12 PSD spectrum of peptide d (m/z 1827.7)

Mascot search with mass lists obtained during MS/MS experiments of peptide d gave no significant peptide or protein hits. Therefore manual spectrum interpretation was carried out, leading to the following conclusions:

- Only the peak at 174.30 gave a hint that the peptide might contain a C-terminal arginine, otherwise no significant immonium ions or mass differences to the precursor could be used to elucidate the terminal amino acids.
- Analysis of the Mascot search results led to the following potential sequence tags: 869.77-W[Q|K][I|L][I|L]-1411.44; 959.00-P[Q|K]N[I|L]A-1483.23; 959.00-PE[I|L][I|L]-1411.44; 1055.65-[Q|K]N[I|L]AT-1583.63
- Manual assignment of potential sequences led to the following tags: 247.04-TY[F|Mox]PN-869.77; 247.04-TA[I|L]-532.39; 325.18-W[F|Mox]PN-869.77; 545.39-[I|L]PN-869.77; 1055.65-[Q|K][I|L]WT-1583.63. Whereas three of the tags gave multiple or no protein results, the first and the third mentioned tag led to single proteins. Nevertheless these were also no relevant results as the assigned proteins were both too small, one of them had additionally a completely wrong pI and the other one could be ruled out due to the description of its function as it was a dosage-dependent suppressor of the cell cycle mutation.

3.3.3.5 Peptide e – 1843.7

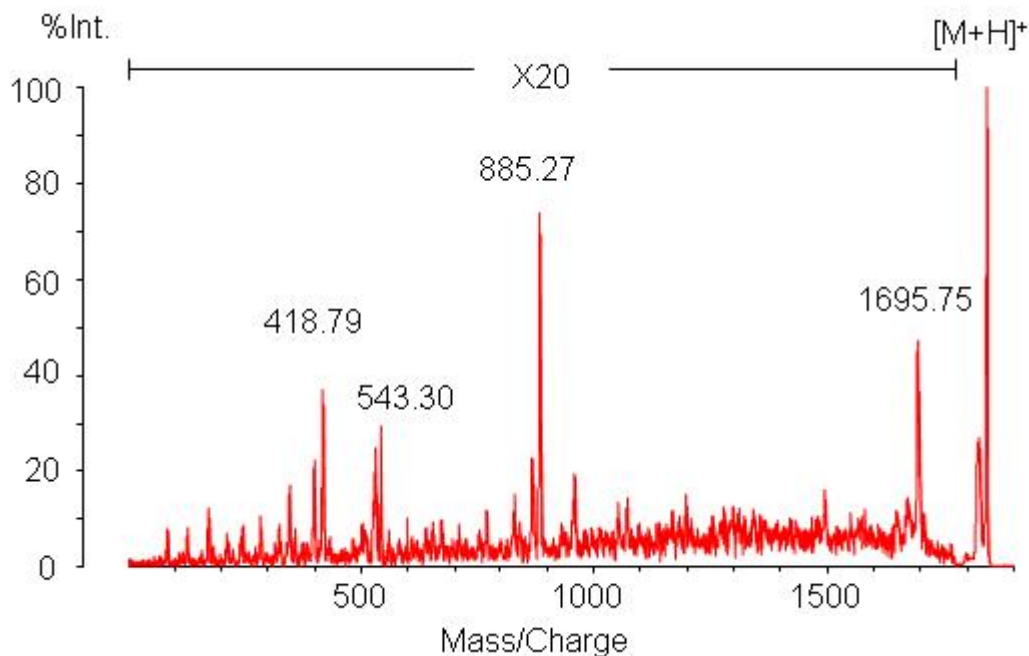


Figure 3.3-13 PSD spectrum of peptide e (m/z 1843.7)

Peptide e has a mass difference of 16 Da to peptide d, so peptide d is probably modified by an oxidation. As the MS/MS information did not lead to sequence identification through Mascot search, manual interpretation was applied to get the following results:

- Peptide e contains a C-terminal arginine due to the same reasons as peptide d. The strong peak at 1695.75 is suspected to indicate an N-terminal oxidized methionine, but this could not explain the absence of 16 Da shifts compared to peptide d in the mass region below m/z 400, locating the oxidation in the middle of the peptide. The mass difference between the parent ion at the peak at 1695.75 could also be explained by phenylalanine (m/z Mox = m/z F).
- Analysis of the Mascot search results led to the following potential sequence tags: 674.40-P[I|L]WV-1169.50; 771.16-[I|L]W[Q|K]V-1298.16; 771.16-NW[Q|K]P-1298.16; 830.75-E[I|L][Q|K]P-1298.16
- Manual interpretation gave one identical and one tag exhibiting a difference of +16 Da for the C-terminal mass value compared to peptide d: 247.48-TA[I|L]-531.58 and 561.56-[I|L]PN-885.27

3.3.3.6 Peptide f – m/z 2177.6

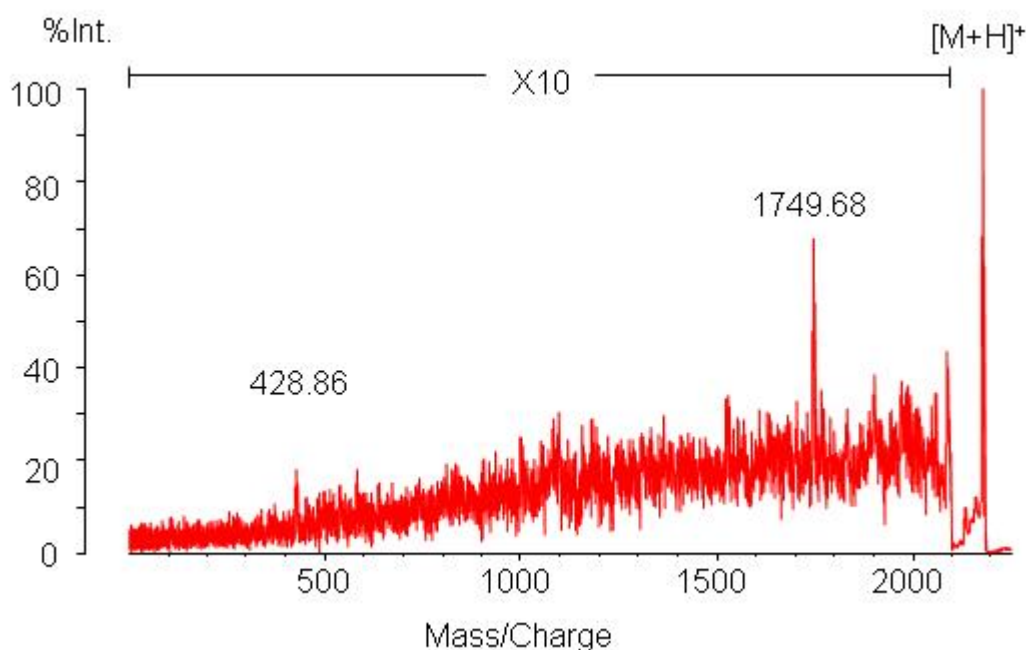


Figure 3.3-14 PSD spectrum of peptide f (m/z 2177.6)

Peptide f was subjected to PSD experiments equally to the other peptides. The resulting spectra had a low signal to noise ratio, therefore only very few peaks could be detected. The m/z lists were submitted to Mascot search, but no significant peptide sequence was identified, neither if the information was used on its own nor if it was combined with data of the other peptide fragmentation experiments. As only very few peaks are visible in the spectrum, manual interpretation was not possible either.

3.3.3.7 Combination of the sequence tags

All manually acquired sequence tags were not only submitted separately but also together in one Mascot sequence query to all databases. None of the results included more than one of the sequence tags, except for the corresponding tags of peptides d and e. As the limitations of the single results were already discussed in the respective chapters, the logical conclusion is the impossibility of protein identification with these data.

3.3.4 Spot P4 – Exoglucanase

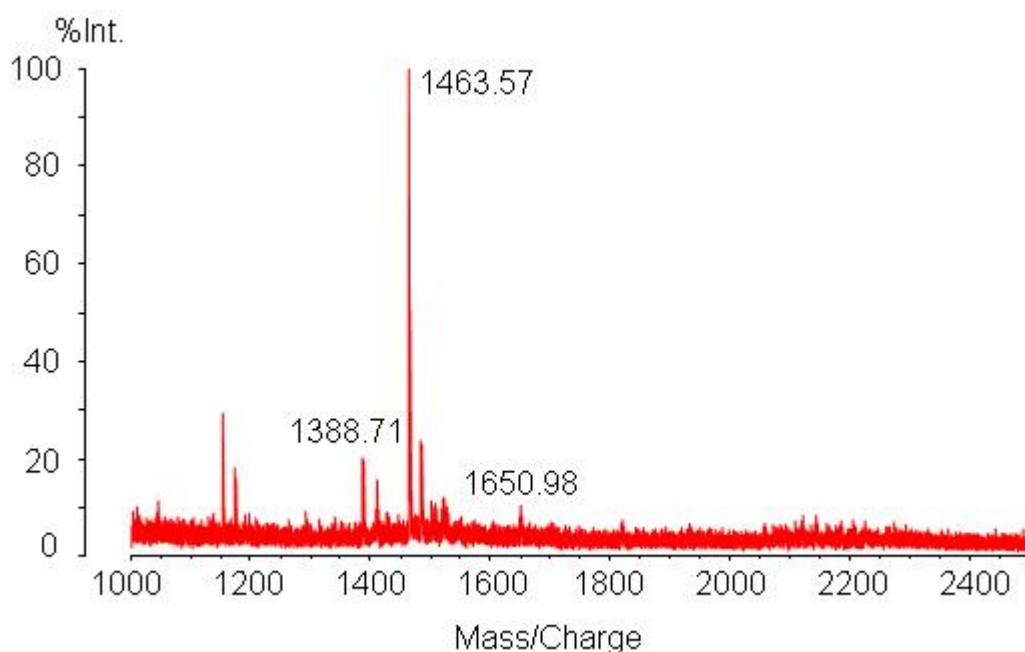


Figure 3.3-15 Peptide mass fingerprint of P4

Table 3.3-4 PMF –monoisotopic m/z values of P4 peptides

Peptide	<i>P.ultimum</i>	Sequence
	1292.76	
a	1388.71	
b	1463.57	YGTGYCDSQCPR
	1507.55	
	1650.98	
	2121.95	

Spot P4 is a protein of approximately 69 kDa with a pI of 4.9. Its peptide mass fingerprint as shown in figure 3.3-15 and table 3.3-4 was submitted to Mascot search, but no significant protein hit was obtained. Therefore MS/MS experiments on two strong peptides were carried out leading to the PSD spectra shown in figures 3.3-16. Both fragment mass lists (see appendix-table 7) were submitted to Mascot search.

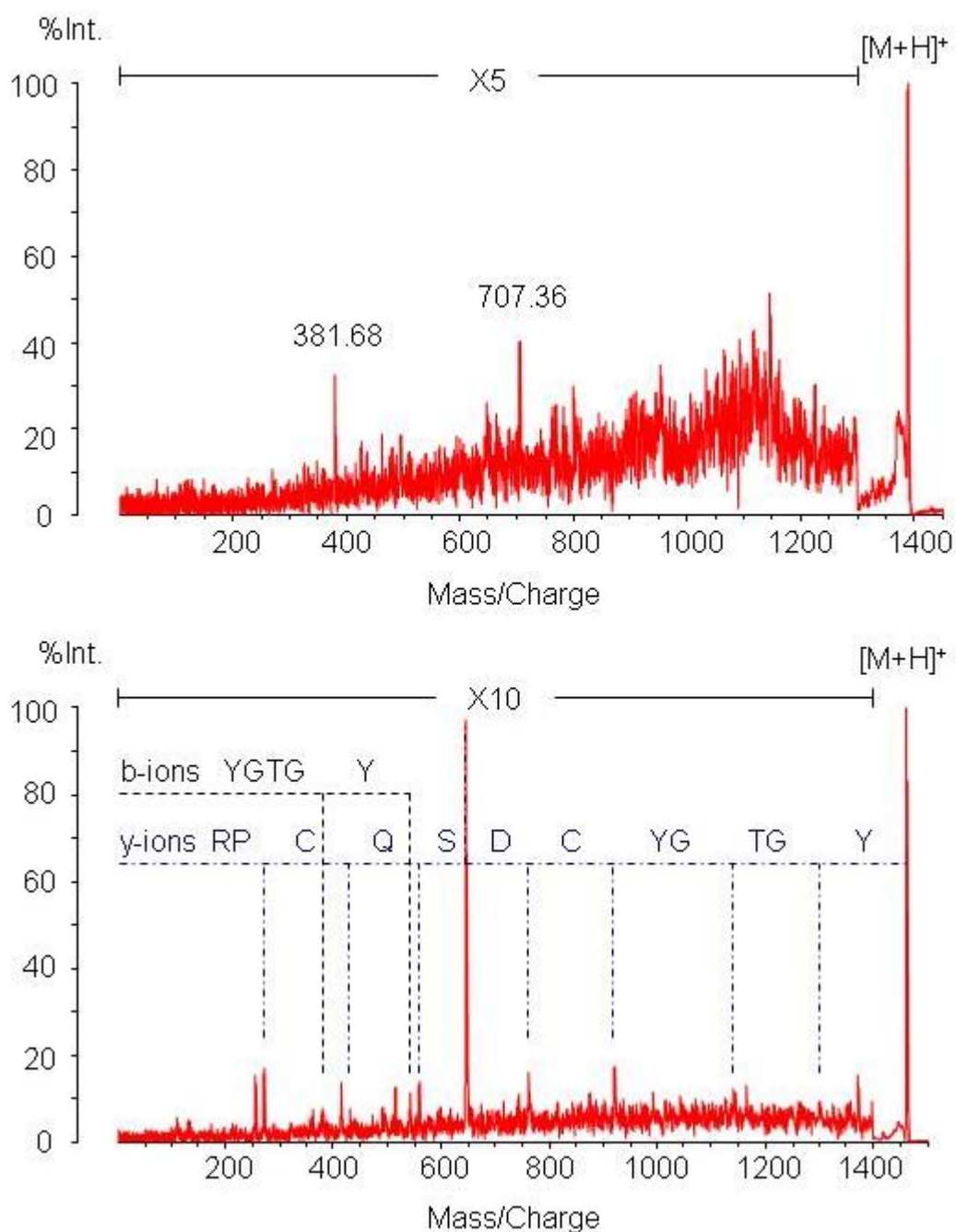


Figure 3.3-16 PSD spectra of peptides a (m/z 1388.71, above) and b (m/z 1463.57, below)

As the spectrum obtained from peptide b had a high quality, the identification of its sequence YGTGYCDSQCPR was possible. This peptide was found through Mascot search in the SWISS-PROT database (search parameters: enzyme trypsin, variable modifications oxidation methionine) with a significant Mascot ions score of 31 (Threshold value: 30). It is part of eight different fungal exoglucanase (EC 3.2.1.91) precursors whose SWISS-PROT primary accession numbers are the following:

- Q92400 of *Agaricus bisporus* (Common mushroom)
- O59843 of *Aspergillus aculeatus*
- Q06886 of *Penicillium janthinellum* (*Penicillium vitale*)
- P13860 of *Phanerochaete chrysosporium* (*Sporotrichum pruinosum*)
- Q9P8P3 of *H. lixii* (*T. harzianum*)
- P62695 of *T. koningii*
- P62694 of *H. jecorina* (*T. reesei*)
- P19355 of *T. viride*

It was tried to assign peptide a to one of these proteins as well, but as its PSD spectrum had a low signal to noise ratio, only few fragment masses could be found. They were not sufficient for generating sequence tags and the peptide could not be identified by error tolerant Mascot search of the results either. Only the knowledge of the sequence of the exoglucanase secreted by *H. atroviridis* P1, which was not in the database, would allow the identification. Therefore the sequence coverage is only approximately 2 % for the found proteins.

The found exoglucanase seems to belong to the glycoside hydrolase family 7 according to the domain identified with NCBI BLASTP. This family shows activities as endoglucanases and reducing end acting cellobiohydrolases. The corresponding *H. jecorina* (*T. reesei*) exoglucanase contains also a fungal-type carbohydrate-binding domain. The beginning of the protein is marked by a signal peptide and it carries potential N-glycosylation sites. The unprocessed protein has a molecular weight of 54 kDa and its theoretical pI is 4.65. Its glycosylation is described among others by Pakula et al. [80].

The possibility of glycosylation is in good accordance with the fact that the spot P4 belongs to a train of spots on the gel, which might be caused by that sort of posttranslational modifications. Also glycosylation increases the molecular weight, which would be an explanation why the protein is found at 69 kDa .

In general the presence of an exoglucanase among the secreted peptides of *H. atroviridis* was expected, as hydrolytic enzymes play a major role in biocontrol. Therefore the hit seems very probable regardless of the low sequence coverage.

3.3.5 Spots R5 & B5

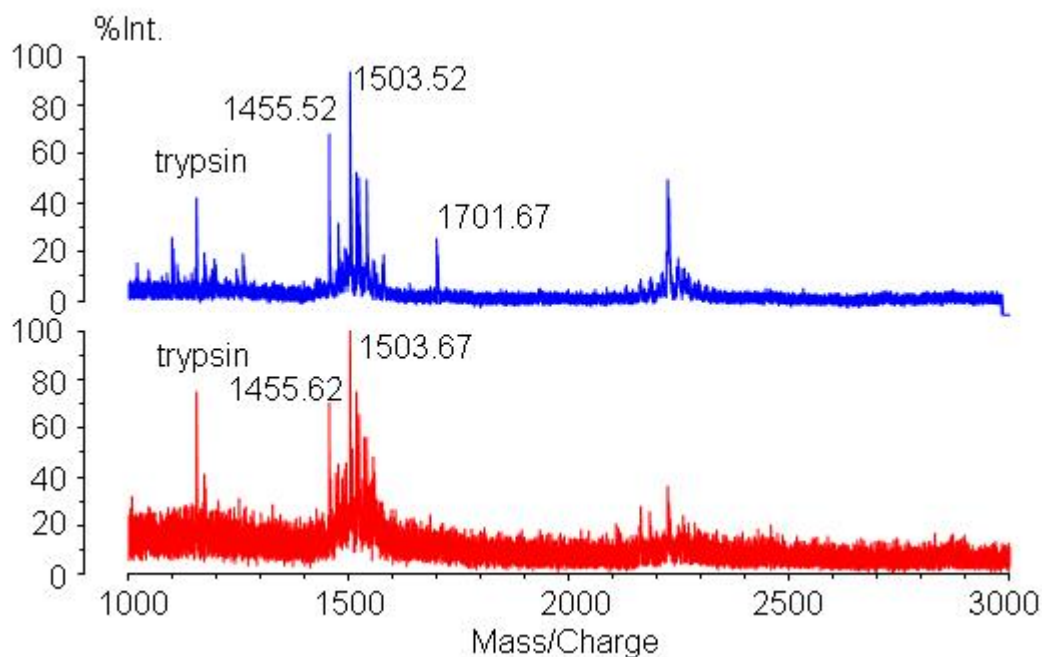


Figure 3.3-17 Peptide mass fingerprints of R5 (above) and B5 (below)

Table 3.3-5 PMF – monoisotopic m/z values of R5 and B5 peptides

Peptide	<i>R. solani</i>	<i>B. cinerea</i>
	1099.62	
	1245.55	
	1261.48	
a	1455.52	1455.62
	1471.55	1471.49
	1484.45	
	1491.48	1491.35
b	1502.55	1502.58
b	1503.52	1503.67
		1507.58
	1519.55	1519.56
	1535.4	1535.61
	1701.67	

Spot R5 and B5 are proteins of approximately 50 kDa with a pI of 4.4 to 4.5. Their PMFs contain a great number of identical m/z values as shown in figure 3.3-17 and table 3.3-5,

The identity of the two proteins was confirmed by the fact that MS/MS experiments of corresponding peptides led to the same fragment ion patterns. On the other hand spot P5 showed a very different peptide mass fingerprint and is therefore discussed in chapter 3.3.6.

Based on the experience with the other proteins, which could not be identified by their peptide mass fingerprints, PSD experiments (mass lists shown in appendix-table 8) were carried out even before the PMFs were studied closer. Therefore m/z 1502.6 was picked as the strongest peak without noticing that its distribution was not normal monoisotopic but probably contained two different peptides. The consequences of this error are described in chapter 3.3.5.2.

Neither the fragmentation pattern of peptide a nor the one from “peptide” b, nor the peptide mass fingerprint a resulted in significant hits during Mascot search. The PSD spectra obtained from spot R5 had a far higher quality as those obtained from spot B5 and were therefore used also for the manual search. B5 PSD spectra could just be used to confirm the identity of the two proteins but had a very low signal to noise ratio.

3.3.5.1 Peptide a – m/z 1455.5

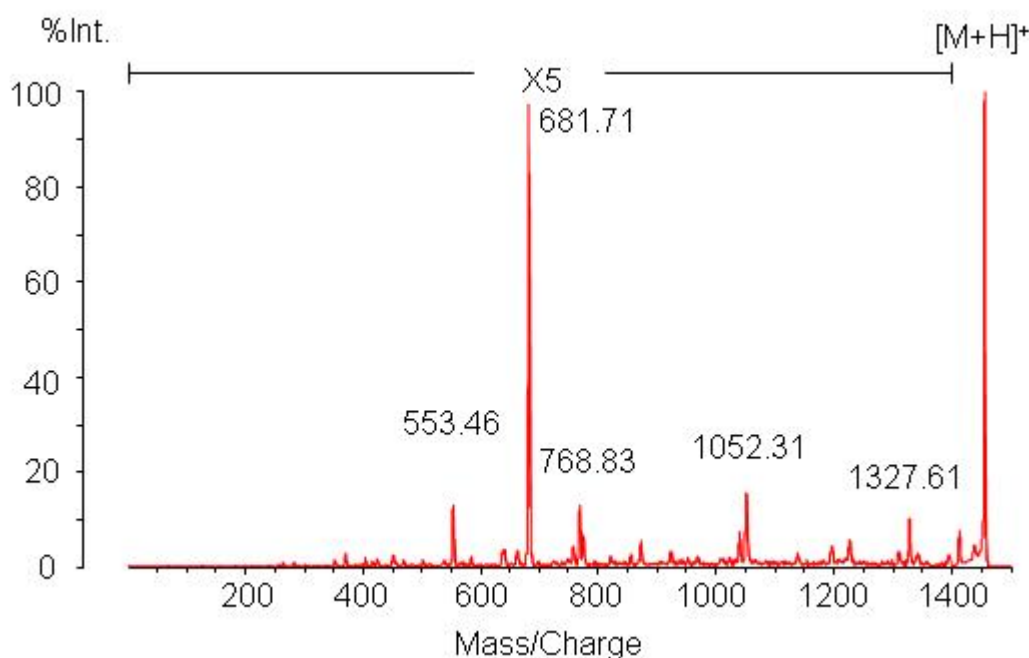


Figure 3.3-18 PSD spectrum of peptide a (m/z 1455.5)

Peptide a was subjected to manual interpretation as Mascot searches of its fragmentation pattern did not result in sequence identification. Even though no full sequence could be discovered, the following conclusions could be drawn:

- The peak at m/z 1310.50 indicates a C-terminal lysine, but it could also be explained as ammonia loss of the peak at m/z 1327.61. 1327.61 Da itself signals an N-terminal glutamic acid. Nevertheless a third peak in that region, at m/z 1342.96 might also hint an N-terminal leucine or isoleucine. As only PSD spectra were recorded, no immonium ions can assist a proper assignment and no potential y_1 or b_1 ions were found either.
- Analysis of the Mascot search results led to the following potential sequence tags: 256.29-PA[I|L]V-637.20; 284.14-SPPP[I|L]PP-969.22; 318.61-PT[Q|K][I|L]-757.86; 553.46-S[Q|K]S[I|L]AR-1196.21; 681.71-SSPAD-1139.53
- Additionally, the following manual sequence tags were discovered: 452.40-T[Q|K]SS-855.71; 775.91-PPAWTE-1456.47 (y-ions); 822.73-TESGN[Q|K]-1456.47 (b-ions). The latter two together can explain all main peaks in the upper m/z region, but the transformation into the corresponding low mass ions does not resemble the spectrum in this area. Therefore at least one of them does not reflect the correct sequence, even though no other possibility to explain the visible peaks was found. Nevertheless all three tags were submitted to Mascot search. None of them gave a unique result when searched in the databases.

3.3.5.2 “Peptide b” – m/z 1503.1

Peptide b was chosen as the strongest peak in the peptide mass fingerprint spectrum. The observation that it probably consists of two different peptides made it necessary to loosen the mass tolerance for the parent ion during Mascot search. It was set to 1.2 instead of 0.7 and the mean value of both monoisotopic m/z values was used as a parent mass. Of course it was not very likely that a significant result could be obtained from this PSD spectrum as its fragment ions belong to different sequences. In case one peptide fragmented significantly easier than the other and gave more intensive peaks, identification was still possible. Therefore it was treated as every other peptide, so as the Mascot search failed it was interpreted manually. It was not possible to obtain a full sequence for peptide b, but some conclusions could be drawn independently of its special case:

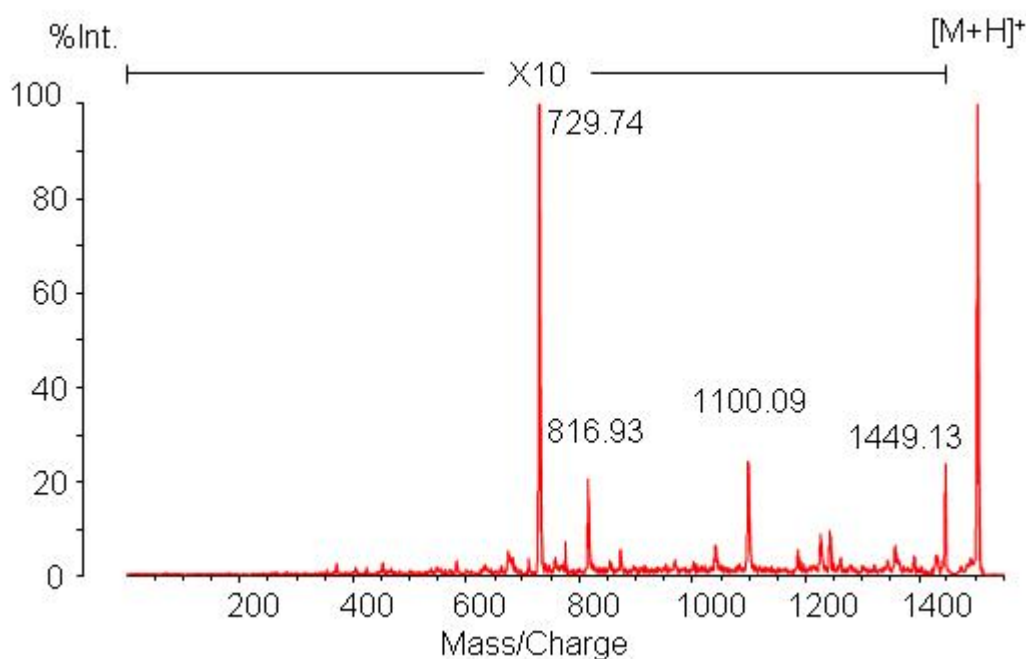


Figure 3.3-19 PSD spectrum of “peptide b” (m/z 1503.1)

- The lower mass region below m/z 300 has a very bad signal to noise ratio, therefore only the mass differences to the precursor can be used to determine the terminal amino acids. As the precursor has two potential masses, both differences had to be considered. The peak at m/z 1358.61 for example could be explained by a C-terminal lysine of the precursor with 1503.6 Da, whereas the peak at m/z 1432.00 hinted to an N-terminal alanine for the peptide with 1502.6 Da. All other peaks in this mass region had no characteristic distance to the precursors.
- Analysis of the Mascot search results led to the following potential sequence tags: 353.57-[F|Mox]V[Q|K]SG-873.14; 353.57-A[F|Mox]RS-816.93; 387.52-NEGS-776.00; 601.31-SSPPPES-1187.15; 601.31-ESGE-1003.25; 969.77-AWM[Q|K]-1504.00 (b-ions); 1100.09-SGN[Q|K]-1504.00 (b-ions); 1100.09-SGD[Q|K]-1504.00 (b-ions)
- Manual interpretation added the following sequence tags: 632.48-PSWPSGD[Q|K]-1504.00 (b-ions); 776.00-PPAW-1227.02. Mascot database search gave no unique result for any of these tags.

3.3.5.3 Combination of the sequence tags

All sequence tags were not only submitted separately but also together in one Mascot sequence query to all listed databases. None of the results included more than one of the sequence tags. As none of the tags was unique for itself, the logical conclusion of this fact is the ambiguity of these combined results as well. Protein identification was therefore not possible.

3.3.6 Spot P5

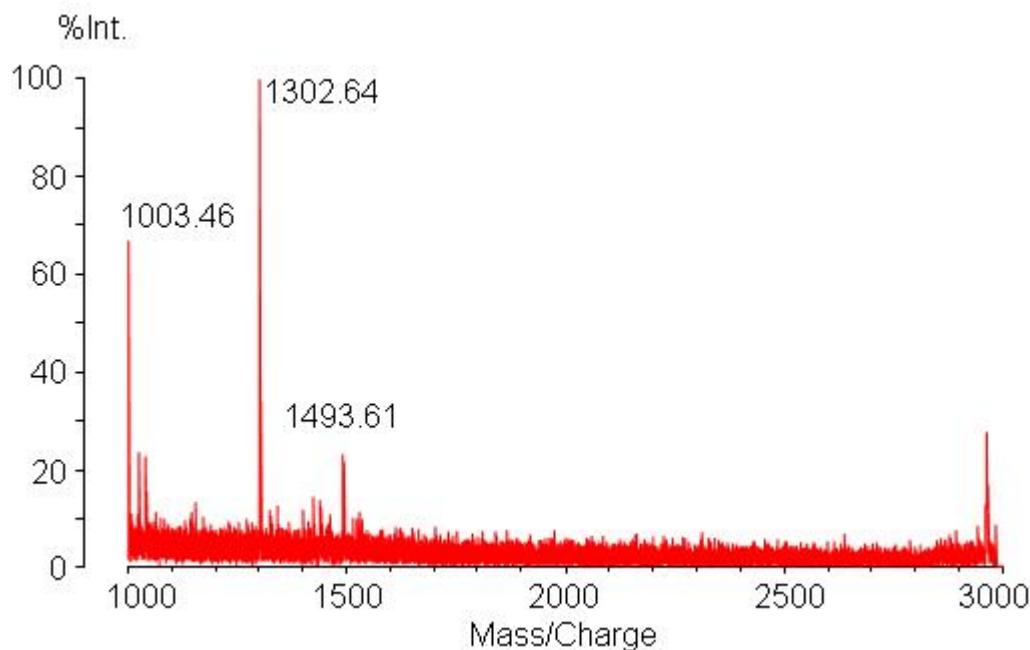


Figure 3.3-20 Peptide mass fingerprint of P5

Table 3.3-6 PMF – monoisotopic m/z values of P5 peptides

Peptide	<i>P.ultimum</i>
a	1003.46
b	1302.64
	1402.34
c	1424.47
	1440.4
	1493.61

Spot P5 is a protein of approximately 43 kDa with a pI of 4.6. Its peptide mass fingerprint is shown in figure 3.3-20 and table 3.3-6. The list of monoisotopic peptide masses was submitted to Mascot search, but no significant protein hit was obtained. Therefore MS/MS experiments were carried out on three different peptides and the resulting fragment mass lists (see appendix-table 9) were submitted to Mascot search. No peptide sequence or protein hit was obtained by this method, so manual interpretation was carried out.

3.3.6.1 Peptide a – m/z 1003.46

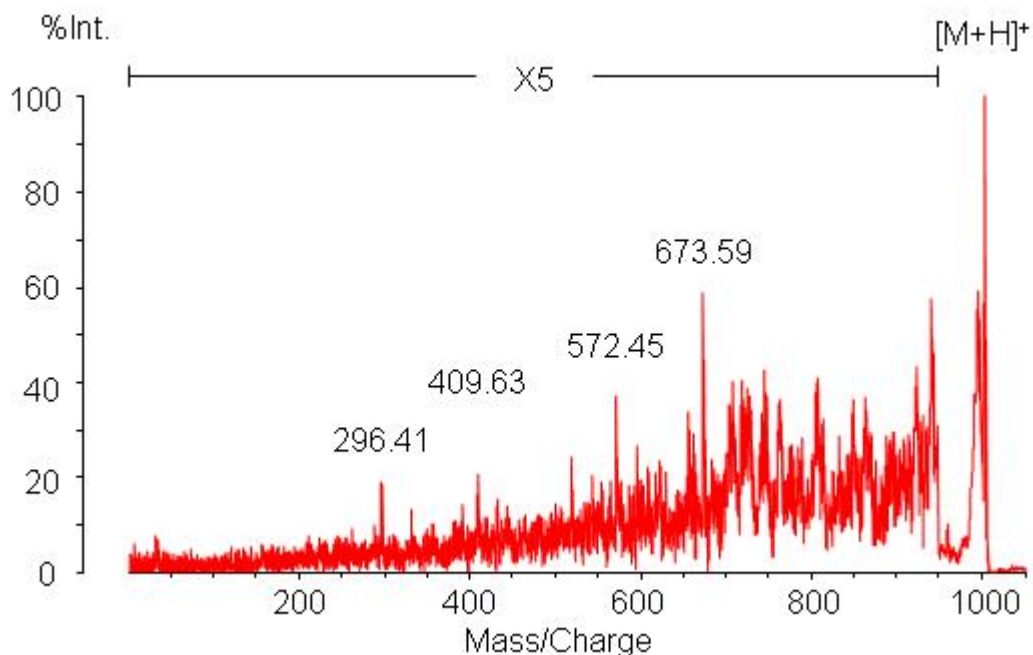


Figure 3.3-21 PSD spectrum of peptide a (m/z 1003.46)

PSD experiments carried out with peptide a only resulted in spectra with a low signal to noise ratio. Therefore it was not possible to identify the sequence neither with Mascot search nor through manual interpretation. The only results obtained were the following:

- None of the terminal amino acids are known, as in the low molecular weight area no peaks were detected and also the mass differences of high mass peaks to the precursor were not characteristic.
- No sequence tags with more than three amino acids could be obtained through analysis of the Mascot search results.
- The only two sequence tags discovered during manual interpretation were 296.41-[I|L]YTA-745.25 and 331.42-TSH-655.80. They both do not give a unique result when submitted to Mascot sequence query.

3.3.6.2 Peptide b – m/z 1302.64

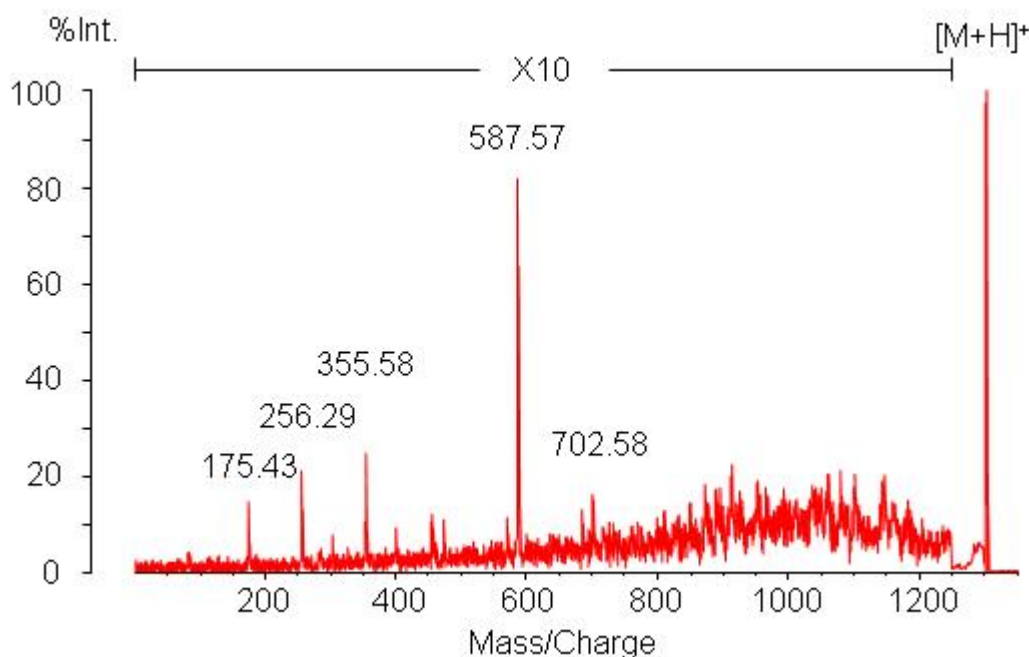


Figure 3.3-22 PSD spectrum of peptide b (m/z 1302.64)

MS/MS experiments of Peptide b resulted in the spectrum shown in figure 3.3-22. The peaks in its lower m/z range show a very good signal to noise ratio, whereas the resolution decreases rapidly over m/z 700. The obtained fragment mass list was submitted to Mascot search, but no peptide or protein could be identified. Therefore manual interpretation was applied to reach the following conclusions:

- The C-terminal amino acid is probably an arginine, as the peak at m/z 175.43 is very probable the y_1 -ion. The corresponding difference between the precursor and a high mass peak can not be found, but the peak at 1146.81 might be explained by an N-terminal arginine. This would fit with the peak at 256.29 which can be caused by a b_2 -ion consisting of arginine and valine.
- Analysis of the Mascot search results did not lead to any sequence tags longer than three amino acids.
- Manual interpretation was applied to obtain the following sequence tags: 0.0-R[Q|K]PG-457.27 (y -ions); 0.0-RVVV-454.55 (b -ions); 474.46-[I|L]D[F|Mox]-849.55. All three tags gave multiple results when submitted to Mascot sequence query.

3.3.6.3 Peptide c – m/z 1424.47

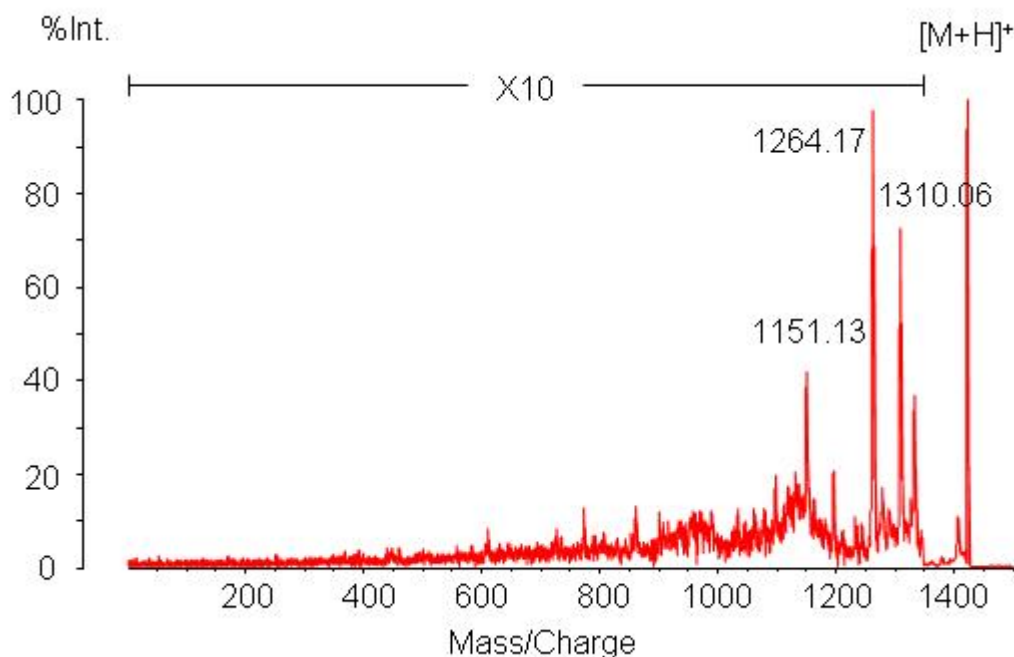


Figure 3.3-23 PSD spectrum of peptide c (m/z 1424.47)

The MS/MS spectrum of peptide c was subjected to manual interpretation as Mascot search did not lead to significant peptide or protein results. Its characteristics can be described as follows:

- The small peak at m/z 1279.49 hints a C-terminal lysine, but also a C-terminal alanine could be indicated by the peak at 1334.68. The peak at m/z 1310.06 might be caused by an N-terminal aspartic acid.
- Analysis of the Mascot search results led to the following potential sequence tags: 775.49-[Q|K]SC-camC-camN-1425.32 (y-ions); 1035.98-[N|D]NAA-1425.32 (b-ions)
- Manual interpretation added two further sequence tags: 449.79-YY[S|Q|K]A-1062.18 and 1035.98-DEK-1425.32, which did not lead to unique results when submitted to Mascot search.

3.3.6.4 *Combination of the sequence tags*

All manually obtained sequence tags were submitted as a single query to Mascot search. No protein was found to contain more than one of the tags, therefore an identification of the unknown protein was not possible

3.3.7 Spot R6 - Alginate lyase

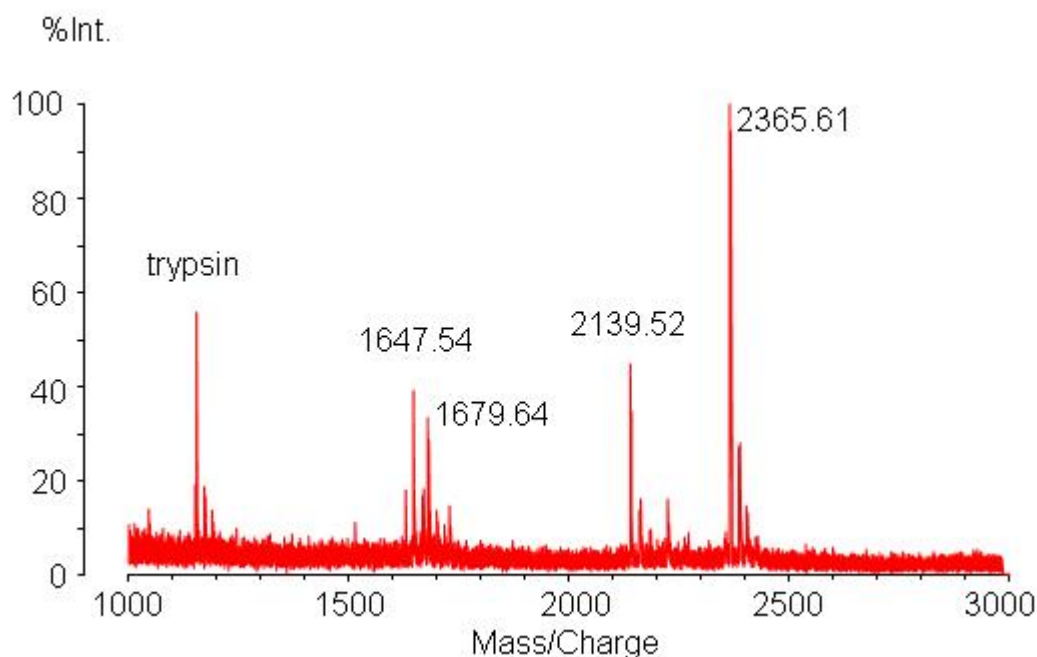


Figure 3.3-24 Peptide mass fingerprint of R6

Table 3.3-7 PMF – monoisotopic m/z values of R6 peptides

Peptide	<i>R. solani</i>	Sequence
	1516.39	
	1629.51	
a	1647.54	
	1679.64	
	1728.56	
b	2139.52	VPGAPSNTGCVTTPNSQHCR
	2365.61	

Spot R6 was picked from the extracellular proteome of *H. atroviridis* when it was grown on *R. solani* cell walls. No corresponding protein was produced in case the other cell walls were used as carbon source. Its properties are a molecular weight of approximately 24 kDa and a pI value of 6.1. The peptide mass fingerprint is shown in figure 3.3-24 and table 3.3-7. It was not possible to obtain a significant protein hit with these data, so MS/MS experiments were carried out with two peptides. The resulting fragment mass lists can be found in appendix-table 10 whereas the spectra are shown in figure 3.3-25.

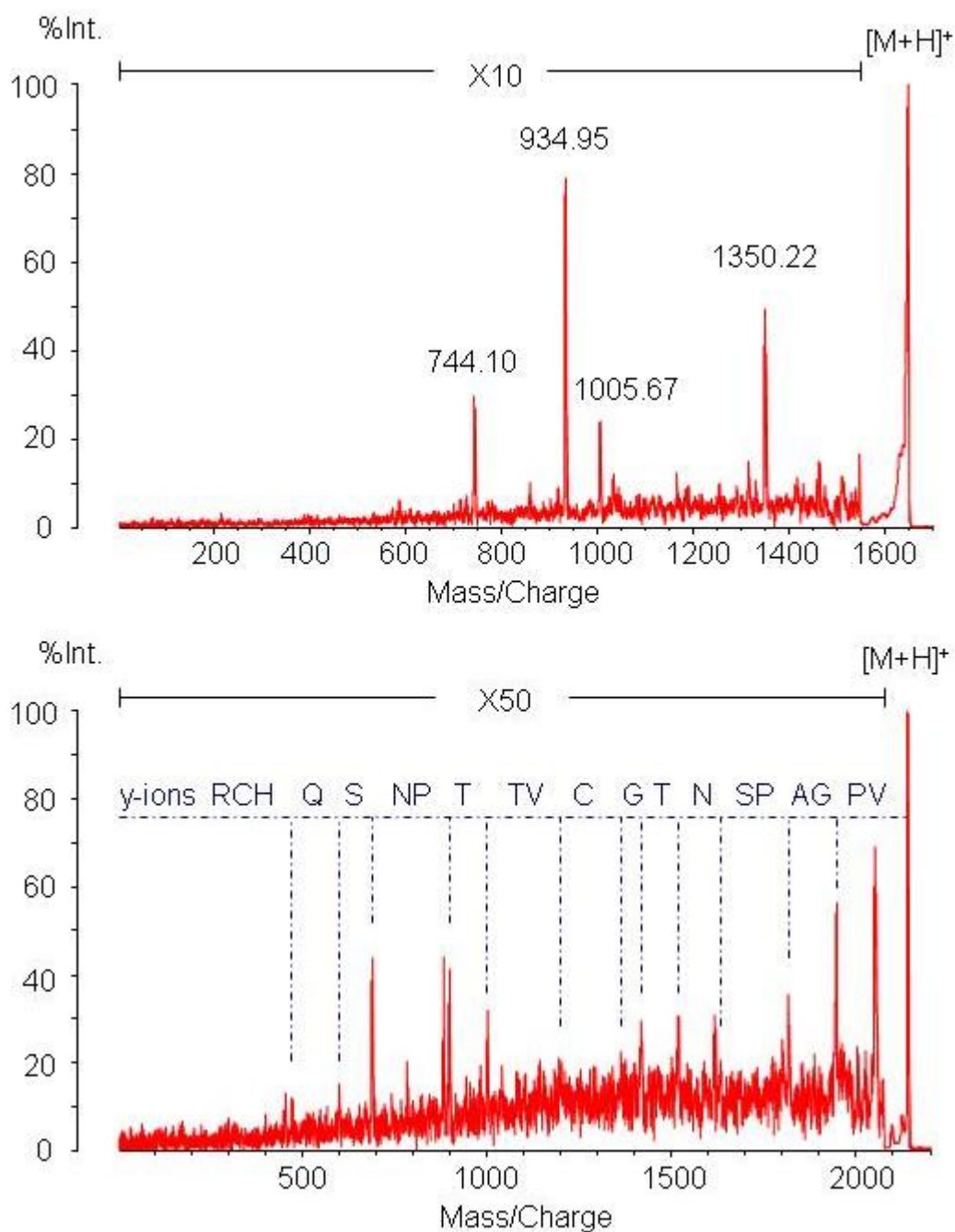


Figure 3.3-25 PSD spectra of peptide a (above, m/z 1647.54) and b (below, m/z 2139.52)

Submitted to Mascot search, peptide b led to a significant peptide hit with a Mascot ion score of 46 (search parameters: enzyme trypsin, variable modification oxidation of methionine, Threshold: 32). The respective sequence VPGAPSNTGCVTTPNSQHCR was found to belong to the protein fgenesh5_pg.C_scaffold_1000666, protein ID 103033 in the TreeseiV12predtran database. The corresponding amino acid sequence can be found in figure 3.3-26. If Mascot search is carried out with the enzyme setting “semitypsin”, peptide a is fitted to the protein as well (ion score 20), the respective sequence being

ALNPSCAPGGNFDLSK. Threshold increases to 45, but the total protein score rises at the same time to 66. The further m/z values of the PMF can not be fitted with any tryptic peptides of the identified protein.

```

MPALKTLIAS GLLGAASALN PSCAPGGNFD LSKWLSLQLPI GNP GSPTTIP
ASQLEGCGNGY QDPGHYFFFT ESGDGALVMK VPGAPSNTGC VTPNSQHCR
TELRESNPSS WSPNNPNNKL TVSLAVEQPD NSGHGTVIGQ IHIDDSVSTK
PVCELYYNSS GVLAMGVEQT RSGGNEIITP VGNVPVGQPF TYTISYSSNV
LSVSINGGAA QTLSTYSLDA PPSYFKVGNY NQGSSASDVH FYSISVSH

```

Figure 3.3-26 Amino acid sequence of the *H. jecorina* “alginate lyase”-like protein – blue: predicted signal peptide, red: identified sequence, bold: cleavage sites

The semispecific cleavage of peptide a hints that the N-terminus of the database sequence might include a signal peptide. Therefore the sequence was submitted to SignalP [81] to verify or discard the theory. The predicted signal peptide can be seen in figure 3.3-26 in blue. It does not fit with the identified sequence for peptide a, as it contains the first 18 not 17 amino acids. Therefore peptide a is either wrongly associated with the sequence or the *H. atroviridis* P1 protein shows a modification which alters the signal peptide cleavage site compared to the *H. jecorina* protein. As the proposed sequence does not explain all fragment mass ions of the PSD spectrum of peptide a, the first possibility seems more likely. Only the exact amino acid sequence data of the *H. atroviridis* P1 protein will solve this question.

Nevertheless an identification of the properties of the identified protein was carried out. An NCBI BLASTP search identified the putative conserved domain Alginate-lyase 2 and aligned the protein to the following gene products (five best scoring hits):

- BAE63841.1 unnamed protein product (*Aspergillus oryzae*) E value: $5 \cdot 10^{-79}$
- XP_001258694.1 conserved hypothetical protein (*Neosartorya fischeri* NRRL 181) E value: $8 \cdot 10^{-67}$
- NP_821977.1 hypothetical protein SAV802 (*Streptomyces avermitilis* MA-4680) E value: $1 \cdot 10^{-40}$

- YP_373156.1 hypothetical protein Bcep18194_B2401 (*Burkholderia* sp. 383) E value: 8.10^{-29}
- BAD16656.1 alginate lyase (*Sphingomonas* sp. A1) E value: 9.10^{-15}

As this list shows, most orthologues of this protein can be found in species which are not close relatives but in many cases bacteria. Therefore this gene might have occurred in the *H. atroviridis* genome because of a horizontal gene-transfer from *Streptomyces* spp.. In general alginate lyases are enzymes which cleave the glycosidic linkage of alginate, but the exact function of the *H. atroviridis* protein was not determined.

3.3.8 Spots R7 & B7

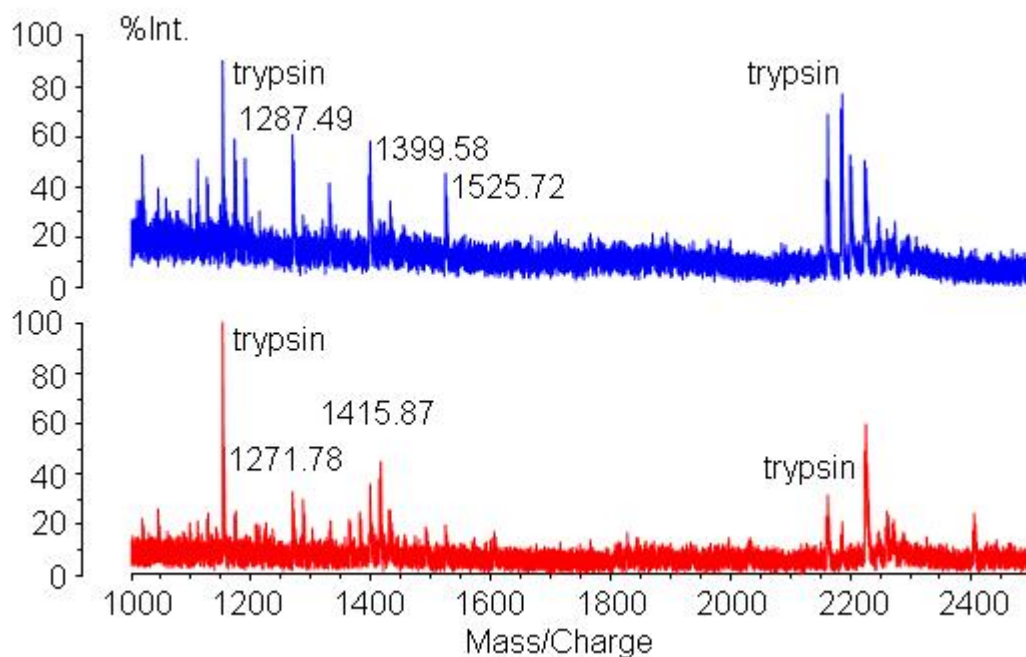


Figure 3.3-27 Peptide mass fingerprints of R7 (above) and B7 (below)

Table 3.3-8 PMF – monoisotopic m/z values of R7 and B7 peptides

<i>R. solani</i>	<i>B. cinerea</i>
1271.51	1271.78
1287.49	1287.74
	1303.70
1331.62	
	1365.85
	1381.89
1399.58	1399.82
1415.59	1415.87
1433.57	
	1491.57
1525.72	1525.98
	1606.74

The proteins analyzed as spots R7 and B7 have a molecular weight of approximately 60 kDa and their pI value is 5.1 to 5.2. There is no spot P7 because it could not be distinguished from spot P4. As spot P4 was different from R4 and B4, its peptide mass

fingerprint was compared with the ones from spot R7 and B7, but no similarities were found in this case either.

The m/z values obtained from the PMFs of spot R7 and B7 are partly the same, but the potential identity of the proteins could not be confirmed as MS/MS experiments did not lead to interpretable fragmentation patterns. The peptide mass fingerprint data of both proteins were submitted to Mascot search, but no significant hit was obtained.

3.3.9 Spots R8, B8 & P8 – Epl1

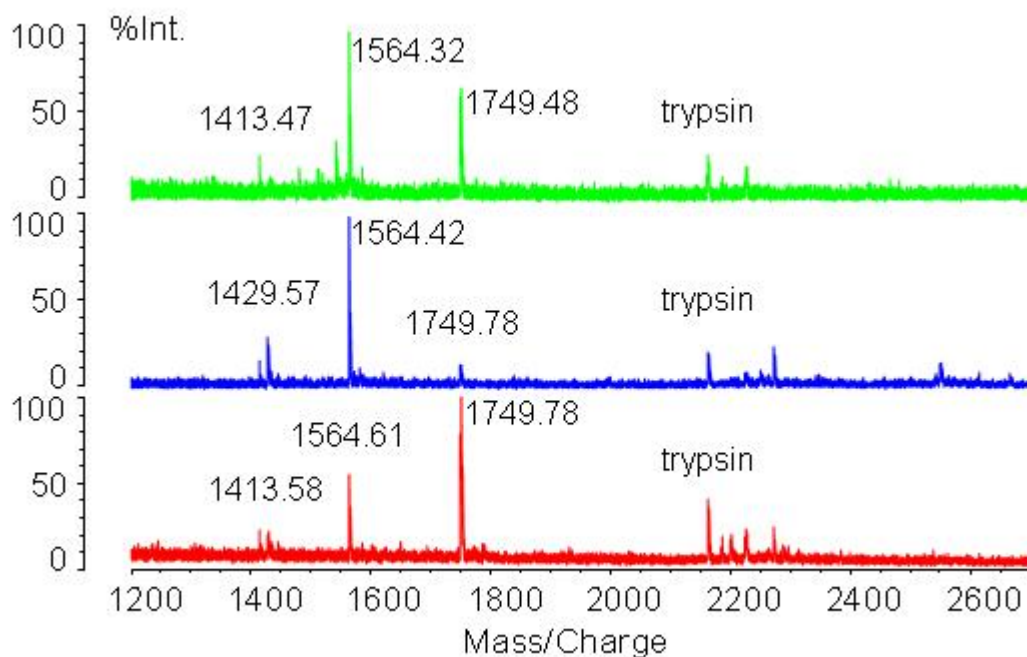


Figure 3.3-28 Peptide mass fingerprints of R8 (green), B8 (blue) and P8 (red)

Table 3.3-9 PMF - monoisotopic m/z values of R8, B8 & P8 peptides (bold = belong to the identified protein)

Peptide	<i>R. solani</i>	<i>B. cinerea</i>	<i>P. ultimum</i>
	1099.47		
	1327.40		
	1339.46		
a	1413.47	1413.55	1413.58
		1429.57	1429.67
			1445.64
	1481.50		
	1543.42		
	1559.50		
b	1564.32	1564.42	1564.61
	1621.31		
c	1749.48	1749.78	1749.78
	1891.52		

The proteins analyzed as spot R8, B8 and P8 have a molecular weight of approximately 16 kDa and a pI value of 5.5 to 5.6. These properties are the same as for spot G1, which was identified as Epl1. The peptide mass fingerprints of the proteins analyzed as spot R8, B8 and P8 showed the values of all main peptides of Epl1, but the data of spot R8 contained some additional masses. Therefore it seems that during growth on *R. solani* cell walls a second protein with similar pI and protein mass is produced additionally. Nevertheless this could not be confirmed, as further experiments showed a far lower resolution, so the small peaks which did not belong to Epl1 could not be detected.

The identification of the three spots R8, B8 and P8 as Epl1 was confirmed by the comparison of PSD spectra of the peptides (mass lists see appendix-table 11). The amount of the protein in the gels representing the growth on different cell walls was not as high as in the proteome while glucose was used as carbon source, so the signal to noise ratios of the spectra were worse. Additionally the mass calibration of the MS/MS fragmentation was misaligned during some experiments leading to a fragment mass tolerance of 2.0 Da for the proteins in spot R8 and P8. Therefore Mascot search could not identify all peptides, but significant results were obtained for the combined information for each of the three proteins (enzyme: semiTrypsin). The similarity of their MS/MS spectra is shown in figure 3.3-29.

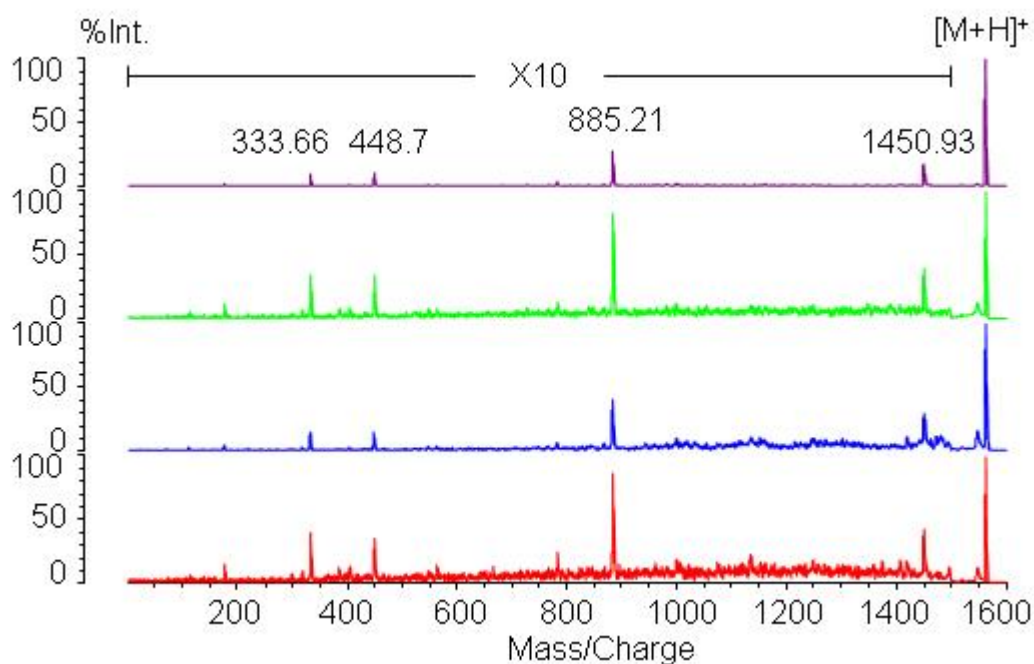


Figure 3.3-29 PSD spectra of peptide b (m/z 1564.5) of G1 (violet), R1 (green), B1 (blue) and P1 (red)

3.3.10 Spots R9, B9 & P9

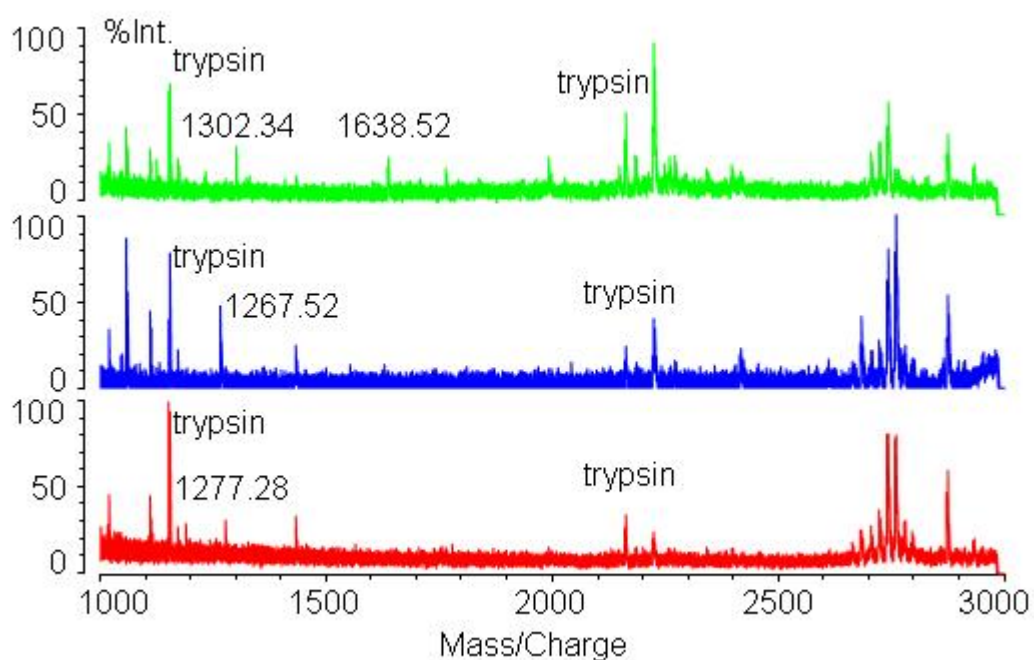


Figure 3.3-30 Peptide mass fingerprint of R9 (green), B9 (blue) and P9 (red)

Table 3.3-10PMF – monoisotopic m/z values of R9, B9 and P9 peptides

<i>R. solani</i>	<i>B. cinerea</i>	<i>P. ultimum</i>
1111.38		
	1267.52	
		1277.28
1302.34		
1433.38		
1638.52		

Spots R9, B9 and P9 are proteins with a molecular weight of approximately 13 kDa. The pI value of R9 and B9 is approximately 4.1, whereas it is 4.5 for P9. The identity or difference between the proteins can not be clearly stated, as the peptide mass fingerprints of B9 and P9 only consist of one monoisotopically determinable peptide each, whereas R9 is more or less running with the electrophoresis front and is therefore potentially contaminated. Nevertheless one peptide of R9 with an m/z of 2876.45 (average mass) was

successfully subjected to MS/MS analysis. Its PSD spectrum is shown in figure 3.3-30, the mass list can be found in appendix-table 12.

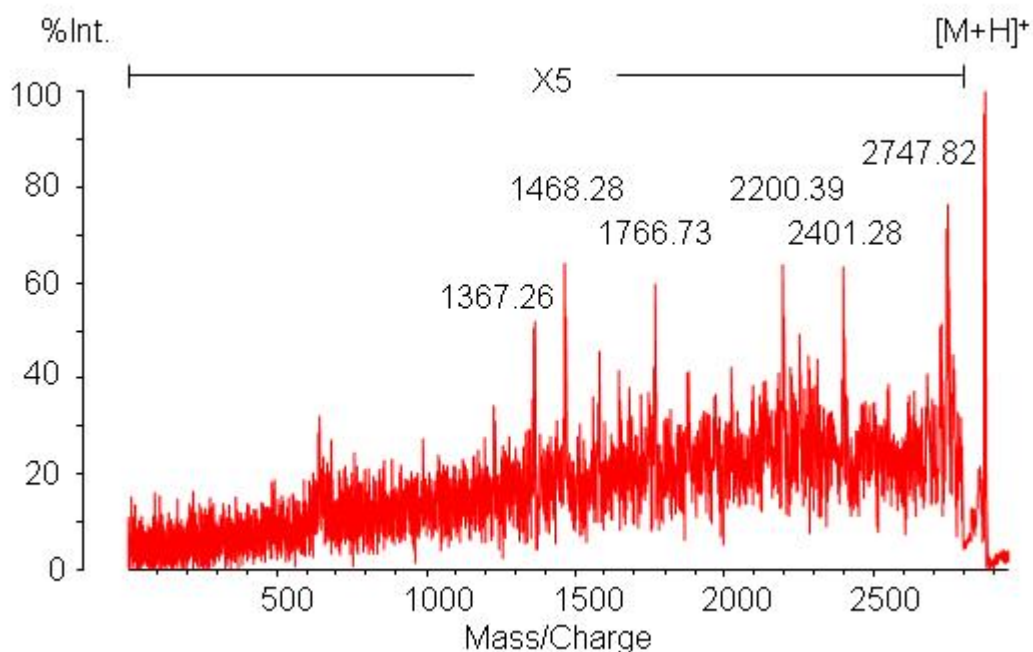


Figure 3.3-31 PSD spectrum of peptide R9a (m/z 2876.45)

Even though part of the spectrum showed clear peaks, Mascot search did not lead to any significant hit, not even a sequence tag with more than three amino acids could be obtained through analysis of the results. Manual interpretation was slightly more successful, resulting in two different tags: 1230.22-HT[IL]-1580.89 and 1344.22-VHVSN-1880.75. Whereas the first gave multiple results when submitted to Mascot search, the second led to a single protein hit. Nevertheless the assigned hypothetical protein was considered a false positive as it had not only a molecular weight of nearly 80 kDa and a pI of 9.14. In this pI region no extracellular protein was found on any of the gels and additionally the wrongly identified peptide had a mass of 5600 Da which is a difference of more than 2700 Da. So no identification was possible for spot R9.

3.4 Experiments with PNGase F

Experiments with PNGase F were carried out with the three proteins G1, G2 and G3.

To prove the activity of the enzyme, an invertase standard was used as control. Invertase from *Sacchararomyces cerevisiae* has a molecular weight of 61 kDa for the unprocessed precursor, but it is normally heavily glycosylated as it has 14 potential N-glycosylation sites. Two N-glycosylated tryptic peptides are normally recommended as indicators if the deglycosylation was successful, even though they are not the only ones which can be used for this purpose: (glycosylation site printed in bold italic)

- NPVLAAN**ST**QFR monoisotopic m/z 1317.6910
- AEPILNISNAG**P**WSR monoisotopic m/z 1624.8442

In case the deglycosylation and digestion of the invertase – and therefore normally also of the samples – was successful, additional peaks should be visible in the peptide mass finger print compared to a PMF of native glycosylated invertase. In this experiment not all potentially present or absent peaks were studied but only the two recommended indicators were considered.

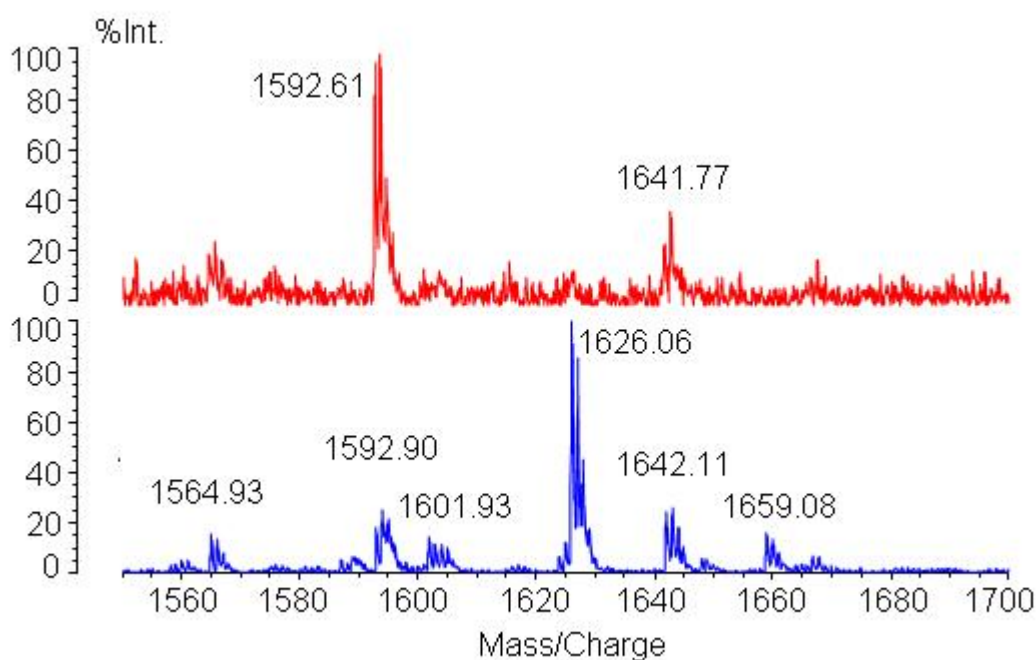


Figure 3.4-1 Relevant m/z area of the peptide mass fingerprints of invertase before (above) and after (below) PNGase F digest

Two different methods - digestion in the microwave oven and digestion over night at 37°C - were applied for deglycosylation, only the second being recommended by the manufacturer of the enzyme. The solutions resulting from both experiments were analyzed by MALDI mass spectrometry, but whereas the product of the overnight digest with invertase clearly showed one of the additional peaks (see figure 3.4-1), the digestion in the microwave oven did not work at all. The second indicator peak was not visible after neither of the two digestion conditions.

Nevertheless the samples of both experiments were analyzed. None of them showed a different peptide mass fingerprint compared to the simple tryptic digest without employment of PNGase F. Therefore no asparagine-bound N-glycanes seem to be present in these three proteins, an information which was confirmed for G1 and G2 through the knowledge of their complete sequence. As none of the proteins was part of a train of isoforms, the result had been expected that way.

4 Conclusion

Hypocrea atroviridis is an important biocontrol species and therefore especially its interactions with other organisms are of interest. The analysis of extracellular proteomes produced under different growth conditions is one of the possibilities to gain this kind of information and was the focus of this work.

The first experiments were carried out to identify the extracellular proteins produced while glucose was supplied as carbon source. Interestingly the most abundant protein was Epl1 (eliciting plant response-like protein 1). This discovery led to further investigations focused on the detailed characterization of this protein and these respective findings were already published by Seidl et al [74]. The experiments carried out in this study showed the protein presence under all other investigated growth conditions.

Furthermore the extracellular proteom secreted by *H. atroviridis* during growth on glucose was compared with those produced while *R. solani*, *B. cinerea* and *P. ultimum* cell walls were used as carbon source. As expected, the host cell walls induced the production of many different additional proteins compared to glucose because they are needed to degrade the carbohydrate polymers before the uptake. The protein pattern derived from experiments with *R. solani* and *B. cinerea* cell walls, which both contain chitin as major component of their cell wall, were very similar. Different proteins were expressed when *P. ultimum* cell walls, consisting of cellulose, were used as carbon source. Even though some of the differences might also represent artefacts resulting from sample preparation (e.g. pI differences of the same protein apparently depending on the cell wall material), the main building blocks of the host cell walls are for sure an important factor in the induction of distinct proteins or protein isoforms.

The identification of the main proteins produced under the different growth conditions was a major goal of this work. The achievement of such results was complicated by the fact that the genomic data of the examined species *H. atroviridis* were only available as an EST (expressed sequence tag) database. A close relative, *H. jecorina* (*T. reesei*) is fully sequenced, but the exchange of a single amino acid can easily complicate protein identification. Nevertheless the data which did not result in significant protein hits at the current time can be used for further searches when more sequence information is present in the databases in the future.

The proteins that could be identified by mass spectrometry pointed in the direction that a wide range of hydrolytic enzymes such as the exoglucanase or the alginate lyase are indeed

an important part of the extracellular proteome during growth on complex carbon sources. Furthermore it was shown that also proteins with other functions as for example Epl1 can form a major extracellular compound.

Future investigations can be directed either towards the analysis of further protein spots or towards increasing the knowledge about the structure and function of the already identified proteins. Additionally the pattern of posttranslational modifications of some of the proteins (especially those being part of protein isoform trains) could reveal much of their specific function and regulation. Nevertheless all these potential investigations rely on the availability of genomic data as base for the application of bioinformatics search tools. This prospect increases the probability that proteomic information might be available on a bigger scale in the near future. Only the detailed analysis of great parts of the proteomic data can lead to an understanding of the biological processes, which is important not only from a scientific point of view but also for potential industrial application.

5 References

1. Gams, W. and J. Bissett, Morphology and Identification of *Trichoderma*, in *Trichoderma and Gliocladium*, Vol. 1, Basic biology, taxonomy and genetics, C.P. Kubicek and G.E. Harman, Editors. 1998: London.
2. Druzhinina, I. and C.P. Kubicek, Species concepts and biodiversity in *Trichoderma* and *Hypocrea*: from aggregate species to species clusters? *Journal of Zhejiang University. Science. B.*, 2005. **6**(2): p. 100-12.
3. Samuels, G.J., *Trichoderma*: a review of biology and systematics of the genus. *Mycological Research*, 1996. **100**: p. 923-935.
4. Klein, D. and D.E. Eveleigh, Ecology of *Trichoderma*, in *Trichoderma and Gliocladium*, Vol. 1 Basic Biology, Taxonomy and Genetics, C.P. Kubicek and G.E. Harman, Editors. 1998: London. p. 57-74.
5. Hjeljord, L. and A. Tronsmo, *Trichoderma* and *Gliocladium* in biological control: an overview, in *Trichoderma and Gliocladium*, Vol. 2, Enzymes, biological control and commercial applications, C.P. Kubicek and G.E. Harman, Editors. 1998: London. p. 131-152.
6. Paulitz, T.C. and R.R. Belanger, Biological control in greenhouse systems. *Annual review of phytopathology*, 2001. **39**: p. 103-33.
7. Kubicek, C.P. and M. Penttila, Regulation of production of plant polysaccharide degrading enzymes by *Trichoderma*, in *Trichoderma and Gliocladium*, Vol. 2, Enzymes, biological control and commercial applications, C.P. Kubicek and G.E. Harman, Editors. 1998: London. p. 49-72.
8. Kullnig-Gradinger, C., G. Szakacs, and C.P. Kubicek, Phylogeny and evolution of the fungal genus *Trichoderma*—a multigene approach. *Mycological Research*, 2002. **106**: p. 757–767.
9. Dodd, S., E. Lieckfeldt, and G. Samuels, *Hypocrea atroviridis* sp. nov., the teleomorph of *Trichoderma atroviride* *Mycologia*, 2003. **95**: p. 27-40.
10. Cook, R.J. and K.F. Baker. The nature and practice of biological control of plant pathogens (conference proceedings). in *Conference of the American Phytopathological Society*. 1983. St. Paul, MN, USA.
11. Deacon, J.W., *Modern Mycology*. 1997: Blackwell Scientific, Oxford.
12. Weindling, R., Studies on lethal principle effective in the parasitic action of *Trichoderma lignorum* on *Rhizoctonia solani* and other soil fungi. *Phytopathology*, 1934. **24**.
13. Harman, G.E., et al., *Trichoderma species* - opportunistic, avirulent plant symbionts. *Nature Reviews Microbiology*, 2004. **2**(1): p. 43-56.
14. Benitez, T., et al., Biocontrol mechanisms of *Trichoderma* strains. *International Microbiology*, 2004. **7**(4): p. 249-260.
15. Kullnig, C., et al., Enzyme diffusion from *Trichoderma atroviride* (= *T. harzianum* P1) to *Rhizoctonia solani* is a prerequisite for triggering of *Trichoderma* ech42 gene expression before mycoparasitic contact. *Applied and Environmental Microbiology*, 2000. **66**(5): p. 2232-4.

16. Inbar, J., A. Menendez, and I. Chet, Hyphal interaction between *Trichoderma harzianum* and *Sclerotinia sclerotiorum* and its role in biological control. *Soil Biology and Biochemistry*, 1996. **28**: p. 757-763.
17. Chet, I., N. Benhamou, and S. Haran, Mycoparasitism and lytic enzymes, in *Trichoderma and Gliocladium*, Vol. 2, Enzymes, biological control and commercial applications, C.P. Kubicek and G.E. Harman, Editors. 1998: London. p. 153-172.
18. Schirmbock, M., et al., Parallel formation and synergism of hydrolytic enzymes and peptaibol antibiotics, molecular mechanisms involved in the antagonistic action of *Trichoderma harzianum* against phytopathogenic fungi. *Applied and Environmental Microbiology*, 1994. **60**(12): p. 4364-70.
19. Seidl, V., et al., A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases. *FEBS Journal*, 2005. **272**(22): p. 5923-39.
20. Kim, D.J., et al., Cloning and characterization of multiple glycosyl hydrolase genes from *Trichoderma virens*. *Current Genetics*, 2002. **40**(6): p. 374-84.
21. Kredics, L., et al., Extracellular proteases of *Trichoderma* species. A review. *Acta Microbiologica et Immunologica Hungarica*, 2005. **52**(2): p. 169-84.
22. Vizcaino, J.A., et al., Generation, annotation and analysis of ESTs from *Trichoderma harzianum* CECT 2413. *BMC Genomics*, 2006. **7**(193).
23. Vizcaino, J.A., et al., Generation, annotation, and analysis of ESTs from four different *Trichoderma* strains grown under conditions related to biocontrol. *Applied Microbiology and Biotechnology*, 2007.
24. Wisniak, J., Jöns Jacob Berzelius A Guide to the Perplexed Chemist. *The Chemical Educator*, 2000. **5**(6): p. 343-350.
25. Brown, H., F. Sanger, and R. Kitai, The structure of pig and sheep insulins. *The Biochemical Journal*, 1955. **60**(4): p. 556-65.
26. Atkins, J.F. and R. Gesteland, Biochemistry. The 22nd amino acid. *Science*, 2002. **296**(5572): p. 1409-10.
27. Wilkins, M.R., et al., Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology & Genetic Engineering Reviews*, 1996. **13**: p. 19-50.
28. Peng, J. and S.P. Gygi, Proteomics: the move to mixtures. *Journal of Mass Spectrometry*, 2001. **36**(10): p. 1083-91.
29. Tyers, M. and M. Mann, From genomics to proteomics. *Nature*, 2003. **422**(6928): p. 193-7.
30. Pluskal, M.G., Microscale sample preparation. *Nature Biotechnology*, 2000. **18**(1): p. 104-5.
31. Putnam, F.W., Alpha-, beta-, gamma-globulin--Arne Tiselius and the advent of electrophoresis. *Perspectives in Biology and Medicine*, 1993. **36**(3): p. 323-37.
32. Raymond, S. and L. Weintraub, Acrylamide gel as a supporting medium for zone electrophoresis. *Science*, 1959. **130**: p. 711.
33. Hjerten, S., Agarose as an anticonvection agent in zone electrophoresis. *Biochimica et Biophysica Acta*, 1961. **53**: p. 514-7.
34. Vesterberg, O. and H. Svensson, Isoelectric fractionation, analysis, and characterization of ampholytes in natural pH gradients. IV. Further studies on the resolving power in connection with separation of myoglobins. *Acta Chemica Scandinavica*, 1966. **20**(3): p. 820-34.

35. Klose, J., Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, 1975. **26**(3): p. 231-43.
36. O'Farrell, P.H., High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, 1975. **250**(10): p. 4007-21.
37. Görg, A., W. Postel, and S. Gunther, The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, 1988. **9**(9): p. 531-46.
38. Marchetti, M., Characterization of allergy related proteins and peptides in natural latex gloves and flowers of *Sambucus nigra* by mass spectrometry 2002, Ph.D. thesis, University of Vienna: Vienna.
39. Lottspeich, F. and H. Zorbas, *Bioanalytik*. 1998, München: Spectrum Verlag.
40. Görg, A., 2-D Electrophoresis using immobilized pH gradients - Principles and Methods. 1998, München: Amersham Biosciences.
41. Lopez, M.F., et al., A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis*, 2000. **21**(17): p. 3673-83.
42. Rabilloud, T., *Proteome research: Two-dimensional gel electrophoresis and identification methods*. 2000, Berlin, Heidelberg, New York: Springer Verlag.
43. Heukeshoven, J. and R. Dernick, Improved silver staining procedure for fast staining in PhastSystem Development Unit. I. Staining of sodium dodecyl sulfate gels. *Electrophoresis*, 1988. **9**(1): p. 28-32.
44. Shevchenko, A., et al., Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical Chemistry*, 1996. **68**(5): p. 850-8.
45. Karas, M., et al., Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion Processes*, 1987. **78**: p. 53-68.
46. Tanaka, K., et al., Protein and polymer analyses up to m/z 100.000 by laser ionization time of flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 1988. **2**(8): p. 151-170.
47. Hillenkamp, F., et al., Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Analytical Chemistry*, 1991. **63**(24): p. 1193A-1203A.
48. Schuerenberg, M., et al., Prestructured MALDI-MS sample supports. *Analytical Chemistry*, 2000. **72**(15): p. 3436-42.
49. Thomas, J.J., et al., Desorption/ionization on silicon (DIOS): a diverse mass spectrometry platform for protein characterization. *Proceedings of the National Academy of Science of the USA*, 2001. **98**(9): p. 4932-7.
50. Kussmann, M., et al., Matrix-assisted laser desorption/ionization mass spectrometry sample preparation techniques designed for various peptide and protein analytes. *Journal of Mass Spectrometry*, 1997. **32**(6): p. 593-601.
51. Zenobi, R. and R. Knochenmuss, Ion formation in MALDI mass spectrometry. *Mass Spectrometry Reviews*, 1998. **17**(5): p. 337-366.
52. Siuzdak, G., *Mass spectrometry for biotechnology*. 1996, San Diego, CA, USA: Academic Press.

53. Brown, R.S. and J.J. Lennon, Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer. *Analytical Chemistry*, 1995. **67**(13): p. 1998-2003.
54. Mamyrin, B., et al., Mass reflectron. New nonmagnetic time-of-flight high-resolution mass spectrometer. *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki*, 1973. **64**: p. 82 et sqq.
55. Cornish, T.J. and R.J. Cotter, A curved-field reflectron for improved energy focusing of product ions in time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 1993. **7**(11): p. 1037-40.
56. Papayannopoulos, I.A., The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrometry Reviews*, 1995. **14**(1): p. 49-73.
57. Roepstorff, P. and J. Fohlman, Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry*, 1984. **11**(11): p. 601 et sqq.
58. Johnson, R.S., et al., Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Analytical Chemistry*, 1987. **59**(21): p. 2621-5.
59. Henzel, W.J., J.T. Stults, and C. Watanabe. *Proceedings of the Third Symposium of the Protein Society*. 1989. Seattle, WA.
60. Henzel, W.J., et al., Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proceedings of the National Academy of Science of the USA*, 1993. **90**(11): p. 5011-5.
61. James, P., et al., Protein identification by mass profile fingerprinting. *Biochemical Biophysical Research Communications*, 1993. **195**(1): p. 58-64.
62. Mann, M., P. Hojrup, and P. Roepstorff, Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biological Mass Spectrometry*, 1993. **22**(6): p. 338-45.
63. Pappin, D.J., P. Hojrup, and A.J. Bleasby, Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology*, 1993. **3**(6): p. 327-32.
64. Yates, J.R., 3rd, et al., Peptide mass maps: a highly informative approach to protein identification. *Analytical Biochemistry*, 1993. **214**(2): p. 397-408.
65. Mann, M. and M. Wilm, Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical Chemistry*, 1994. **66**(24): p. 4390-9.
66. Eng, J.K., A.L. McCormack, and J.R. Yates, 3rd, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 1994. **5**: p. 976-989.
67. Clauser, K.R., P. Baker, and A.L. Burlingame, Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Analytical Chemistry*, 1999. **71**(14): p. 2871-82.
68. Zhang, W. and B.T. Chait, ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Analytical Chemistry*, 2000. **72**(11): p. 2482-9.
69. Perkins, D.N., et al., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 1999. **20**(18): p. 3551-67.
70. Yates, J.R., 3rd, J.K. Eng, and A.L. McCormack, Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Analytical Chemistry*, 1995. **67**(18): p. 3202-10.

71. Shevchenko, A., et al., Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proceedings of the National Academy of Science of the USA, 1996. **93**(25): p. 14440-5.
72. Creasy, D.M. and J.S. Cottrell, Unimod: Protein modifications for mass spectrometry. Proteomics, 2004. **4**(6): p. 1534-6.
73. Seidl, V., Molecular and physiological investigations of biocontrol by the genus *Hypocrea/Trichoderma* 2006, Ph.D. thesis, Vienna University of Technology: Vienna.
74. Seidl, V., et al., Epl1, the major secreted protein of *Hypocrea atroviridis* on glucose, is a member of a strongly conserved protein family comprising plant defense response elicitors. FEBS Journal, 2006. **273**(18): p. 4346-59.
75. Bienvenut, W.V., et al., Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. Proteomics, 2002. **2**(7): p. 868-76.
76. Pazzagli, L., et al., Purification, characterization, and amino acid sequence of cerato-platanin, a new phytotoxic protein from *Ceratocystis fimbriata* f. sp. platani. Journal of Biological Chemistry, 1999. **274**(35): p. 24959-64.
77. Pazzagli, L., et al., Cerato-platanin, the first member of a new fungal protein family: cloning, expression, and characterization. Cell Biochemistry and Biophysics, 2006. **44**(3): p. 512-21.
78. Altschul, S.F., et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 1997. **25**(17): p. 3389-402.
79. Margolles-Clark, E., et al., Cloning of genes encoding alpha-L-arabinofuranosidase and beta-xylosidase from *Trichoderma reesei* by expression in *Saccharomyces cerevisiae*. Applied and Environmental Microbiology, 1996. **62**(10): p. 3840-6.
80. Pakula, T.M., et al., Monitoring the kinetics of glycoprotein synthesis and secretion in the filamentous fungus *Trichoderma reesei*: cellobiohydrolase I (CBHI) as a model protein. Microbiology, 2000. **146** (Pt 1): p. 223-32.
81. Nielsen, H., et al., A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. International Journal of Neural Systems, 1997. **8**(5-6): p. 581-99.

6 Appendix

Appendix - Table 1: MH⁺ and m/z lists from CID MS/MS experiments of protein spots G1 and G2

peptide	MH ⁺ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
G12a	1414.29	58.89:12.42, 70.01:265.89, 84.26:117.31, 86.61:222.09, 101.23:127.25, 110.24:224.26, 112.19:47.58, 130.26:65.81, 136.28:54.67, 159.31:63.15, 166.31:49.11, 170.29:33.94, 175.32:22.1, 185.35:18.57, 186.29:19.45, 201.31:23.21, 229.64:76.36, 255.37:49.92, 270.35:57.8, 272.37:77.09, 273.34:64.9, 284.3:27.35, 296.33:18.85, 301.28:117.37, 327.34:19.33, 340.26:77.25, 439.68:65.18, 487.19:24.19, 496.26:20.14, 570.12:17.36, 624.76:65.39, 627.27:29.39, 698.72:80.46, 725.18:26.59, 753.82:63.06, 854.02:133.96, 927.92:23.49, 953.27:34.86, 982.21:55.96, 1069.46:22.45, 1114.96:23.84, 1168.46:34.5, 1284.3:32.76, 1306.79:216.69, 1371.06:77.55, 1398.82:229.8
G12a ₁	1430.29	58.79:35.87, 69.95:641.67, 83.99:352.98, 86.46:423.71, 100.96:274.68, 110.32:565.31, 112.37:95.38, 129.16:26.27, 130.29:35.28, 136.37:101.47, 138.24:31.46, 157.95:87.12, 166.38:153.76, 175.39:90.1, 183.24:44.95, 185.51:120.52, 201.43:48.78, 212.48:18.24, 229.19:217.36, 255.54:181.97, 270.46:182.02, 272.2:168.93, 273.44:113.56, 284.71:68.2, 301.31:257.59, 313.46:25.16, 327.36:42.76, 340.43:182.24, 354.5:14.42, 368.83:52.96, 385.55:63.16, 414.43:17.41, 439.68:142.71, 485.03:34.71, 496.32:81.29, 503.5:71.78, 512.23:27.9, 539.7:14.27, 569.55:97.67, 596.25:16.12, 624.26:213.32, 698.33:186.67, 725.41:89.36, 754.06:117.57, 767.19:52.71, 783.14:32.08, 799.1:45.76, 808.63:51.01, 853.88:238.58, 910.9:77.43, 917.6:20.93, 927.72:88.22, 953.15:84.87, 981.78:238.08, 1045.78:32.67, 1109.47:106.68, 1118.73:49.01, 1130.46:90.16, 1139.44:38.85, 1158.31:49.85, 1184.89:35, 1226.23:101.52, 1250.82:26.08, 1266.97:39.04, 1283.86:197.07, 1300.07:37.54, 1307.54:47.79, 1323.11:221.42, 1371.68:91.93, 1386.84:100.34
G12a ₂	1446.26	558.78:55.79, 69.92:1017.63, 83.72:678.88, 86.42:506.18, 100.76:466.31, 110.27:764.97, 136.35:186.55, 166.14:271.19, 174.43:155.04, 181.52:39.42, 185:183.59, 201.6:98.23, 229.28:318.52, 255.35:260.72, 270.53:815.04, 282.8:49.41, 284.64:111.7, 299.6:270.98, 301.35:227.4, 310.65:24.83, 327.56:50.77, 340.29:261.44, 354.3:23.37, 356.24:56.4, 368.46:99.07, 385.02:91.73, 412.45:42.19, 439.17:216.87, 483.83:40.67, 496.7:50.5, 501.32:128.91, 511.13:70.56, 519.17:106.21, 569.89:126.34, 624.34:349.2, 698.25:306.6, 725.94:89.8, 753.44:121.46, 766.71:56.09, 798.43:109.34, 853.84:360.6, 910.16:48, 927.58:257.81, 952.31:65.72,

		981.24:364.51, 1043.48:83.19, 1061.69:63.73, 1107.37:70.11, 1119.18:73.08, 1126.31:209.96, 1146.73:40.5, 1174.46:106.98, 1225.96:60.68, 1283.26:312.55, 1297.44:54.89, 1338.79:118.36
G12b	1565.58	59.76:43.04, 70.09:186.42, 72.6:368.63, 87.37:366.44, 100.3:87.73, 112.47:120.11, 116.29:22.94, 136.46:298.8, 147.54:11.66, 158.98:110.46, 175.52:174.36, 185.46:38.67, 189.47:47.22, 201.35:27.6, 217.49:77.62, 230.25:92.68, 245.71:84.43, 262.39:35.02, 298.47:27.31, 300.43:50.48, 317.38:197.22, 333.66:353.71, 388.04:83.19, 403.34:60.96, 432.24:26.49, 448.7:260.71, 503.28:16.82, 563.15:23.1, 618.74:30.17, 784.17:201.77, 839.69:102.77, 852.25:99.12, 868.28:33.44, 885.21:514.87, 982.55:63.85, 999.9:48.22, 1162.73:49.29, 1250.83:45.17, 1349.1:32.21, 1450.93:229.5, 1505.12:93.5, 1522.88:244.7, 1548.6:588.62
G12c	1750.65	59.21:24.67, 71.66:444.07, 86.38:825.41, 100.33:105.72, 112.39:121.52, 133.21:24.34, 158.4:47.11, 173.41:173.72, 175.36:252.2, 199.44:28.33, 201.48:201.92, 230.41:68.51, 243.86:66.78, 259.65:113.22, 276.38:32.25, 284.85:90.22, 302.63:159.6, 331.46:54.7, 344.6:97.56, 383.26:17.74, 401.05:22.63, 443.88:142.86, 501.25:98.7, 559.83:559.87, 613.62:61.77, 629.02:56.91, 657.06:59.15, 674.8:86.18, 743.52:84.02, 802.43:603.96, 917.34:46.04, 971.9:77.03, 1058.78:73.64, 1164.95:52.42, 1218.91:43.47, 1449.24:54.87, 1662.85:23.75, 1707.86:311.62
G12d	2543.90	933.59:99.99, 995.49:56.87, 1033.08:70.83, 1083.19:88.01, 1197.6:123.18, 1382.53:98.73, 1440.87:139.63, 1512.39:301.18, 1683.95:183.45, 1809.78:227.32, 2023.06:115.82, 2045.92:114.77, 2393.15:275.64, 2450.07:63.62

Appendix - Table 2: MH⁺ and m/z lists from CID MS/MS experiments of protein spot G3

peptide	MH ⁺ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
G3a	742.77	30.19:71.86, 39.16:33.68, 43.37:212.84, 44.35:217.11, 55.38:95.78, 59.43:102.17, 60.36:26.49, 69.49:80.55, 70.53:264.19, 72.43:412.27, 84.52:68.79, 86.58:1053.61, 87.55:254.37, 100.46:46.77, 102.58:92.4, 112.56:84.32, 158.53:18.78, 175.68:379.86, 199.89:376.69, 215.67:77.31, 227.84:484.51, 243.67:303.34, 257.76:203.48, 274.76:82.18, 329.01:161.06, 357.02:947.29, 370.94:136.27, 387.94:294.69, 442.19:152.37, 469.73:186.02, 570.69:69.85, 629.94:76.56, 685.71:41.46, 698.67:38.3
G3b	1198.70	43.63:19.82, 55.57:31.68, 59.14:23.23, 60.14:15.42, 70.33:90.41, 72.48:157.49, 86.46:204.11, 87.41:92.9, 100.47:20.56, 112.9:42.23, 133.5:19.33, 143.61:11.35, 169.55:10.54, 175.6:28.79, 201.66:10.02, 230.03:49.69, 260.67:26.3, 275.57:17.27, 375.50:63, 412.29:29.12, 454.83:16.26, 483.37:36.42, 488.11:25.1, 525.77:22.37, 582.63:33.73, 599.44:13.72, 654.48:17.04, 668.14:30.14,

		695.8:12.64, 767.21:47.7, 825.84:12.92, 880.49:31.89, 995.58:49.2, 1108.46:58.04, 1125.41:31.83, 1156.61:53.9
G3c	1246.00	70.21:282.08, 72.44:552.24, 86.64:1268.25, 100.19:21.66, 104.31:62.81, 112.43:110.34, 133.39:80.98, 143.58:47.15, 175.56:196.23, 200.96:28.74, 229.63:179.66, 232.69:28.12, 260.63:190.87, 275.34:88.28, 283.82:81.82, 328.96:29.91, 374.75:495.09, 388.34:98.39, 399.86:148.61, 459.58:157.46, 487.58:284.72, 530.93:37.57, 647.34:86.05, 657.94:43.69, 702.04:61.75, 715.23:50.97, 759.99:52.48, 813.94:112.08, 927.57:48.79, 1075.73:77.22, 1078.52:61.74, 1156.01:134.22
G3d	1397.07	70.22:230.86, 72.15:126.01, 73.96:39.6, 86.9:295.57, 101.1:61.02, 120.45:150.38, 175.56:92.4, 221.65:82.98, 229.61:104.8, 249.65:74.93, 342.79:146.85, 363.71:217.29, 448.87:197.12, 462.63:107.64, 644:161.15, 660.44:142.96, 715.14:102.13, 788.5:43.34, 1354.3:191.58
G3e	2057.29	69.74:90.88, 85.44:167.83, 120.53:44.09, 129.77:72.45, 136.87:77.04, 141.69:20.46, 147.89:8.19, 170.01:36.65, 183.98:13.14, 186.9:25.78, 212.15:25.05, 222.06:3.89, 283.01:6.73, 311.33:4.59, 326.13:21.17, 340.3:8.51, 355.26:10.3, 426.01:2.96, 459.37:24.91, 487.48:14.38, 539.66:15.15, 597.03:10.6, 601.67:35.11, 643.72:11.83, 669.47:5.53, 697.71:3.22, 715.12:37.78, 743.87:30.71, 766.03:39.24, 829.14:16.59, 857.28:46.84, 879:15.71, 926.25:22.02, 954.32:10.94, 971.53:115.92, 1089.87:210.36, 1154.17:24.02, 1182.28:88.02, 1204.33:75.04, 1267.32:19.4, 1295.15:126.89, 1316.56:19.09, 1430.63:31.77, 1457.34:98.62, 1520.76:81.69, 1678.53:38.72, 1748.77:38.52, 1802.49:101.81, 1913.92:45.66, 1930.61:25.08, 1986.38:17.16

Appendix - Table 3: MH⁺ and m/z lists from PSD MS/MS experiments of protein spot G3

peptide	MH ⁺ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
G3b	1198.70	175.65:8.8, 260.78:4.38, 288.63:1.96, 345.82:4.65, 374.92:34.09, 388.72:2.99, 396.76:3.69, 411.97:11.97, 460.11:9.4, 488.23:27.22, 525.02:6.99, 558.46:6.25, 582.42:29.78, 599.28:19.81, 658.23:19.66, 695.67:15.85, 712.99:16.21, 808.72:9.73, 825.77:16.94, 922.79:19.78, 997.07:53.87, 1107.83:59.66, 1124.12:64.35
G3e		539.65:9.95, 602.21:11.63, 715.06:27.14, 743.55:21.53, 765.68:24.3, 828.56:6.6, 857.26:47.19, 878.87:12.58, 925.81:29.29, 943.09:8.82, 953.55:25.25, 971.5:148.98, 1039.14:19.81, 1068.47:23.26, 1089.43:262.76, 1153.61:27.46, 1165.57:26.51, 1172.24:5.46, 1182.08:122.2, 1203.64:128.9, 1267.44:43.96, 1294.84:229.39, 1317.81:51.12, 1457.95:154.38, 1515.07:65.23, 1520.51:83.01, 1634.84:49.38, 1649.55:45.44, 1662.62:19.71, 1678.14:80.18, 1706.05:12.05, 1723.73:10.82, 1748.91:85.17, 1802.56:170.65, 1912.52:101.71, 1930.97:55.04

Appendix - Table 4: MH⁺ and m/z lists from PSD MS/MS experiments of protein spots R1, R2, B1, B2, P1 and P2

peptide	MH ⁺ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
R1a	1302.4	181.42:4.24, 482.93:5.41, 568.02:14.3, 571.66:7.72, 575.45:5.63, 585.06:5.83, 606.6:14.76, 632.42:13.32, 639.1:13.93, 645.99:8.62, 680.68:18.85, 687.13:14.96, 697.15:15.07, 757.68:22.95, 808.7:34.5, 830.44:11.17, 859.47:51.02, 866.38:54.79, 878.98:42.03, 886.55:26.42, 895.87:60.17, 923.35:53.53, 940.44:46.43, 967.13:104.38, 979.98:16.53, 1022.2:17.13, 1038.79:64.1, 1049.44:21.99, 1057.29:31.84, 1079.58:41.39, 1121.02:29.36, 1156.86:53.99, 1165.62:64.15, 1180.69:57.44, 1198.78:44.81, 1214.95:53.72, 1222.97:23.35, 1241.2:50.99
R1b	1318.4	484.23:3.74, 494.07:4.77, 521.35:3.77, 548.23:4.59, 569.56:6.13, 584.04:7.48, 636.15:9.14, 678.66:16.41, 724.2:10.64, 728.38:7.05, 743.33:21.18, 755.57:27.06, 832.45:41.58, 877.14:45.63, 894.99:30.54, 921.54:49.18, 939.63:43.98, 964.72:56.05, 984.17:66.37, 1055.91:55.85, 1109.44:33.2, 1135.08:33.75, 1154.48:49.44, 1171.76:66.38, 1197.63:58.72, 1212.01:58.89, 1231.24:46.85, 1255.84:84.68
R1c	1856.9	435.56:0.55, 459.36:0.77, 461.68:0.71, 505.88:0.17, 578.61:17.31, 778.45:17.14, 1462.34:29.93, 1751.06:56.43
R2a	1302.4	451.51:13.98, 496.28:26.76, 539.49:39.06, 757.22:208.71, 807.17:80.09, 834.64:191.54, 852.38:136.43, 879.82:129.82, 896.09:203.24, 923.03:196.36, 966.71:367.35, 1038.74:281.67, 1112.83:185.72, 1157.01:147.48, 1179.86:108.02, 1201.19:148.38, 1214.16:119.13, 1240.13:92.03, 1245.03:117.5
R2b	1318.4	494.44:16.29, 509.75:14.43, 538.84:27.97, 584.15:46.27, 636.91:17.69, 662.27:30, 671.86:58.87, 725.54:22.6, 755.99:125.4, 777.44:48.13, 806.98:49.63, 834.18:158.56, 851.51:123.02, 867.35:116.9, 875.31:90.12, 893.32:159.39, 920.72:256.97, 939.54:184.94, 964.82:254.65, 983.65:135.07, 996.42:113.87, 1010.38:106.28, 1054.83:147.57, 1129.33:120.7, 1159.01:185.86, 1173.73:188.38, 1198.32:183.39, 1217.44:118.83, 1230.23:140.04, 1255.79:186.41
R2c	1856.9	503.76:4.02, 536.34:2.67, 578.33:13.81, 623.92:4.39, 638.75:4.56, 650.06:4.37, 669.28:4.26, 711.82:5.23, 883.17:9.12, 1034.34:22.02, 1397.9:19.36, 1404.06:73.77, 1429.34:45.46, 1466.7:53.93, 1534.28:18.47, 1652.07:59.02, 1763.37:13.63, 1799.99:84.93
B1a	1302.4	450.15:5.71, 472.97:12.23, 495.07:30.42, 539.47:50.91, 654.42:29.91, 683.07:31.54, 700.17:75.49, 713.45:81.95, 721.2:27.64, 726.34:33.43, 739.47:29.61, 757.92:66.32, 805.92:147.93, 828:191.72, 851.58:386.94, 878.92:183.84, 895.95:362.99, 913.97:171.29, 923.43:226.79, 942.33:85.33,

		967.13:308.9, 994.72:97.75, 1012.27:129.73, 1038.49:486.56, 1062.79:571.01, 1083.91:262.12, 1102.75:181.96, 1156.59:401.57, 1170.11:123.89, 1182.71:88.43, 1200.56:203.46, 1214.21:148.85, 1240.98:193.04
B1b	1318.4	451.64:5.32, 474.54:7.45, 496.11:16.68, 540.33:41.06, 584.93:30.62, 653.5:30.02, 655.42:24.04, 673.09:15.82, 699.01:51.69, 702.83:40.24, 742.28:24.8, 755.84:75.61, 774.24:29.61, 806.74:60.62, 830.35:246.44, 851.27:275.51, 895.15:275.69, 896.9:255.35, 913.48:340.37, 921.43:173.49, 927.76:130.65, 939.38:300.71, 957.25:268.68, 965.54:246.46, 984.58:166.36, 1013:287.69, 1056.48:275.38, 1060.51:338.06, 1086.97:351.1, 1102.58:376.63, 1108.67:176.75, 1116.39:102.78, 1119.68:166.84, 1129.87:163.31, 1212.52:115.79, 1217.94:59.39, 1230.74:95.53, 1256.08:146.64
B2a	1302.4	923.16:36.70, 967.03:73.68, 1038.98:22.06
B2b	1318.4	492.5:7.64, 495.81:12.63, 521.21:4.78, 539.04:10.57, 584.88:26.21, 653.19:22.73, 663.67:10.99, 756.89:48.54, 775.44:53.01, 807.88:66.1, 851.79:80.13, 876.52:110.78, 894.77:144.75, 912.28:131.66, 920.87:193.15, 939.26:149.39, 964.66:171.82, 983.49:103.19, 1055.7:244.55, 1110.1:147.47, 1128.81:47.17, 1154.54:108.29, 1172.14:147.4, 1199.73:97.54, 1217.47:142.73, 1230.9:86.47, 1255.38:213.2
P1a	1302.4	184.57:4.17, 470.44:15.29, 493.89:27.96, 539.62:39.26, 677.27:50.86, 715.02:61.29, 757.67:208.84, 804.3:172.65, 851.45:219.45, 875.44:102.46, 896.04:181.03, 922.75:241.14, 941.12:53.18, 950.88:42.84, 966.84:188.11, 1039.17:133.99, 1158.24:377.75, 1201.39:97.73, 1213.89:88.78, 1241.4:134.02
P1b	1318.4	495.33:87.81, 521.54:12.9, 539.84:100.68, 585.26:132.48, 651.46:44.95, 699.09:55.39, 713.08:44.39, 743.52:67.3, 756.13:238.47, 774.7:76.3, 798.52:26.42, 806.93:58.63, 850.15:117.47, 868.33:107.37, 877.92:178.38, 895.96:167.72, 913.66:123.17, 921.24:215.54, 939.56:153.62, 965.64:313.61, 984.05:204.05, 1055.89:439.17, 1085.17:75.07, 1111.43:157.06, 1129.77:129.29, 1156.88:152.46, 1172:208.87, 1213.46:105.9, 1218.16:99.78, 1230.95:105.77, 1256.35:352.27, 1261.01:234.67
P1c	1856.9	174.67:1.7, 261.93:1.06, 333.78:0.51, 360.06:2.19, 448.62:4.49, 477.4:6.34, 531.46:7.48, 578.32:49.81, 639.2:10.96, 764.54:94.58, 785.17:83.37, 1052.38:134.68, 1226.77:171.91, 1299.38:130.77, 1595.18:252.16, 1673.34:227.33, 1717.85:95.11, 1750:123.06
P2a	1302.4	180.56:2.41, 194.52:0.9, 437.92:1.08, 449.8:2.85, 474.95:2.58, 495:26.56, 516.96:4.25, 538.12:39.14, 635.59:9.44, 650.33:14.44, 678.28:8.43, 681.42:13.5, 694.04:23.47, 699.13:21.87, 713.52:58.15, 726.7:35.81, 731.83:31.27, 750.48:30.69, 757.05:36.41, 806.43:76.94, 819.89:30.83, 831.51:113.36, 850.4:147.09, 863.62:45.8, 877.58:86.05, 895.49:126.89, 906.08:43.68, 908.45:91.24, 922.38:143.27,

		934.42:47.72, 941.77:45.82, 952.54:98.82, 966.65:141.35, 994.82:52.77, 1026.16:35.67, 1038.68:73.73, 1097.08:42.97, 1110.95:48.04, 1155.13:189.3, 1180.03:104.46, 1200.55:150.16, 1213.36:91.67, 1242.41:143.33
P2b	1318.4	179.68:9.08, 193.17:6.75, 437.37:9.74, 450.75:15.8, 495.31:73.61, 539.68:109.41, 584.73:148.33, 627.69:38.72, 653.02:41.79, 668.12:24.85, 677.99:31.84, 699.5:69.14, 712.9:45.69, 725.09:46.35, 742.84:88.67, 755.61:273.9, 774.26:133.49, 799.75:22.84, 806.65:37.4, 824.06:95.67, 832.18:141.09, 835.59:89.52, 850.58:146.62, 862.44:36.23, 867.97:126.3, 875.71:186.36, 894.77:222.24, 913.27:169.85, 921.71:169.85, 939.75:255.94, 957.49:118.4, 966.19:169.14, 983.61:260.49, 1012.29:180.95, 1055.14:312.63, 1068.04:137.85, 1110.56:108.97, 1129.39:169.8, 1154.2:100.1, 1158.29:186.3, 1171.31:287.3, 1187.04:97, 1217.23:224.66, 1230.96:190.05, 1256.62:327.46
P2c	1856.9	140.39:2.96, 174.58:2.84, 262.63:4.67, 389.69:3.43, 476.71:7.32, 504.8:13.42, 534.7:13.11, 554.57:6.78, 578.46:93.56, 639.71:19.37, 667.02:11.7, 694.07:12.58, 785.41:10.38, 807.11:8.14, 840.24:17.6, 938.54:23.44, 1026.33:28.65, 1171.42:46.56, 1356.21:42.87, 1467.36:30.55, 1595.16:26.16, 1749.98:26.16

Appendix - Table 5: MH⁺ and m/z lists from PSD MS/MS experiments of protein spot R3

peptide	MH ⁺ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
R3a	1603.57	291.24:0.18, 340.84:5.34, 437.56:11.42, 476.49:7.51, 504.22:30.8, 575.42:112.49, 592.47:24.13, 659.83:40.92, 688.81:215.44, 706.52:90.06, 759.87:63.96, 769.38:52.02, 786.7:245.25, 803.87:29.94, 821.09:150.38, 884.66:43.81, 901.17:119.96, 918.64:134.9, 985.89:41.26, 996.85:31.76, 1014.93:157.32, 1031.73:78.26, 1048.28:52.89, 1059.54:27.69, 1086.71:212.07, 1103.13:79.71, 1200.6:196.45, 1313.24:200.2, 1370.62:78.28, 1380.23:22.58, 1456.99:85.02, 1466.4:28.19, 1475.61:58.06, 1542.06:145.85

Appendix - Table 6: MH⁺ and m/z lists from PSD MS/MS experiments of protein spots R4 and B4

peptide	MH ⁺ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
R4a	1056.1	70.19:12.1, 112.32:0.73, 159.26:4.77, 175.35:89.11, 184.48:4.59, 212.35:119.28, 229.39:12.26, 258.31:1.1, 271.29:19.67, 284.34:4.18, 288.39:9.85, 297.38:15.36, 300.51:1.71, 325.42:179.23, 342.43:27.63, 359.33:21.06, 371.43:13.79, 398.38:74.97, 414.54:0.76, 454.55:2.49, 466.68:37.59, 483.69:97.09, 494.84:20.97, 511.59:260.23, 528.54:39.52, 545.77:37.27, 555.73:13.72, 582.74:74.87,

		599.44:1.74, 638.48:3.35, 642.07:5.96, 652.46:23.67, 658.63:41.41, 670.54:28.04, 681.94:9.67, 697.47:69.64, 710.9:1.8, 722.38:5.54, 740.85:9.84, 751.46:21.48, 755.68:51.76, 768.5:67.8, 852.44:24.18, 869.84:597.92, 888.41:7.6, 890.34:5.79, 906.06:63.52, 926.16:5.48, 941.85:5.24, 960.7:2.58, 983.78:2.36, 1002.88:4.56
R4b	1492.5	112.28:4.7, 175.13:53.78, 245.33:2.24, 264.12:5.52, 289.15:0.5, 306.71:3.42, 321.09:25.68, 346.84:6.9, 378.15:8.57, 394.82:21.14, 404.12:36.29, 421.75:71.33, 441.72:1.34, 452.1:1.22, 469.67:11.28, 479.21:30.34, 491.21:4.79, 496.14:26.51, 524.27:10.65, 541.03:66.67, 551.07:51.75, 568.89:65.01, 583.1:18.41, 609.13:6.21, 627.35:10.13, 638.47:15.74, 654.4:55.87, 664.58:6.22, 681.71:33.64, 699.02:17.45, 706.93:12.19, 724.06:239.85, 741.26:135.89, 752.69:22.32, 770.06:5.39, 795:105.42, 812.67:89.34, 854.37:6.81, 868.07:7.27, 908.45:12.23, 924.98:51.21, 973.28:8.43, 983.16:12.12, 996.26:54.42, 1027.68:26.76, 1065.67:34.82, 1072.3:104.9, 1154.86:24.52, 1173.21:30.82, 1212.05:36.9, 1229.5:78.41, 1286.4:63.65, 1369.96:82.54, 1383.93:187.15, 1392.96:43.98, 1431.55:10.97
R4c	1607.6	135.51:9.33, 175.16:31.92, 201.37:3.93, 239.61:2.44, 257.7:2.69, 290.79:2.78, 308.25:4.02, 367.66:8.21, 404.98:8.33, 421.68:6.74, 479.44:5.99, 490.38:18.82, 509.64:11.42, 517.91:12.05, 531.72:10.09, 549.51:28.72, 565.54:13.74, 576.8:8.83, 586.74:11.08, 624.59:12.22, 643.96:30.58, 653.97:87, 679.06:22.21, 685.69:32.7, 698.62:15.75, 723.48:2.84, 730.04:68.27, 785.91:48.76, 797.37:17.11, 805.54:17.23, 818.18:39.22, 822.88:93.6, 843.41:30.37, 893.5:49.39, 919.34:61.28, 938.98:34.54, 1006.07:25.04, 1014.52:47.72, 1061.44:52.91, 1099.75:62.65, 1192.13:89.91, 1368.46:169.13, 1442.52:65.02, 1470.92:77.2
R4d	1828.8	127.65:6.62, 174.3:19.24, 247.04:5.03, 286.26:7.61, 325.18:28.76, 348.42:46.59, 401.31:77.45, 419.37:126.65, 504.54:18.29, 511.44:31.38, 532.39:68.72, 545.39:35.93, 582.48:28.81, 612.78:16.21, 639.19:20.81, 658.64:34.99, 756.22:44.1, 826.59:48.87, 830.29:52.57, 852.55:35.76, 869.77:266.46, 959:84.08, 1055.65:89.3, 1184.49:58.59, 1297.59:41.16, 1310.57:31.71, 1350.97:26.92, 1362.64:16.21, 1411.44:41.51, 1483.23:49.73, 1541.58:18.44, 1577.92:20.73, 1583.63:47.37, 1603.26:18.32, 1686.8:84.36
R4f	2178.9	429.92:10.07, 500.29:12.85, 525.64:8.23, 906.34:29.73, 1099.41:72.84, 1716.68:61.26, 1749.9:93.42, 1859.39:67.41, 1953.79:33.86, 1968.08:58.71, 2090.51:59.33, 2121.58:27.34
B4b	1492.5	72.82:3.1, 112.34:15.08, 121:4.41, 175.64:58.59, 218.05:5.96, 235.98:8.33, 246.45:4.75, 264.28:24.6, 288.86:6.5, 306.95:23.38, 322.55:48.71, 347.05:22.71, 362.03:17.12, 378.26:30.76, 396.33:38.83, 404.35:76.88, 422.48:99.59, 480.79:39.88, 496.16:35.49, 524.49:36, 542.2:117.95, 551.78:74.03, 569.04:74, 584.89:35.42, 610.5:24.41,

		627.38:31.99, 654.54:59.59, 665.34:62.08, 682.67:59.77, 698.72:35.56, 724.87:181.59, 742.45:145.76, 752.61:53.84, 795.59:94.76, 813.22:121.57, 925.86:247.65, 982.46:158.44, 1073.83:117.05, 1178.81:280.76, 1232.6:330.73, 1315.73:279.62, 1384.59:345.27
B4c	1607.6	112.73:7.61, 136.61:33.27, 175.77:94.78, 201.63:9.81, 258.88:11.9, 275.45:15.94, 291.72:30.28, 309.48:10.65, 365.23:21.02, 385.66:25.65, 403.49:43.11, 405.47:40.23, 421.58:34.89, 423.06:35.59, 491.14:42.57, 508.88:28.2, 532.52:22.42, 549.65:40.72, 566.67:42.87, 657.07:183.55, 767.27:55.9, 786.5:55.55, 823.61:216.65, 844.04:54.98, 892.74:119.43, 914.27:88.67, 920.54:132.48, 938:66.66, 996.41:81.48, 1008.15:103.99, 1028.62:123.2, 1100.93:78.03, 1245.56:190.6, 1370.11:373.65, 1407.78:210.63, 1470.46:146.77
B4d	1828.8	175.27:15.1, 248.12:3.19, 326.35:4.75, 349.22:16.39, 402.57:47.65, 420.31:87.77, 534.44:41.43, 546.93:12.72, 583.85:7.25, 612.94:7.53, 641.35:12.09, 659.58:17.88, 692.68:6.63, 713.71:4.61, 736.06:9.01, 738.1:8.51, 757.16:29.11, 780.66:9.37, 793.57:11.84, 805.95:18.62, 822.17:23.67, 829.58:55.57, 852.38:12.06, 871.55:205.97, 909.92:32.17, 961.29:50.61, 1056.88:48.04, 1680.06:123.88, 1689.37:158.15, 1715.74:33.71, 1786.76:309.47
B4e	1844.8	85.7:1.11, 129.38:3.36, 175.48:28.99, 212.05:0.76, 247.48:3.08, 286.15:11.56, 325.22:6.01, 348.56:49.1, 359.19:2.76, 401.28:90.99, 418.79:190.26, 504.49:10.15, 509.81:3.68, 531.58:135.93, 543.3:125.47, 561.56:1.51, 598.09:6.68, 638.67:7.37, 656.1:6.52, 674.4:14.35, 712.51:3.01, 754.46:3.15, 771.16:19.36, 830.76:27.58, 842.17:5.77, 867.41:71.89, 870.89:68.45, 885.27:325.68, 932.9:3.48, 959.54:75.67, 1053.47:22.68, 1071.9:28.46, 1169.5:14.21, 1183.56:16.25, 1199.47:30.76, 1298.16:43.77, 1312.4:19.17, 1343.44:28.69, 1496.31:45.17, 1551.55:13.65, 1567.74:11.01, 1574.72:15.11, 1583.04:9.89, 1650.17:23.26, 1672.21:24.67, 1695.75:299.09, 1707.92:14.97, 1800.45:617.76
B4f	2178.9	428.86:66.11, 540.55:50.32, 586.52:51.9, 1001.54:92.67, 1083.58:233.45, 1100.27:422.85, 1331.52:460.65, 1367.73:417.63, 1529.13:1237.28, 1705.83:129.52, 1750.5:239.89, 1749.68:175.39, 1835.27:575.29, 1902.67:664.76, 2061.23:117.12, 2088.6:228.95

Appendix - Table 7: MH⁺ and m/z lists from PSD MS/MS experiments of protein spot P4

peptide	MH ⁺ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
P4a	1389.54	270.53:3.43, 326.58:11.68, 381.68:39.03, 426.95:16.25, 463.7:12.73, 495.61:29.05, 648.52:104.61, 665.53:35.67, 707.36:58.11, 764.54:50.16, 782.55:42.21, 801.66:82.26, 951.95:385.64, 1119.46:186.81, 1147.05:173.87, 1225.51:69.96,

		1240.86:107.83, 1295.23:74.39, 1327.22:217.64, 1338.02:77.01
P4b	1464.44	255.53:19.99, 272.44:25.03, 362.51:3.52, 379.79:7.05, 415.49:16.77, 432.66:2.5, 474.24:0.94, 491.56:2.69, 497.26:1.24, 508.79:1.36, 514.34:18.06, 542.86:10.5, 552.45:0.79, 560.58:14.5, 598.08:2.89, 618.51:2.25, 629.9:1.91, 647.58:267.1, 674.41:3.45, 694.49:2.28, 715.65:4.2, 744.37:14.08, 762.41:28.99, 817.19:7.83, 844.01:5.57, 860.26:5.29, 874.62:5.78, 877.11:8.42, 887.6:13.76, 903.93:4.21, 921.64:38.75, 993.89:11.95, 1015.79:10.17, 1043.78:11.16, 1127.35:19.12, 1142.4:37.44, 1166.84:16.56, 1248.93:13.69, 1301.39:14.09, 1325.83:7.88, 1373.7:28.89

Appendix - Table 8: MH^+ and m/z lists from PSD MS/MS experiments of protein spots R5 und B5

peptide	MH^+ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
R5a	1456.5	70.28:0.29, 185.27:2.9, 212:1.05, 242.35:0.54, 256.29:4.36, 266.27:16.05, 284.14:25.89, 301.84:9.54, 318.61:0.31, 335.46:0.65, 343.43:1.98, 353.83:41.53, 371.54:70.21, 387.52:7.18, 399.09:9.49, 405.66:42.03, 416.08:25.94, 424.89:42.7, 452.4:113.35, 468.61:37.34, 474.91:5.4, 484.26:11.66, 487.98:1.43, 497.69:5.05, 502.38:29.83, 515.7:13.05, 539.13:52.47, 553.46:420.14, 565.33:12.45, 570.5:40.92, 575.73:9.99, 584.9:47.54, 600.44:3.68, 622.19:8.76, 637.2:85.18, 640.66:95.09, 644.92:13.37, 655.03:16.75, 663.04:150.32, 681.71:2906.8, 699.92:23.07, 724.32:49.25, 748.69:95.83, 757.86:108.41, 768.83:440.07, 775.91:108.47, 822.73:85.29, 855.71:77.45, 873.26:249.89, 923.46:133.97, 953.95:35.2, 969.22:116.07, 1009.26:52.7, 1012.34:43.76, 1023.42:60, 1041.13:179.77, 1052.31:409.58, 1139.53:84.95, 1196.21:191.64, 1227.56:116.04, 1310.5:115.19, 1327.61:321.4, 1342.96:115.17, 1357.96:23.34, 1395.55:110
R5b	1504.0	353.57:1.26, 371.42:11.29, 387.52:0.03, 405.15:6.15, 424.65:7.71, 452.59:18.27, 464.29:1.54, 469.06:2.81, 502.2:0.84, 503.28:0.6, 539.42:2.72, 548.76:8.54, 554.82:4.11, 565.52:0.76, 572.02:0.89, 575.81:0.68, 583.9:25.89, 601.31:2.23, 626.22:1.88, 632.48:31.31, 644.68:1.7, 662.5:9.93, 675.67:120.23, 682.75:14.62, 688.17:7.81, 695.39:1.14, 711.86:19.98, 729.74:1265.02, 740.43:29.61, 747.83:10.97, 757.7:49.59, 776:71.13, 816.93:271.45, 855.19:34.79, 873.14:61.37, 969.77:43.6, 1003.25:23.14, 1041.5:110.49, 1100.09:337.28, 1187.15:67.13, 1227.02:119.61, 1244.21:121.35, 1262.61:42.56, 1345.77:44.31, 1358.61:107.24, 1392.11:33.17, 1432:55.72, 1449.13:492.84, 1454.7:0.13
B5a	1456.5	452.31:5.04, 553.92:19.14, 590.8:6.14, 681.55:139.06, 768.53:47.03, 913.3:41.82, 922.23:42.16, 1053.16:45.41, 1256.82:46.87, 1360.91:19.22

Appendix - Table 9: MH⁺ and m/z lists from PSD MS/MS experiments of protein spot P5

peptide	MH ⁺ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
P5a	1004.06	296.41:17.35, 331.42:5.95, 391.91:14.95, 409.63:25.41, 432.73:11.13, 519.59:33.07, 544.27:25.33, 572.45:58.12, 595.98:30.98, 609.27:17, 622.51:55.74, 655.8:83.32, 663.54:26.45, 673.59:90.38, 708.56:262.98, 725.79:355.08, 745.25:231.55, 763.86:168.65, 807.33:306.21, 849.77:128.55, 863.67:209.67, 922.74:238.31, 940.82:338.99, 959.53:304.89
P5b	1303.42	175.43:20.82, 256.29:42.95, 303.44:5.41, 355.58:62.57, 400.34:11.33, 454.55:13.12, 457.27:8.74, 474.46:10.11, 570.71:17.08, 587.57:190.52, 685.37:21.6, 702.58:21.15, 799.91:14.94, 810.02:20.39, 849.55:59.8, 872.85:86.81, 913.78:159.48, 952.97:149.97, 1061.97:428.45, 1080.48:215.61, 1102.17:170.86, 1146.81:164.01, 1162.38:169.59, 1182.78:57.75, 1205.45:22.37
P5c	1425.32	449.79:4.28, 501.65:3.87, 558.56:5.7, 612.62:35.14, 726.8:37.73, 736.92:14.91, 775.49:91.05, 807.73:31.77, 862.82:90.04, 902.11:91.65, 990.59:127.51, 1035.98:106.3, 1062.18:132.14, 1081.64:110.93, 1098.79:243.38, 1120.87:414.05, 1132.74:261.72, 1151.13:894.95, 1196.94:117.35, 1233.85:59.37, 1236.41:36.84, 1243.9:41.14, 1246.43:70.36, 1264.17:588.09, 1279.49:198.17, 1291.65:92.23, 1310.06:781.92, 1334.68:368.75, 1363.63:137.08

Appendix - Table 10: MH⁺ and m/z lists from PSD MS/MS experiments of protein spot R6

peptide	MH ⁺ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
R6a	1648.53	215.77:0.89, 299.23:0.23, 370.04:0.13, 388.23:0.55, 397.09:0.66, 403.69:0.64, 416.19:0.95, 462.67:1.37, 532.76:2.99, 573.48:8.9, 588.45:15.12, 608.87:10.42, 624.43:4.68, 661.75:5.65, 689.92:7.3, 701.3:8.99, 714.46:16.09, 727.16:30.29, 744.1:126.78, 860.52:50.08, 888.97:22.8, 918.32:46.61, 934.95:322.06, 1005.67:160.36, 1034.38:120.37, 1165.51:81.33, 1254.55:78.37, 1290.99:125.15, 1317.48:66.23, 1331.93:41.44, 1350.22:190.43, 1419.1:66.06, 1430.82:43.45, 1456.97:68.26, 1463.82:54.68, 1474.78:27.83, 1511.31:67.23, 1548.18:71.65, 1575.13:247.76
R6b	2140.80	455.58:3.62, 471.85:1.64, 600.10:3.06, 687.64:43.27, 740.16:0.97, 783.75:7.96, 880.82:47.59, 899.13:40.79, 914.44:1.39, 946.20:3.54, 955.23:3.38, 981.81:10.07, 999.89:19.07, 1040.89:8.86, 1085.62:9.49, 1101.46:12.63, 1144.41:15.15, 1155.13:10.69, 1182.83:8.48, 1197.60:26.47, 1361.86:29.60, 1400.57:16.58, 1417.37:43.00, 1520.03:44.52, 1615.50:29.94, 1816.14:68.28, 1945.51:140.18, 2050.90:112.08

Appendix - Table 11: MH⁺ and m/z lists from PSD MS/MS experiments of protein spots R8, B8 and P8

peptide	MH ⁺ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
R8a	1414.3	274.76:2.11, 303.99:33.35, 489.81:23.96, 572.44:2.62, 578.12:3.08, 618.51:10.29, 700.6:19.97, 717.66:13.77, 729.14:13.18, 735.45:8.42, 766.53:8.77, 800.49:30.66, 841.8:25.35, 844.93:24.58, 973.63:62.39, 979.04:24.65, 1030.65:104.98, 1056.89:51.21, 1115.08:81.82, 1142.63:62.57, 1150.3:96.21, 1207.91:103.59, 1218.49:107.54, 1249.9:86.92, 1284.3:23.43, 1299.22:21.7, 1353.78:28.61
R8b	1565.4	114.35:6.12, 137.87:0.33, 159.67:0.26, 176.8:21.67, 219.03:1.7, 247.16:0.26, 263.67:0.23, 299.35:1.01, 300.27:2.4, 317.81:12.39, 334.8:125.49, 367.71:1.99, 386.93:12.44, 396.04:2.69, 404.64:9.98, 432.55:4.64, 438.86:4.51, 449.81:119.13, 508.54:2.81, 523.08:9.85, 548.45:21.43, 553.06:5.49, 564.67:13.43, 656.91:6.44, 710.72:7.74, 727.87:11.83, 749.37:6.48, 762.52:6.17, 767.01:14.87, 785.19:50.06, 796.56:10.55, 840.84:41.03, 849.62:25.89, 867.34:93, 885.67:334.1, 983.55:60.63, 994.78:42.56, 999.85:64.58, 1057.34:32.91, 1136.06:127.09, 1162.37:104.11, 1250.6:81.69, 1348.79:69.33, 1390.61:201.43, 1406.36:91.92, 1433.33:34.22, 1450.53:207.08, 1503.98:123.85
R8c	1750.7	177.16:3.41, 304.16:0.42, 306.01:0.01, 403.98:0.07, 561.02:20.01, 675.94:1.19, 803.84:115.53, 900.59:9.7, 918.36:9.97, 927.49:4.9, 983.14:3.84, 989.26:8.73, 991.59:7.75, 1006.34:10.15, 1050.41:6.85, 1063.67:14.34, 1080.81:4.87, 1091.39:4.77, 1154.44:9.68, 1199.82:9.36, 1259.06:17.56, 1292.7:12.05, 1301.77:9.16, 1357.31:25.36, 1416.14:7.24, 1539.91:67.98, 1567.6:90.9, 1580:19.36, 1667.8:65.02
B8a	1414.3	257.06:4.76, 273.75:32.2, 302.83:88.27, 488.34:88.04, 571.61:25.04, 616.48:67.69, 658.98:29.69, 699.42:53.54, 717.3:101.66, 800.15:56.18, 903.71:73.02, 1030.2:137.55, 1090.2:89.41, 1115.01:95.65, 1249.16:366.5, 1291.68:31.23
B8b	1565.4	112.14:22.52, 136.53:3.9, 175.51:52.96, 217.51:8.79, 245.3:8.49, 262.38:4.42, 298.61:13.92, 316.5:20.61, 333.52:161.17, 343.1:4.69, 366.95:5.8, 385.55:21.28, 403.69:30.86, 431.07:22.02, 448.45:167.7, 465.81:14.86, 504.76:14.32, 521.78:39.61, 547.87:76.24, 563.1:76.78, 590.5:13.54, 622.36:42.51, 709.39:69.75, 741.51:54.65, 749.05:113.26, 765.19:78.9, 783.28:108.9, 844.28:153.19, 867.85:130.03, 884.95:514.42, 945.69:164.64, 982.56:178.37, 1000.03:229.06, 1137.19:345.46, 1153.57:107.7, 1252.91:115.86, 1421.06:205.07, 1443.4:154.99, 1451.26:459.56, 1482.58:251.25, 1503.98:447.95
B8c	1750.7	112.41:2.75, 175.43:88.98, 200.81:2.95, 260.25:4.95,

		284.77:24.73, 302.47:94.21, 372.6:8.75, 383.91:49.66, 402.03:64.41, 420.21:9.6, 448.4:7.81, 484.89:14.63, 501.03:29.12, 559.91:127.02, 588.95:42.61, 656.38:75.86, 673.68:45.45, 732.49:37.06, 758.52:82.34, 780.48:90.69, 802.19:463.19, 835.72:92.89, 874.26:211.69, 917.47:139.53, 967.31:178.96, 1005.85:222.4, 1017.56:134.86, 1177.87:603.82, 1251.48:274.08, 1290.33:444.27, 1406.08:473.82, 1450.12:199.35, 1476.81:174.61, 1521.45:495.57, 1536.7:590, 1568.39:1287.13, 1697.35:275.95
P8b	1565.4	72.82:3.66, 113.68:7.94, 138.09:4.16, 159.66:7.84, 176.87:46.39, 219.03:5.72, 299.86:21.17, 318.08:37.67, 334.85:148.17, 386.4:44.2, 404.8:44.32, 449.79:160.85, 466.7:27.42, 507.29:27.41, 548.57:55.08, 564.9:42.55, 667.01:32.02, 767.06:66.13, 784.53:102.64, 868.88:50.94, 885.46:347.92, 962.33:271.78, 999.78:378.72, 1074.71:231.41, 1135.8:241.06, 1250.02:412.99, 1376.25:306.76, 1406.22:108, 1420.69:204.37, 1450.27:249.39, 1503.67:171.27
P8c	1750.7	177.08:5.7, 276.98:4.15, 303.26:4.66, 430.03:14.1, 515.97:27.44, 528.16:12.85, 591.77:16.16, 687.77:20.55, 709.76:46.49, 778.48:72.26, 803.24:100.81, 950.06:218.18, 1005.31:218.23, 1032.13:207.65, 1055.39:302.54, 1069.82:213.94, 1089.59:116.68, 1110.49:142.89, 1120.91:306.08, 1151.12:352.3, 1203.1:537.31, 1407.71:219.26, 1444.31:440.25, 1506.53:427.04, 1534.39:551.78, 1601.61:612.52, 1688.34:55.83

Appendix - Table 12: MH⁺ and m/z lists from PSD MS/MS experiments of protein spot R9

peptide	MH ⁺ (av.)	fragment mass ₁ :intensity ₁ , fragment mass ₂ :intensity ₂ , ...
R9a	2876.54	642.74:9.85, 685.21:5.29, 991.29:6.6, 1199.56:8.61, 1230.22:20.53, 1335.68:7.04, 1344.22:10.79, 1352.84:6.87, 1361.43:13.55, 1367.26:58.03, 1443.44:23.19, 1468.28:72.89, 1562.84:21.54, 1580.89:51.62, 1649.57:24.44, 1680.01:32.98, 1723.22:25.14, 1748.42:50.44, 1766.73:79.28, 1880.75:51.63, 1971.33:36.36, 2025.05:71.82, 2200.39:108.78, 2252.74:74.08, 2401.28:80.36, 2550.1:29.73, 2678.24:74.44, 2697.91:28.6, 2706.63:19.68, 2726.71:82.83, 2747.82:127.36, 2767.65:66.89

Epl1, the major secreted protein of *Hypocrea atroviridis* on glucose, is a member of a strongly conserved protein family comprising plant defense response elicitors

Verena Seidl¹, Martina Marchetti², Reingard Schandl^{1,2}, Günter Allmaier² and Christian P. Kubicek¹

¹ Research Area Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, Vienna University of Technology, Austria

² Institute of Chemical Technologies and Analytics, Vienna University of Technology, Austria

Keywords

cerato-platanin; elicitor; *Hypocrea* (*Trichoderma*); plant defense responses

Correspondence

V. Seidl, Research Area Gene Technology and Applied Biochemistry, Institute of Chemical Engineering, Vienna University of Technology, Getreidemarkt 9/166-5, A-1060 Vienna, Austria
Fax: +43 1 58801 17299
Tel: +43 1 58801 17227
E-mail: vseidl@mail.zserv.tuwien.ac.at/
Website: <http://www.vt.tuwien.ac.at/>

(Received 30 March 2006, revised 25 July 2006, accepted 27 July 2006)

doi:10.1111/j.1742-4658.2006.05435.x

We used a proteomic approach to identify constitutively formed extracellular proteins of *Hypocrea atroviridis* (*Trichoderma atroviride*), a known biocontrol agent. The fungus was cultivated on glucose and the secretome was examined by two-dimensional gel electrophoresis. The two predominant spots were identified by MALDI MS utilizing peptide mass fingerprints and amino acid sequence tags obtained by postsource decay and/or high-energy collision-induced dissociation (MS/MS) experiments, and turned out to be the same protein (12 629 Da as determined with MS, pI 5.5–5.7), probably representing the monomer and the dimer. The corresponding gene was subsequently cloned from *H. atroviridis* and named *epl1* (eliciting plant response-like), because it encodes a protein that exhibits high similarity to the cerato-platanin family, which comprises proteins such as cerato-platanin from *Ceratocystis fimbriata* f. sp. *platani* and Snodprot1 of *Phaeosphaeria nodorum*, which have been reported to be involved in plant pathogenesis and elicitation of plant defense responses. Additionally, based on the similarity of the N-terminus to that of *H. atroviridis* Epl1, we conclude that a previously identified 18 kDa plant response elicitor isolated from *T. virens* is an ortholog of *epl1*. Our results showed that *epl1* transcript was present under all growth conditions tested, which included the carbon sources glucose, glycerol, L-arabinose, D-xylose, colloidal chitin and cell walls of the plant pathogen *Rhizoctonia solani*, and also plate confrontation assays with *R. solani*. *epl1* transcript could even be detected under osmotic stress, and carbon and nitrogen starvation.

Fungi belonging to the genus *Hypocrea/Trichoderma* are highly interactive in root, soil and foliar environments; they compete with other soil microorganisms for nutrients, produce antibiotic substances, and parasitize other fungi. In addition, they have recently been shown to be able to enhance root and plant growth and to induce systemic and localized resistance in plants [1–4]. The latter property may be crucially important for agricultural uses and for understanding

the roles of *Hypocrea/Trichoderma* in natural and managed ecosystems.

The ability of *Trichoderma* spp. to induce local and systemic resistance has been shown with *Hypocrea lixii* (*Trichoderma harzianum*) in agricultural crops such as bean, cotton, tobacco, lettuce, tomato and maize [5–9], with *T. asperellum* in cucumber [10–12], and with *H. virens* (*T. virens*) in cotton [13]. However, little is known about the elicitors of this response. Harman

Abbreviations

CID, collision-induced dissociation; 2D-GE, two-dimensional gel electrophoresis; Epl1, eliciting plant response-like protein 1; EST, expressed sequence tag; GRAVY, grand average of hydropathicity; IT, ion trap; PMF, peptide mass fingerprint; PSD, postsource decay; UTR, untranslated region.

et al. [2] defined three different classes of compound that are produced by *Hypocrea/Trichoderma* and induce resistance in plants: proteins with enzymatic functions, avirulence proteins, and oligosaccharides and low-molecular-weight compounds released from fungal or plant cell walls by hydrolytic enzymes. Despite increasing knowledge about the ability of *Hypocrea/Trichoderma* spp. to induce defense responses in a variety of plants, the molecular basis of this mechanism is still unclear and the number of identified elicitors remains low. So far, there is only published evidence for three proteins that are able to induce resistance. Two of them are enzymes, namely a 22 kDa xylanase of *T. viride*, which induces ethylene synthesis and pathogenesis-related protein production in tobacco leaves [14,15] and a 54 kDa cellulase of *T. longibrachiatum*, which induces various defense mechanisms in melon cotyledons [16]. The third elicitor is an 18 kDa protein secreted by *H. virens*, which is able to induce systemic resistance in cotton seedlings and was putatively identified as a serine protease through the similarity of its N-terminal sequence to that of a serine proteinase from *Fusarium sporotrichioides* [8]. To our knowledge, no other plant defense response elicitors from *Hypocrea/Trichoderma* have been characterized to date.

In this work, we investigated the secretome of the biocontrol strain *H. atroviridis* P1 (*T. atroviride*) in order to identify constitutively expressed proteins. We used a proteomics approach including two-dimensional gel electrophoresis (2D-GE), peptide mass fingerprinting and MS-generated sequence tags. Interestingly, the major protein found is a member of the recently identified cerato-platanin protein family, which contains proteins from plant pathogenic fungi that have been demonstrated to act as elicitors of plant defense responses and as virulence factors. The *H. jecorina* and *H. atroviridis* orthologs have an almost identical processed N-terminus as the above cited 18 kDa elicitor from *H. virens*, which we therefore also believe to be a member of this family. In this study, the *H. atroviridis* protein was characterized in detail, its expression pattern under growth on various carbon sources and other cultivation conditions was investigated, and its phylogenetic relationship to other proteins of the cerato-platanin family was analyzed.

Results

Analysis of the secretome of *H. atroviridis* during cultivation on glucose

Hypocrea atroviridis was grown on glucose, and the culture supernatant was harvested during the phase of

fast growth (after 20 h). A 2D-GE analysis of proteins secreted under these conditions is shown in Fig. 1. Only a small number of proteins was detected, and by far the most abundant spot (g1 in Fig. 1) was a small protein (approximately 16 kDa, pI 5.5–5.7), and this was followed by spot g2, with a similar pI but with a molecular mass of approximately 27 kDa. Comparison of the *H. atroviridis* secretome under a number of other cultivation conditions, such as growth on colloidal chitin, under nitrogen starvation, or on cell walls of several plant pathogenic fungi (*Rhizoctonia solani*, *Botrytis cinerea* and *Pythium ultimum*), revealed a much higher number of secreted proteins in 2D-GE. This can be explained by the fact that glucose is directly taken up by the fungus, but for growth on more complex carbon sources, such as fungal cell walls, *H. atroviridis* needs to produce several different extracellular enzymes to hydrolyze the corresponding substrates. However, in the area of 15–20 kDa and pI 5.2–6.2, only one protein, at exactly the same location as g1, was present, as can be seen in the respective sections of those 2D gels in Fig. 1. Results from

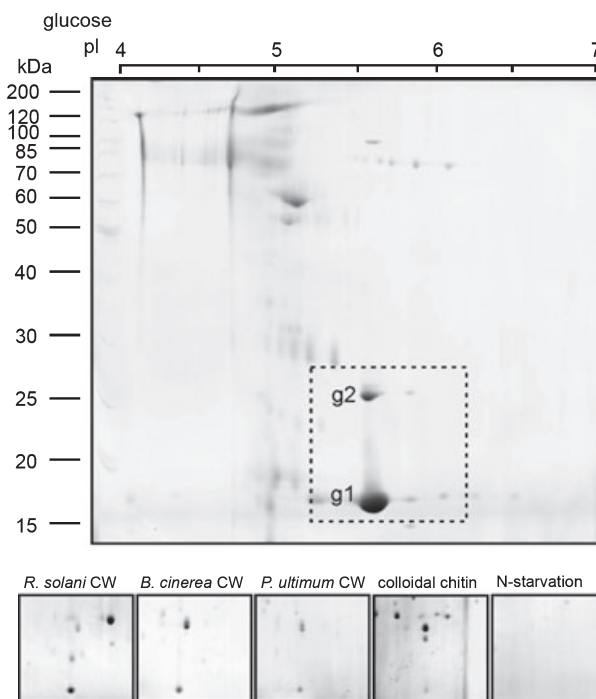


Fig. 1. Two-dimensional gel electrophoresis (2D-GE) of extracellular proteins of *Hypocrea atroviridis*. The large picture shows a representative 2D gel of culture filtrates from glucose cultivations. The region containing the two largest spots (g1 and g2) is framed with a dashed line, and the respective sections of 2D gels from cultures grown on *Rhizoclonia solani*, *Botrytis cinerea* and *Pythium ultimum* cell walls (CW), colloidal chitin and under nitrogen starvation are shown below.

2D-GE thus implied that the 16 kDa/pI 5.5–5.7 protein (g1), abundantly present in cultures grown on glucose as carbon source, was also secreted during growth on *R. solani*, *Botrytis cinerea* and *Pythium ultimum* cell walls and on colloidal chitin, but was absent during growth under nitrogen limitation. The identity of these protein spots in the 2D gels for different growth conditions was confirmed by MALDI-RTOF MS analysis of the corresponding spots. The peptide mass fingerprinting and postsource decay (PSD) experiments with the most prominent tryptic peptide of spot g1 (see below) from the glucose cultivations gave the same results as for the protein spots from other growth conditions (data not shown).

Identification of the two major components of the secretome on glucose via cross-species identification

For protein identification of spots g1 and g2, the spots were cut out of the gels and digested with trypsin,

and the resulting extracted peptides were analysed by MALDI-RTOF MS. Interestingly, the peptide mass fingerprints (PMFs) of g1 and g2 did not differ significantly, as shown in Fig. 2A,B, except for the peptides at m/z 1429.73, 1445.73, 2558.56 and 2574.57, respectively. They were only found in the PMF of g1 and represented two oxidized forms each ($[M + H + 16]^+$ and $[M + H + 32]^+$). Although the information content of the PMF based on the number of detected peptides was high with respect to the size of the protein (five detected peptides out of seven theoretical peptides), a search of the databases with corresponding mass lists gave no significant protein hit for g1 and g2.

For protein identification within spot g1, PSD and high-energy collision-induced dissociation (CID) MS/MS experiments with six prominent peptides (m/z 1413.72 (P1), 1429.73 (P1a), 1445.73 (P1b), 1564.69 (P2), 1749.95 (P3), 2542.48 (P4); Table 1) were performed. The peptide P1 (Fig. 2c) matched well but not significantly enough with the theoretical ion values of a tryptic peptide of EST L12T11P105R09908

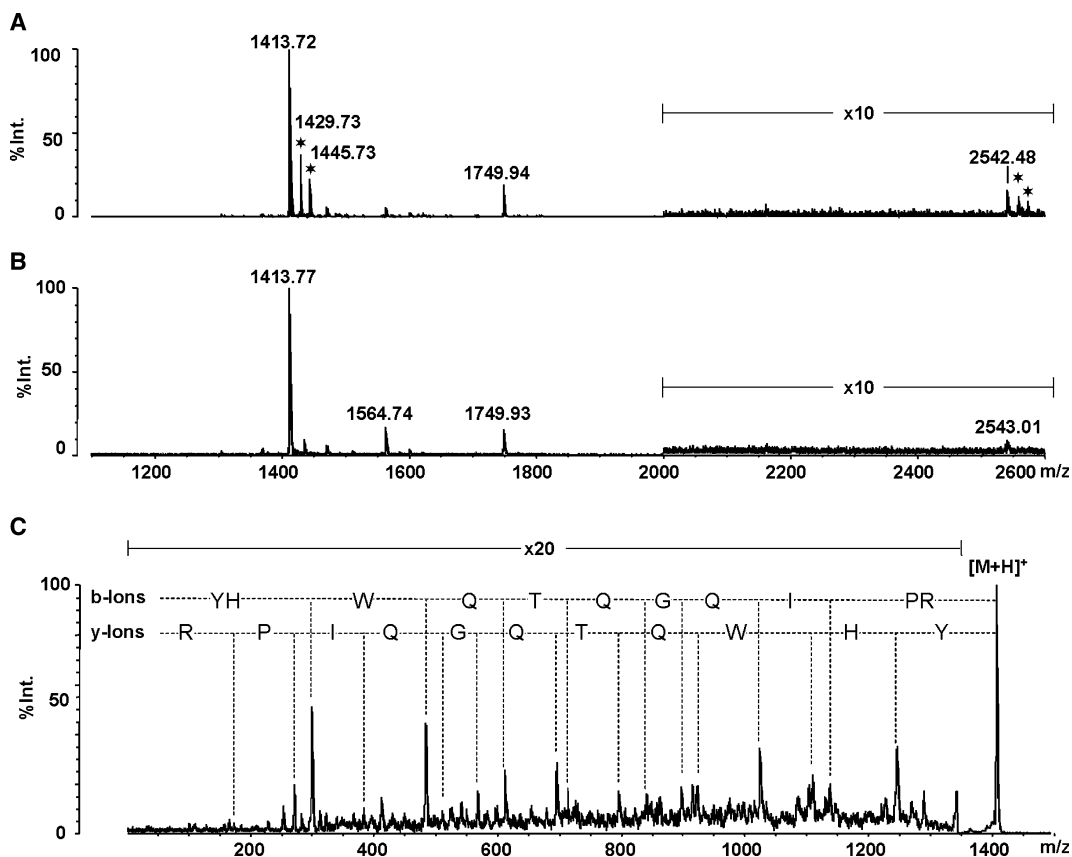


Fig. 2. (A) Positive ion peptide mass fingerprint (PMF) of gel spot g1 (16 kDa/pI 5.5–5.7) by MALDI reflectron MS. Two particular peptides were mono-oxidized and di-oxidized (indicated by asterisks). (B) PMF of gel spot g2 (27 kDa/pI 5.5–5.7). (C) Positive ion postsource decay (PSD) spectrum of peptide P1 (precursor ion at m/z 1413.72; deduced sequence YHWQTQGQIPR).

Table 1. Identified peptides and sequence tags of spots g1 and g2 and matching EST sequences in the TrichoEST database, identified with the MASCOT search engine.

Spot	Selected precursor ion [M + H] ⁺ _{monoisot.} ([M + H] ⁺ _{calculated})	Peptide sequence	MASCOT ion score	Match to
g1				
P1	1413.72 (1413.70)	YHWQTQGQIPR	(34) ^a	L14T53P106R00046 L12T11P105R09908
P1a	1429.73 (1429.70)	YHWQTQGQIPR + 1 Ox (HW)	49	L14T53P106R00046 L12T11P105R09908
P1b	1445.73 (1491.71)	YHWQTQGQIPR + 2 Ox (HW)	49	L14T53P106R00046 L12T11P105R09908
P2	1564.69 (1564.64)	DTVSYDTGYDDASR	124	L14T53P106R00046 L12T11P105R09908
P3	1749.95 (1749.88)	SLTVWSCSDGANGLITR	50	L14T53P106R00046 L12T11P105R09908
P4	2542.48 (2542.16)	FPYIGGVQAVAGWNSPCGTCWK	Not identified by MASCOT	L14T53P106R00046 L12T11P105R09908
P5	1491.70 (1491.71)	<i>m/z</i> value fits to theoretical value of tryptic peptide		
g2				
P6	1413.77 (1413.70)	YHWQTQGQIPR	(34) ^a	L14T53P106R00046 L12T11P105R09908
P7	1564.74 (1564.64)	DTVSYDTGYDDASR	124	L14T53P106R00046 L12T11P105R09908
P8	1749.93 (1749.88)	SLTVWSCSDGANGLITR	50	L14T53P106R00046 L12T11P105R09908
P9	2542.31 (2542.16)	<i>m/z</i> value fits to theoretical value of tryptic peptide		

^a Below significant threshold ($P = 0.05$).

(DDBJ/EMBL/GenBank accession number AJ901879) of *H. atroviridis* 11 (IMI 352941 [17]), and EST L14T53P106R00046 (DDBJ/EMBL/GenBank accession number AJ902344) of *T. asperellum* of the TrichoEST database (<http://www.trichoderma.org>). Considering up to two oxidations on tryptophan and/or histidine increased the MASCOT ion scores above the threshold (significant threshold $P < 0.05$), resulting in significant hits for the two peptides P1a and P1b, respectively. The mono-oxidation was clearly located at the tryptophan, generating hydroxytryptophan, as determined by high-energy CID experiments. The location of the second oxidation could not be clearly elucidated, but localization on the already mono-oxidized tryptophan, giving *N*-formylkynurenine, was more likely than one on the less reactive histidine. Results of PSD experiments with the peptide P3 were again in good agreement (MASCOT ion score 50) with the database entries of expressed sequence tags (ESTs) L12T11P105R09908 and L14T53P106R00046, but clearly showed the substitution T → A (Fig. 3A).

The PSD spectrum of peptide P4 did not give a reliable MASCOT search result, but by manual interpretation of the acquired spectrum, a partial sequence tag (PYIGGVQAVAGWNSP) was obtained, which fitted to a calculated tryptic peptide (FPYIGGVQAVAGWNSPCGTCWK) of the sequence of the respective *H. atroviridis* protein, as deduced from the respective DNA sequences (see below), but comprised two amino acid changes (A → V, N → S) in comparison to the previously identified ESTs. The two signals representing two oxidized forms (m/z 2558.56 [M + H + 16]⁺ and m/z 2574.57 [M + H + 32]⁺) that were detected in the PMF could be explained by a double oxidation on either of the two tryptophans present in this sequence. The PSD mass spectrum of peptide P2 was identified as DTVSYDTGYDDASR by omitting enzymatic cleavage of the database entries (MASCOT ion score 124) in the same ESTs.

For protein identification of gel spot g2, which showed, as mentioned above, a similar PMF (Fig. 2B) except for the two double-oxidized tryptophans, three

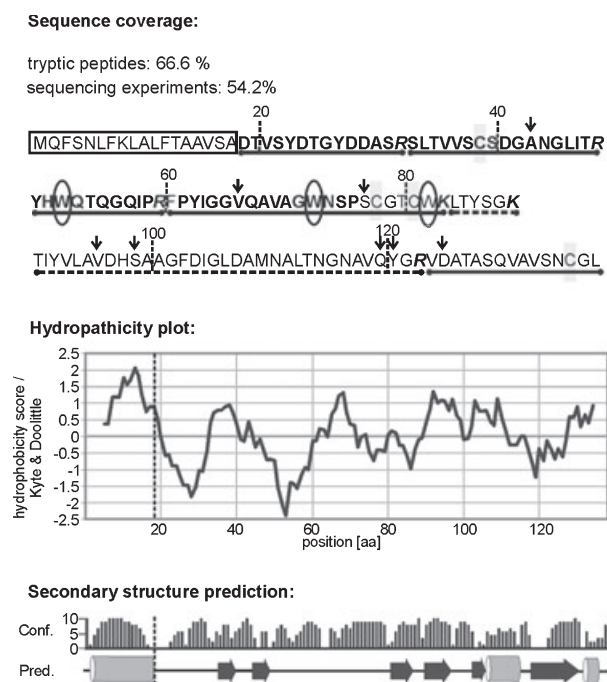


Fig. 3. (A) Sequence coverage by MS experiments of *Hypocrea atroviridis* Epl1. The signal peptide is marked with a box, the tryptic peptides are underlined and the respective basic amino acid residues, R and K, are indicated in italics. A solid line indicates peptides that were positively identified by MS; peptides that were not found are marked with a dashed line. Amino acids covered by sequencing experiments are highlighted in bold, amino acids that were found to be exchanged in comparison to the EST sequences L14T53P106R00046 and L12T11P105R09908 are marked with an arrow, oxidized tryptophans are encircled, and the four conserved cysteines of the cerato-platanin protein family are indicated by a gray box. (B) Hydropathicity plot (Kyte & Doolittle). The vertical dashed line shows the signal peptide-cleavage site. (C) Secondary structure prediction of Epl1 with PSIPRED. Gray barrels represent helices, broad, black arrows indicate strands, and the black line indicates coiled, unstructured regions. The bars at the location of the corresponding amino acids indicate the confidence of the secondary structure prediction. The vertical dashed line shows the signal peptide-cleavage site.

peptides were chosen for sequencing experiments (P6, P7, and P8). All of these peptides showed the same MASCOT ion score as spot g1 for the identified amino acid sequences (Table 1), indicating that these gel spots represent the same protein. Taken together, spots g1 and g2 could be clearly identified, with a sequence coverage of 66.6% by tryptic peptides and 54.2% by sequencing experiments, as the *H. atroviridis* homologs of EST L12T11P105R09908 (*H. atroviridis* 11) and EST L14T53P106R00046 (*T. asperellum*). The two peptides that were not detected by peptide mass fingerprinting were either too small (calculated monoisotopic $[M + H]^+$ ion m/z 668.36) to be clearly differentiated

from matrix background ions, or too large (calculated monoisotopic $[M + H]^+$ ion m/z 3536.76) to be detected at a reliable signal-to-noise ratio by MALDI-TOF MS. With the presence of two tryptophans in a double-oxidized form in spot g1 as the only difference in the MS spectra, spots g1 and g2 possibly represented the monomer and dimer of the same protein.

The matching EST sequences were used for a TBLASTX search of the genome database of *H. jecorina* (*T. reesei*; <http://gsphere.lanl.gov/trire1/trire1.home.html>), which is so far the only *Hypocrea/Trichoderma* species for which the whole genome sequence is available. We identified three different ORFs, among which tre46514 encodes the protein with highest similarity to the EST sequences from *H. lixii* and *T. asperellum* mentioned above. A number of additional EST sequences from other *Hypocrea/Trichoderma* spp. (Fig. 4) could consequently be identified in the TrichoEST database by conducting further TBLASTX searches. Interestingly, the N-termini of the mature *Hypocrea/Trichoderma* proteins (after cleavage of the signal peptide as predicted with SIGNALP [18]) showed strong similarity (15 of 19 amino acids) to the N-terminal sequence of a plant response elicitor from *H. virens* [8]. Because the size of this protein (18 kDa in SDS/PAGE) is comparable to that of the protein identified in this study (16 kDa in SDS/PAGE), we concluded that this elicitor is a homolog of the protein identified from *H. atroviridis* in this study, which we therefore named Epl1 (eliciting plant response-like protein 1). Furthermore, the protein sequence of the recently submitted UniProtKB entry Snodprot1 of *H. virens* (UniProtKB accession number Q1KHY4) is highly similar to *H. atroviridis* Epl1 and has the same N-terminus of the mature protein as the plant response elicitor described by Hanson and Howell [8], and therefore supports the conclusion that we cloned the corresponding ortholog of this elicitor in our study.

Cloning of *epl1* from *H. atroviridis* and characterization of the protein

Using conserved primers designed from the ESTs that were identified in the MS analysis, the cDNA and genomic DNA of the corresponding gene was cloned from *H. atroviridis* P1 as described in Experimental procedures. The *epl1* gene contains an ORF of 417 bp interrupted by one intron (63 bp), and the lengths of the 5'UTR (untranslated region) and 3'UTR are 122 bp and 227 bp, respectively, as determined by analysis of the cDNA. The gene encodes a precursor protein of 138 amino acids. SIGNALP [18] predicts an 18 amino acid N-terminal signal sequence

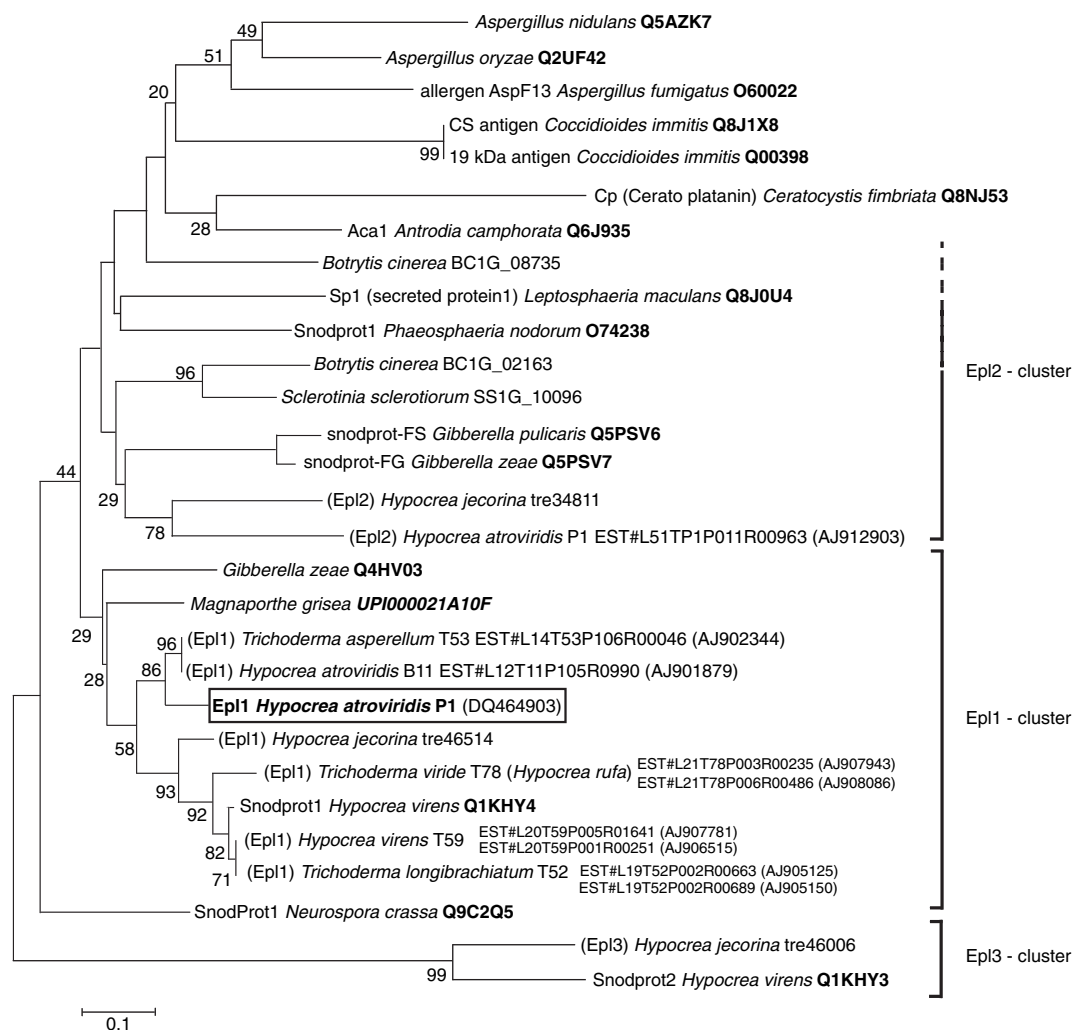


Fig. 4. Phylogeny of the cerato-platanin family. Proteins similar to Epl1 were identified by a BLASTP search. The mature proteins (after cleavage of the signal peptide as predicted with SIGNALP) were used for phylogenetic analysis using neighbor joining. The bar marker indicates the genetic distance, which is proportional to the number of amino acid substitutions. Protein names, as listed in the respective database entries, if available, are shown before the species name. UniProtKB accession numbers are given in bold, and UniParc accession numbers in bold and italics. If only entries in the respective genome databases were available (<http://www.broad.mit.edu/annotation/fgi/> for *Botrytis cinerea* and *Sclerotinia sclerotiorum* and <http://gsphere.lanl.gov/trire1/trire1.home.html> for *Hypocrea jecorina*), the respective protein accession numbers are shown. The ESTs of the various *Hypocrea/Trichoderma* sequences were derived from the TrichoEST database (<http://www.trichoderma.org>), and the respective DDBJ/EMBL/GenBank accession numbers are given in parentheses.

which targets Epl1 to the secretory pathway and leads to D as the N-terminus of the mature protein. This was confirmed by the MS data, which identified the corresponding peptide correctly. The mature protein has a theoretical pI of 5.3 and a calculated average molecular mass of 12 627 Da (predicted with the pI/MW tool [19]), which is slightly below the value (16 kDa) determined by 2D-GE. As no obvious targets for post-translational processing such as N-glycosylation were detected, and also 66.6% of the protein sequence coverage identified in the MS experiments as

well as the intact protein carried no post-translational modifications except for the tryptophan oxidations, this suggested that the protein did not unfold completely during 2D-GE. To verify this finding, the Epl1 protein was purified from the cell culture supernatant by ion exchange chromatography, and the molecular mass of the protein was measured by LC-ESI-IT MS, giving an average molecular mass of 12 629 Da, which is in very good agreement with the calculated value. Two minor components representing two oxidations could also be detected.

The grand average of hydropathicity (GRAVY) was determined by PROTPARAM to be -0.062 , indicating a well-soluble, nonhydrophobic protein. A hydropathicity plot for Epl1 is given in Fig. 3B, which shows that the protein contains hydrophobic and hydrophilic domains.

The secondary structure of Epl1 was predicted with PSIPRED, which is based on position-specific scoring matrices [20,21] (Fig. 3C). The majority of the protein folds to a random coil, interrupted by short, mostly 4–7 amino acid, stretches of strands. The C-terminus of the protein contains two helices, separated by a 14 amino acid strand.

INTERPROSCAN analysis [22] of Epl1 showed the affiliation of this protein to the cerato-platanin family (IPR010829). This is a group of low molecular weight, 4-cysteine-containing fungal proteins that are characterized by high sequence similarity, but do not always have clear functional similarities. Some of these proteins have been reported to act as phytotoxins [e.g. cerato-platanin of *Ceratocystis fimbriata* f. sp. *platani*, Snodprot1 of *Phaeosphaeria nodorum* and Sp1 of *Leptosphaeria maculans*] or human allergens and pathogenesis-related proteins (As-CG of *Coccidioides immitis*, Aca1 of *Antrodia camphorata* and Aspfl3 of *Aspergillus fumigatus*).

It should be noted that a low similarity of *H. atroviridis* Epl1, but not its orthologs from other *Hypocrea/Trichoderma* species (just below the INTERPROSCAN cut-off value), to the domain structure of Barwin-related endoglucanases (IPR009009) was also detected. Members of this group include, for example, expansins, which are involved in plant cell wall extension, and pollen allergens.

Phylogenetic relationship of Epl1 to other members of the cerato-platanin family

An NCBI BLASTP search with *H. atroviridis* Epl1 revealed highest similarity to hypothetical proteins from *H. virens* (UniProtKB accession number Q1KHY4, 2e-60, 86% positives), *Gibberella zeae* (UniProtKB accession number Q4HV03, expected 5e-56, 84% positives), *Magnaporthe grisea* (UniParc accession number UPI000021A10F, 9e-51, 79% positives) and *Neurospora crassa* (UniProtKB accession number Q9C2Q5, 7e-51, 77% positives), followed by snodprot-FS from *G. pulicaris* (UniProtKB accession number Q5PSV6, 2e-44, 75% positives) and snodprot-FG from *G. zeae* (UniProtKB accession number Q5PSV7, 8e-44, 73% positives), and other members of the cerato-platanin family.

The identified proteins were aligned and subjected to neighbor-joining analysis using MEGA3.1 (Fig. 4).

Bootstrap support for most branches was low, which indicates that these proteins reflect little phylogenetic history because important members in the tree are not known or extinct. However, at the intrageneric clade, some clustering was apparent, such as the branches leading to all *Magnaporthe/Gibberella/Hypocrea/Trichoderma* Epl1 orthologs, or the branch containing the Epl-like proteins from *Aspergillus* spp. Taking only fungi for which the complete genomic sequence is available into account, it is interesting that the *Aspergillus* spp. contain only a single member of this protein family, whereas the pyrenomycetes *Gibberella* and *Hypocrea* display two and three, respectively, different clusters of orthologs. We suggest that the *Hypocrea* orthologs should consequently be named Epl2 and Epl3 (Fig. 4). Epl2 is unlikely to be a pseudogene in *Hypocrea/Trichoderma* spp., because in *H. jecorina* it is supported by EST sequences (<http://gsphere.lanl.gov/trire1/trire1.home.html>), and ESTs encoding the Epl2 in *H. atroviridis* can be found in the TrichoEST database (<http://www.trichoderma.org>) (Fig. 4). Interestingly, *H. jecorina* Epl3 and an orthologous protein of *H. virens* form a basal clade in the analysis, which also exhibits the highest genetic distance to the proteins from all other fungi. Nevertheless, sequence analysis of these two proteins clearly identifies a four-cysteine-containing cerato-platanin domain, and a BLASTP search always yielded the members of the cerato-platanin family as the best hits. It is possible that they represent an ancestral cerato-platanin member that is no longer present in the other genera.

Transcription of *ep11* is modulated by specific growth conditions

Epl1 was identified as the major protein formed by *H. atroviridis* during growth on glucose. To characterize the transcript pattern of *ep11* in more detail and to test whether it was constitutively expressed, a number of different growth conditions were chosen for transcript analysis of *ep11* (Fig. 5). We demonstrated that the *ep11* transcript was present during growth on glucose, and it was even present during carbon source-induced and salt-induced osmotic stress. Growth on other soluble carbon sources revealed a weak *ep11* signal on glycerol, whereas the *ep11* transcript was abundantly present on L-arabinose and D-xylose. Growth on colloidal chitin and on the cell walls of *R. solani*, cultivation conditions under which a spot with the same molecular mass and pI as Epl1 in 2D-GE could be detected, also showed a rather high transcript abundance of *ep11*. Under nitrogen starvation, where no corresponding spot was found in the 2D gels, only

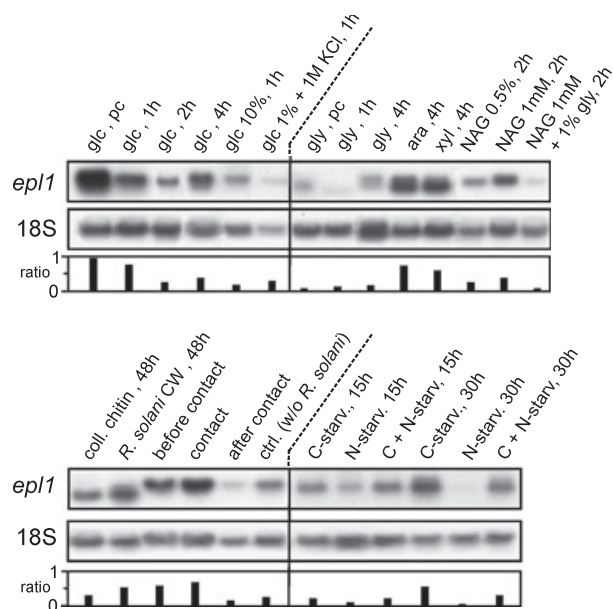


Fig. 5. Analysis of transcript formation of *Hypocrea atroviridis* *epl1*. The culture conditions were: growth on 1% of the carbon sources glucose (glc), glycerol (gly), L-arabinose (ara), D-xylose (xyl), colloidal chitin (coll. chitin) and *Rhizoclonia solani* cell walls (CW); preculture (pc) and replacements on fresh medium for the given time are shown. Additionally, induction experiments with *N*-acetylglucosamine (NAG) and plate confrontation assays with the plant pathogen *R. solani* at the time points before contact, during contact and after contact of the mycelia and *H. atroviridis* alone on plates (ctrl.) are shown. Osmotic stress was applied with 10% glucose or 1 M KCl (+ 1% glucose). Carbon or/and nitrogen starvation experiments were carried out on 0.1% glucose or/and one-tenth of the nitrogen source [0.14 g·L⁻¹ (NH₄)₂SO₄], respectively. 18S rRNA was used as loading control. The bars below the RNA tracks represent the corresponding densitometric scanning of the *epl1* signal, normalized to that of the 18S rRNA. The values are shown relative to the highest value.

a faint signal was detected after 15 h and 30 h, whereas under carbon starvation, even after 30 h a moderately strong signal was still visible. In addition, *epl1* transcript was found in induction experiments with *N*-acetylglucosamine and during plate confrontation assays with *R. solani*, where it was more abundant before contact and upon contact with the host than after contact and in the control (*H. atroviridis* without host). The *epl1* signals in the northern analysis resulted in the hybridization of two bands of slightly different size. It seems unlikely that these signals originate from unspecific hybridization of other genes encoding the proteins of the cerato-platanin family, if the respective *H. jecorina* DNA sequences are compared. Alternative transcription start sites could be detected neither in the available *H. jecorina* ESTs nor upon amplification of *H. atroviridis* *epl1* cDNA. This suggests that, eventu-

ally, spliced and unspliced mRNA species were present, as was recently demonstrated for *H. atroviridis* chitinases [23]. However, our data showed that *epl1* was transcribed under all cultivation conditions tested, although the intensity of the signal varied and was lowest during growth on glycerol and under nitrogen starvation.

Discussion

In this work, we identified a small protein, Epl1, which is the major component of the secretome of *H. atroviridis* on glucose and was expressed under all growth conditions tested, including various carbon sources, plate confrontation assays, osmotic stress and starvation. Although the TrichoEST database comprises ESTs of several *Hypocrea/Trichoderma* species, and the genome database of *H. jecorina* is available, it was impossible to identify Epl1 via peptide mass fingerprinting. This was due to amino acid exchanges that changed the molecular mass of the peptides. Only because of the strong similarity of Epl1 to its orthologs was an identification of spots g1 and g2 via peptide sequence tags and cross-species identification possible.

Analysis of Epl1 revealed it to be a member of the novel cerato-platanin family (IPR010829), which are small proteins that share high sequence similarities, and all of which have four conserved cysteine residues. Cerato-platanin induces phytoalexin production and/or plant cell death in host and nonhost plants [24–26]. Snodprot1 of *Phaeosphaeria nodorum* is produced during infection of wheat leaves [27], and Sp1 of *L. maculans* during infection of *Brassica napus* cotyledons [28]. The Asp13 allergen from *Aspergillus fumigatus* has been characterized as an allergen of human bronchopulmonary aspergillosis [29], and the CS-Ag from *Coccidioides immitis*, which is produced by the saprophytic and the parasitic phases of *Coccidioides immitis*, the causative agent of the human respiratory disease San Joaquin Valley fever, was proposed as a *Coccidioides*-specific antigen for the diagnosis of this fungus [30,31].

The restricted description of members of the cerato-platanin protein family might lead to the conclusion that they may be specifically involved in plant and human pathogenesis and allergic reactions, but members of this family are also found in nonpathogenic filamentous fungi such as *Aspergillus nidulans* and *N. crassa*. They also seem to be abundantly expressed in other fungi, as evidenced by the presence of, for example, approximately 60 ESTs for *N. crassa* and 30 ESTs for *M. grisea* (<http://www.broad.mit.edu/annotation/fgi/>), both cultivated under laboratory conditions. The amino acid sequences of proteins with a

cerato-platanin domain are highly conserved, which further indicates that the occurrence of members of this protein family is not restricted to pathogenic fungi but is universal and may therefore have an important function for filamentous fungi, e.g. involvement in cell wall morphogenesis. In accordance with such a role, Boddi *et al.* [24] located cerato-platanin, which was originally identified from culture filtrates of *Ceratocystis fimbriata* f. sp. *platani*, in the cell walls of ascospores, hyphae and conidia of this fungus. The authors suggested that cerato-platanin may have a similar role to hydrophobins [24,25]. As shown in Fig. 3B, Epl1 contains hydrophobic as well as hydrophilic domains, and its GRAVY is -0.062 , well below the value for hydrophobins, e.g. Hfb2 of *H. jecorina* with 0.694 . However, Hfb1 of *H. jecorina*, which has been reported to be involved in hyphal development [32,33], has a GRAVY of 0.091 and a hydrophobicity profile that is similar to that of *H. atroviridis* Epl1, although Epl1 is, according to its amino acid pattern, clearly no hydrophobin. It could be speculated that Epl1 is a member of a novel class of proteins that have an amphiphilic function in fungal growth and interaction of the fungus with its environment.

The two spots, g1 and g2, with molecular masses of 16 and 27 kDa in 2D-GE, respectively, were both identified as Epl1 on the basis of MS data. The only difference that could be detected between the PMFs comprised two double oxidations, which were solely found in the monomer (g1). The similarity of the PMFs argues against a degraded form of a similar protein or a heterodimer, and rather suggests that spot g2 is an Epl1 dimer. An interesting finding was that the Epl1 monomer contained two double oxidations, both of them most likely located on tryptophans (P1, P4). The oxidations on tryptophan and/or histidine could either be artefacts resulting from sample preparation [34–36], or represent selectively double-oxidized tryptophan residues, which have already been reported for other proteins [37–40]. This is of particular interest because tryptophan oxidation products are themselves capable of generating reactive oxygen species [41], which are responsible for degenerative processes [42], and are involved in plant defense responses [43].

Epl1 is strongly similar (86% positives) to Snodprot1 of *H. virens* (UniProtKB accession number Q1KHY4), which was recently submitted. The N-terminal sequence of the mature *H. virens* protein is identical to the N-terminus of an 18 kDa elicitor that was found in a search of components from *H. virens* that induce terpenoid synthesis (hemigossypol and desoxihemigossypol) in cotton radicles [8]. This elicitor was putatively identified as a serine proteinase, based

on the similarity of the N-terminal sequence tag to a serine protease from *F. sporotrichioides*. However, the UniParc entry of this serine protease (UPI000017B41E) contains only a fragment (24 amino acids), and no published data are associated with it. The high similarity between Epl1 and the 18 kDa elicitor found by Hanson and Howell [8] strongly suggests that Epl1 can indeed function as an elicitor of plant defense responses, which is consistent with the action of other members of this protein family as elicitors and/or even phytotoxins. A glycoside family 11 endoxylanase and a cellulase of *Hypocrea/Trichoderma* has already been shown to elicit defense responses in plants [14,16], but Epl1 would be the first apparently nonenzymatic protein with an elicitor function whose gene has been cloned from any *Hypocrea/Trichoderma* species. With respect to our finding in this study that *epl1* was expressed under all growth conditions tested, and taking into account the fact that we found three cerato-platanin family members in the *H. jecorina* genome database, it will be interesting to study the role of Epl1 and its paralogs in *Hypocrea/Trichoderma* and whether they can functionally compensate for each other.

Experimental procedures

Strains

Hypocrea atroviridis P1 (ATCC 74058) was used in this study and maintained on potato dextrose agar (Difco, Franklin Lakes, NJ, USA). *Escherichia coli* JM109 (Promega, Madison, Germany) was used for plasmid propagation.

Culture conditions

For 2D-GE, shake flask cultures were prepared with a medium containing $0.68 \text{ g}\cdot\text{L}^{-1} \text{ KH}_2\text{PO}_4$, $0.87 \text{ g}\cdot\text{L}^{-1} \text{ K}_2\text{HPO}_4$, $1.7 \text{ g}\cdot\text{L}^{-1} (\text{NH}_4)_2\text{SO}_4$, $0.2 \text{ g}\cdot\text{L}^{-1} \text{ KCl}$, $0.2 \text{ g}\cdot\text{L}^{-1} \text{ CaCl}_2$, $0.2 \text{ g}\cdot\text{L}^{-1} \text{ MgSO}_4\cdot 7\text{H}_2\text{O}$, $2 \text{ mg}\cdot\text{L}^{-1} \text{ FeSO}_4\cdot 7\text{H}_2\text{O}$, $2 \text{ mg}\cdot\text{L}^{-1} \text{ MnSO}_4\cdot 7\text{H}_2\text{O}$ and $2 \text{ mg}\cdot\text{L}^{-1} \text{ ZnSO}_4\cdot 7\text{H}_2\text{O}$, and incubated on a rotary shaker (150 r.p.m.) at 25°C . Cultures were pre-grown for 20 h on 1% (w/v) glucose, harvested by filtering through Miracloth (Calbiochem, Darmstadt, Germany), washed with medium without nitrogen or carbon source, and transferred to a new flask containing 1% (w/v) glucose for another 20 h. For nitrogen starvation experiments, the medium contained only $0.17 \text{ g}\cdot\text{L}^{-1} (\text{NH}_4)_2\text{SO}_4$ after the replacement. Cultivations with colloidal chitin (1% w/v) and *R. solani*, *Pythium ultimum* and *Botrytis cinerea* cell walls were grown for 48 h directly on the respective carbon sources.

For northern analysis, cultures were pregrown on the various carbon sources, and the mycelia were washed and transferred to the growth conditions specified in Results (Fig. 5). Experiments were carried out as previously described by Seidl *et al.* [23] for growth on various carbon sources and under starvation conditions, and also for plate confrontation assays and the preparation of colloidal chitin and fungal cell walls. Osmotic stress experiments were carried out as described by Seidl *et al.* [44].

Preparation and purification of extracellular proteins from *H. atroviridis* culture filtrates

Culture supernatants were isolated by filtration of the *H. atroviridis* cultures through two sheets of Miracloth and subsequent filtration through a 0.22 µm filter (Steritop Filter, Millipore, Billerica, MA, USA) to remove spores and mycelial residues prior to further purification steps, so that the extracellular protein extracts were not contaminated with proteins not of genuinely extracellular origin. The extracts were stored at -80 °C. For concentration, the protein extracts were thawed and always kept at 2 °C during the following steps. Protein concentration was carried out in an Amicon stirred cell 8400 (Millipore) with an Ultracel Amicon YM3 3000 Da NMWL membrane (Millipore), and continued with Amicon Ultra-15 Centrifugal Filter Units with 3000 Da NMWL membranes (Millipore). Dialysis was also carried out in the Amicon centrifugal filter units (according to the manufacturer's instructions) by refilling the tubes three times with cold, distilled water. Proteins were further purified by trichloroacetic acid/acetone precipitation. The pellets were resolubilized in 2D sample buffer containing 9 M urea, 2% Chaps, 1% dithiothreitol, 0.5% carrier ampholytes and 0.1% (v/v) of a protease inhibitor cocktail (Sigma, St Louis, MO, USA), by vortexing and vigorous shaking at room temperature for several hours, and centrifuged at 60 000 g at 20 °C for 30 min (Sigma 3-18K, rotor 12154). The protein concentration of the supernatant was determined with the modified Bio-Rad assay (Bio-Rad, Hercules, CA, USA), and the protein solutions were stored at -80 °C until 2D-GE.

For cation exchange chromatography, proteins were concentrated and purified as described above but resolubilized in acetate buffer (10 mM, pH 4.5). The proteins were loaded onto a MonoS HR 5/5 (GE Healthcare, Little Chalfont, UK) column equilibrated with the same buffer. They were then eluted with a linear sodium chloride gradient (0–250 mM). Progress of the chromatography was monitored by measuring the absorbance at 280 nm. One-milliliter fractions were collected, and the fraction containing the most abundant peak contained the Epl1 protein, as determined by SDS/PAGE (data not shown), was consequently precipitated using chloroform/methanol precipitation [45].

2D-GE

The protein samples that were dissolved in 2D sample buffer as described above were used for overnight in-gel rehydration of pH 4–7, 17 cm immobilized pH gradient (IPG) strips (Bio-Rad) by applying 300 µg of protein, solubilized in 300 µL of 2D buffer. IPGs were focused using the IEF cell (Bio-Rad). The focusing program included a linear ramp to 300 V over 1 h, a linear ramp to 1000 V over 1 h, a linear ramp from 1000 to 10 000 V over 2 h, and 60 000 Volt-hours at 10 000 V_{\max} with a limit of 50 µA per IPG strip. The IPG strips were equilibrated for 15 min in equilibration buffer (6 M urea, 2% SDS, 0.05 M Tris/HCl, pH 8.8, 20% glycerol) containing 2% dithiothreitol, and for 15 min in equilibration buffer containing 2.5% iodoacetamide. The strips were then mounted on 12% SDS-polyacrylamide gels. The gels were run at 25 mA for the stacking gel and 35 mA for the separating gel per gel, and stained with Simply Blue (Invitrogen, Paisley, UK). PageRuler Prestained Molecular Weight Marker (Fermentas, St Leon-Rot, Germany) was used for molecular mass determination of proteins. At least three to five gels were run on each sample. The 2D gels were matched and analyzed with the PDQUEST software (Bio-Rad).

Protein analysis by MS (MALDI-RTOF MS, MALDI-TOF/RTOF MS, HPLC-ESI-IT MS)

The spots of interest from the 2D gels were excised manually with a stainless steel scalpel and subjected to in-gel digestion [46] using trypsin (bovine pancreas, modified; sequencing grade; Roche, Madison, Germany). Extracted tryptic peptides were desalted and purified utilizing ZipTip® technology (C₁₈ reversed phase, standard bed; Millipore) [47]. Sample preparation for MALDI MS was carried out on a stainless steel target, using a thin-layer preparation technique [48] with α -cyano-4-hydroxy-cinnamic acid (Fluka, Buchs, Switzerland) as matrix dissolved in acetone (6 mg·mL⁻¹).

Positive ion mass spectra for peptide mass fingerprinting were recorded on a MALDI-TOF/RTOF instrument (Axima TOF²; Shimadzu Biotech, Manchester, UK) equipped with a nitrogen laser ($\lambda = 337$ nm) by accumulating 200–1000 single unselected laser shots. The instrument was operated throughout all peptide mass fingerprinting experiments in the reflectron mode, applying 20 kV acceleration voltage and delayed extraction (optimized setting for ions of m/z 2500). External calibration was performed using an aqueous solution of standard peptides (bradykinin fragment 1–7, human angiotensin II, somatostatin and adrenocorticotrophic hormone fragment 18–39). The lists of monoisotopic m/z values derived from the MALDI mass spectra of in-gel digested spots were submitted to the MASCOT search engine [49] for a PMF search in several proprietary and public genomic databases using a tailor-made bioinformatics facility. The MASCOT search was run against all proteins and DNA sequence information from public databases

(MSDB, Swiss-Prot, NCBIInr) and the genome sequence information from the fungi *Aspergillus nidulans*, *G. zeae*, *M. grisea*, *N. crassa*, *Ustilago maydis*, *H. jecorina*, and the TrichoEST database (<http://www.trichoderma.org>) containing 26 different cDNA libraries derived from 12 strains of seven species of *Hypocrea/Trichoderma*, including *H. atroviridis* P1. Restrictions for peptide mass tolerance (± 0.7 Da), fixed modifications (carbamidomethylation) and variable modifications (oxidation of M, W, H) were set for the PMF MASCOT search.

In all cases, seamless PSD and/or high-energy CID MS/MS experiments were performed by accumulating 1000–5000 single unselected laser shots to collect sequence tags for protein identification, selecting characteristic tryptic peptides. PSD or high-energy CID studies were carried out on the same instrument mentioned above, with helium as collision gas in the latter case. For PSD and/or high-energy CID database searches, the same tailor-made MASCOT search engine was used, applying the same settings for species and modifications as mentioned above but without using trypsin as a specific enzyme and adding precursor (± 0.7 Da) and product ion tolerances (± 1 Da). Proteins were identified based on PSD and/or high-energy CID experiments in which the database search result gave a significant hit in terms of the probability-based MOWSE score (significance threshold $P < 0.05$) [50].

For determination of the molecular weight of the intact Epl1, it was purified by ion exchange chromatography as described above, and the lyophilized sample was reconstituted in 40 μ L of 4% trifluoroacetic acid. Reverse-phase separation (HPLC) of 10 μ L of protein solution was performed on an Elite La Chrome HPLC System (Hitachi, Tokyo, Japan) using a C4 column (4.6×250 mm, 300 \AA ; Advance Chromatography Technology, Aberdeen, UK) at a flow rate of 500 $\mu\text{L}\cdot\text{min}^{-1}$. Solvent A was 0.3% formic acid in double-distilled water, and solvent B was 0.3% formic acid in acetonitrile. The gradient consisted of isocratic conditions at 5% solvent B for 10 min, a linear gradient to 90% solvent B over 50 min, a linear gradient to 95% solvent B over 5 min, and then a linear gradient back to 5% solvent B over 5 min. The HPLC was connected online to a Bruker Esquire 3000+ quadrupole ion trap mass spectrometer (Bruker Daltonik, Bremen, Germany). Nitrogen was used as nebulizer gas (17 $\text{lb}\cdot\text{in}^{-2}$ pressure) and as drying gas (12 $\text{L}\cdot\text{min}^{-1}$ at 350 $^{\circ}\text{C}$). The spray voltage was set to 4.5 kV. Capillary exit, tube lens, skimmer and quadrupole/octopole voltages were tuned for maximum transmission of multiply charged cytochrome *c* ions. Deconvolution of the mass spectrum was performed using the software provided by the manufacturer.

PCR-aided methods

PCR reactions were carried out in a total volume of 50 μL containing 2.5 mM MgCl_2 , 10 mM Tris/HCl (pH 9.0),

50 mM KCl, 0.1% (v/v) Triton X-100, 0.4 μM each primer, 0.2 mM each dNTP and 0.5 units of *Taq* polymerase (Promega, Madison, WI, USA). The amplification program consisted of: 1 min of initial denaturation (94 $^{\circ}\text{C}$), 30 cycles of amplification (1 min at 94 $^{\circ}\text{C}$, 1 min at primer-specific annealing temperature, 1 min at 72 $^{\circ}\text{C}$), and a final extension period of 7 min at 72 $^{\circ}\text{C}$.

Cloning of the *H. atroviridis epl1* gene

cDNA was synthesized with the Creator SMART cDNA library construction kit (BD Biosciences, Palo Alto, CA, USA) from RNA from *H. atroviridis* cultures grown on glucose. The conserved *H. lixii/T. asperellum* primers snod-fw (5'-TGTCCAACCTCTTCAAGC-3') and snod-rv (5'-TAGAGGCCGCGAGTTGC-3') were used to clone a gene fragment of *H. atroviridis epl1* from the cDNA. Additionally, combinations of the 5'PCR and CDSIII primers from the cDNA kit with snod-rv and snod-fw and the nested primer snod-fwnest (5'-GTCTCTGCTGATACC GTCTCG-3'), respectively, were used to amplify the 5'- and 3'-cDNA ends of *epl1*.

To amplify the genomic DNA of *epl1*, primers 5'-GGGAGCCTTCATCACAAC-3' and 5'-TAATTTAGT AGTAGCGTCTGCC-3', which are located in the 5'UTR and 3'UTP of *epl1*, were used.

The resulting fragments were cloned into pGEMT-Easy (Promega) and sequenced at MWG Biotech (Ebersberg, Germany).

The assembled DNA sequences of *epl1* were deposited in DDBJ/EMBL/GenBank (accession number DQ464903).

Phylogenetic analysis

Protein sequences were aligned first with CLUSTALX 1.8 [51] and then visually adjusted using GENEDOC 2.6 [52]. Phylogenetic analyses were performed in MEGA 2.1 using neighbor joining, a distance algorithmic method. Stability of clades was evaluated by 1000 bootstrap rearrangements. Bootstrap values lower than 20% are not displayed in the cladogram.

RNA isolation and hybridization

Fungal mycelia were harvested by filtration through Miracloth (Calbiochem), washed with cold tap water, squeezed between two sheets of Whatman filter paper, shock frozen and ground in liquid nitrogen. Total RNA was extracted as described previously [53]. Standard methods [54] were used for electrophoresis, blotting and hybridization of nucleic acids.

The 409 bp *epl1* PCR fragment generated with the above-described primers snod-fw and snod-rv was used as the probe for northern hybridizations, and a 297 bp PCR fragment of

the 18S rRNA gene (DDBJ/EMBL/GenBank accession number Z48932) was used as the hybridization control. The relative abundance of transcripts was determined by densitometric measurements of autoradiographs derived from different exposure times with the GS-800 densitometer (Bio-Rad) and analysis with the QUANTITY ONE1-D ANALYSIS software (Bio-Rad). The values are integrated peaks and were corrected by global background subtraction.

Acknowledgements

This work was supported by the EU-funded TrichoEST project (QLK3-2002-02032) and formed part of the mass spectrometric investigations by the Austrian Science Foundation (P15008 to GA). The authors wish to acknowledge the important contribution of their colleagues from the TrichoEST consortium to the generation of the EST database, and especially Patrizia Ambrosino and Luis Sanz for providing purified cell walls of plant pathogenic fungi. The authors also wish to thank Christian Gamauf for his help in Epl1 purification with ion exchange chromatography.

Note added in proof

After acceptance of this manuscript, a paper by Djonovic S, Pozo MJ, Dongott LJ, Howell CR and Kenerley CM (2006) *Mol Plant Microbe Interact* **19**, 838–853 was published, which provides genetic evidence that the *T. virens* orthologue of Epl1 is indeed an elicitor of plant defense responses.

References

- Benitez T, Rincon AM, Limon MC & Codon AC (2004) Biocontrol mechanisms of *Trichoderma* strains. *Int Microbiol* **7**, 249–260.
- Harman GE, Howell CR, Viterbo A, Chet I & Lorito M (2004) *Trichoderma* species – opportunistic, avirulent plant symbionts. *Nat Rev Microbiol* **2**, 43–56.
- Howell CR (2003) Mechanisms employed by *Trichoderma* spp. in the biological control of plant diseases: the history and evolution of current concepts. *Plant Dis* **87**, 4–10.
- Kubicek CP, Mach RL, Peterbauer CK & Lorito M (2001) *Trichoderma*: from genes to biocontrol. *J Plant Pathol* **83**, 11–23.
- Ahmed SA, Sanchez CP & Candely ME (2000) Evaluation of induction of systemic resistance in pepper plants (*Capsicum annuum*) to *Phytophthora capsici* using *Trichoderma harzianum* and its relation with capsidiol accumulation. *Eur J Plant Pathol* **106**, 817–824.
- Bigirimana J, De Meyer G, Poppe J, Elad Y & Höfte M (1997) Induction of systemic resistance on bean (*Phaseolus vulgaris*) by *Trichoderma harzianum*. *Med Fac Landbouww University Gent* **62**, 1001–1007.
- De Meyer G, Bigirimana J, Elad Y & Höfte M (1998) Induced systemic resistance in *Trichoderma harzianum* T39 biocontrol of *Botrytis cinerea*. *Eur J Plant Pathol* **104**, 279–286.
- Hanson LE & Howell CR (2004) Elicitors of plant defense responses from biocontrol strains of *Trichoderma virens*. *Phytopathology* **94**, 171–176.
- Harman GE, Petzold R, Comis A & Chen J (2004) Interactions between *Trichoderma harzianum* strain T22 and maize inbred line Mo17 and effects of these interactions on diseases caused by *Pythium ultimum* and *Colletotrichum graminicola*. *Phytopathology* **94**, 147–153.
- Yedidia I, Benhamou N, Kapulnik Y & Chet I (2000) Induction and accumulation of pathogenesis related protein activity during the early stage of root colonization by *Trichoderma harzianum*. *Plant Physiol Biochem* **38**, 863–873.
- Yedidia I, Shoshani M, Kerem Z, Benhamou N, Kapulnik Y & Chet I (2003) Concomitant induction of systemic resistance to *Pseudomonas syringae* pv. *lachrymans* in cucumber by *Trichoderma asperellum* (T-203) and accumulation of phytoalexins. *Appl Environ Microbiol* **69**, 7343–7353.
- Yedidia II, Benhamou N & Chet II (1999) Induction of defense responses in cucumber plants (*Cucumis sativus* L.) by the biocontrol agent *Trichoderma harzianum*. *Appl Environ Microbiol* **65**, 1061–1070.
- Howell CR, Hanson LE, Stipanovic RD & Puckhaber LS (2000) Induction of terpenoid synthesis in cotton roots and control of *Rhizoctonia solani* by seed treatment with *Trichoderma virens*. *Phytopathology* **90**, 248–252.
- Fuchs Y, Saxena A, Gamble HR & Anderson JD (1989) Ethylene biosynthesis-inducing protein from cellulysin is an endoxylanase. *Plant Physiol* **89**, 138–143.
- Lotan T & Fluhr R (1990) Xylanase, a novel elicitor of pathogenesis-related proteins in tobacco, uses a non-ethylene pathway for induction. *Plant Physiol* **93**, 811–817.
- Martinez C, Blanc F, Le Claire E, Besnard O, Nicole M & Baccou JC (2001) Salicylic acid and ethylene pathways are differentially activated in melon cotyledons by active or heat-denatured cellulase from *Trichoderma longibrachiatum*. *Plant Physiol* **127**, 334–344.
- Hermosa MR, Grondona I, Diaz-Minguez JM, Iturriaga EA & Monte E (2001) Development of a strain-specific SCAR marker for the detection of *Trichoderma atroviride* 11, a biological control agent against soilborne fungal plant pathogens. *Curr Genet* **38**, 343–350.
- Nielsen H, Engelbrecht J, Brunak S & von Heijne G (1997) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst* **8**, 581–599.

- 19 Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD & Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* (Walker JM, ed.), pp. 571–607. Humana Press, Iotowa, NJ, USA.
- 20 Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS & Jones DT (2005) Protein structure prediction servers at University College London. *Nucleic Acids Res* **33**, W36–W38.
- 21 Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195–202.
- 22 Zdobnov EM & Apweiler R (2001) InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848.
- 23 Seidl V, Huemer B, Seiboth B & Kubicek CP (2005) A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases. *FEBS J* **272**, 5923–5939.
- 24 Boddi S, Comparini C, Calamassi R, Pazzagli L, Cappugi G & Scala A (2004) Cerato-platanin protein is located in the cell walls of ascospores, conidia and hyphae of *Ceratocystis fimbriata* f. sp. *Platani*. *FEMS Microbiol Lett* **233**, 341–346.
- 25 Pazzagli L, Cappugi G, Manao G, Camici G, Santini A & Scala A (1999) Purification, characterization, and amino acid sequence of cerato-platanin, a new phytotoxic protein from *Ceratocystis fimbriata* f. sp. *Platani* *J Biol Chem* **274**, 24959–24964.
- 26 Scala A, Pazzagli L, Comparini C, Santini A, Tegli S & Cappugi G (2004) Cerato-platanin, an early-produced protein by *Ceratocystis fimbriata* f. sp. *platani*, elicits phytoalexin synthesis in host and non-host plants. *J Plant Pathol* **86**, 27–33.
- 27 Sharen AL & Krupinski G (1970) Cultural and inoculation studies of *Septoria nodorum*, cause of Glume Blotch of wheat. *Phytopathology* **60**, 1480–1485.
- 28 Wilson LM, Idnurm A & Howlett BJ (2002) Characterization of a gene (*sp1*) encoding a secreted protein from *Leptosphaeria maculans*, the blackleg pathogen of *Brassica napus*. *Mol Plant Pathol* **3**, 487–493.
- 29 Kurup VP, Banerjee B, Hemmann S, Greenberger PA, Blaser K & Cramer R (2000) Selected recombinant *Aspergillus fumigatus* allergens bind specifically to IgE in ABPA. *Clin Exp Allergy* **30**, 988–993.
- 30 Cole GT, Zhu SW, Pan SC, Yuan L, Kruse D & Sun SH (1989) Isolation of antigens with proteolytic activity from *Coccidioides immitis*. *Infect Immun* **57**, 1524–1534.
- 31 Pan S & Cole GT (1995) Molecular and biochemical characterization of a *Coccidioides immitis*-specific antigen. *Infect Immun* **63**, 3994–4002.
- 32 Askolin S, Penttilä M, Wösten HA & Nakari-Setälä T (2005) The *Trichoderma reesei* hydrophobin genes *hfb1* and *hfb2* have diverse functions in fungal development. *FEMS Microbiol Lett* **253**, 281–288.
- 33 Linder MB, Szilvay GR, Nakari-Setälä T & Penttilä ME (2005) Hydrophobins: the protein-amphiphiles of filamentous fungi. *FEMS Microbiol Rev* **29**, 877–896.
- 34 Bienvenut WV, Deon C, Pasquarello C, Campbell JM, Sanchez JC, Vestal ML & Hochstrasser DF (2002) Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics* **2**, 868–876.
- 35 Karty JA, Ireland MME, Brun YV & Reilly JP (2002) Artifacts and unassigned masses encountered in peptide mass mapping. *J Chromatogr B* **782**, 363–383.
- 36 Thiede B, Lamer S, Mattow J, Siejak F, Dimmler C, Rudel T & Jungblut PR (2000) Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion. *Rapid Commun Mass Sp* **14**, 496–502.
- 37 Dad S, Bisby RH, Clark IP & Parker AW (2005) Identification and reactivity of the triplet excited state of 5-hydroxytryptophan. *J Photochem Photobiol B* **78**, 245–251.
- 38 Hiner AN, Martinez JJ, Arnao MB, Acosta M, Turner DD, Lloyd Raven E & Rodriguez-Lopez JN (2001) Detection of a tryptophan radical in the reaction of ascorbate peroxidase with hydrogen peroxide. *Eur J Biochem* **268**, 3091–3098.
- 39 Rigby SE, Junemann S, Rich PR & Heathcote P (2000) Reaction of bovine cytochrome c oxidase with hydrogen peroxide produces a tryptophan cation radical and a porphyrin cation radical. *Biochemistry* **39**, 5921–5928.
- 40 Svistunenko DA (2005) Reaction of haem containing proteins and enzymes with hydroperoxides: the radical view. *Biochim Biophys Acta* **1707**, 127–155.
- 41 Finley EL, Dillon J, Crouch RK & Schey KL (1998) Identification of tryptophan oxidation products in bovine alpha-crystallin. *Protein Sci* **7**, 2391–2397.
- 42 Osiewacz HD (2002) Genes, mitochondria and aging in filamentous fungi. *Ageing Res Rev* **1**, 425–442.
- 43 Apel K & Hirt H (2004) Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu Rev Plant Biol* **55**, 373–399.
- 44 Seidl V, Seiboth B, Karaffa L & Kubicek CP (2004) The fungal STRE-element-binding protein Seb1 is involved but not essential for glycerol dehydrogenase (*gld1*) gene expression and glycerol accumulation in *Trichoderma atroviride* during osmotic stress. *Fungal Genet Biol* **41**, 1132–1140.
- 45 Wessel D & Flügge UI (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **138**, 141–143.
- 46 Shevchenko A, Wilm M, Vorm O & Mann M (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal Chem* **68**, 850–858.

- 47 Pluskal MG (2000) Microscale sample preparation. *Nat Biotechnol* **18**, 104–105.
- 48 Kussmann M, Nordhoff E, Rahbek-Nielsen H, Haebel S, Rossel-Larsen M, Jakobsen L, Gobom J, Mirgorodskaya E, Kroll-Kristensen A, Palm L *et al.* (1997) Matrix-assisted laser desorption/ionization mass spectrometry sample preparation techniques designed for various peptide and protein analytes. *J Mass Spectrom* **32**, 593–601.
- 49 Perkins DN, Pappin DJ, Creasy DM & Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567.
- 50 Pappin DJ, Hojrup P & Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **3**, 327–332.
- 51 Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F & Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
- 52 Nicholas HBJ & McClain WH (1987) An algorithm for discriminating transfer RNA sequences. *Computer Applications Biosci* **3**, 177–181.
- 53 Chomczynski P & Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**, 156–159.
- 54 Sambrook J & Russell DW (2001) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Plainview, NY.

Lebenslauf

Name: Reingard Schandl

Geboren: 17. September 1981, in Wien

Staatsangehörigkeit: Österreich

Familienstand: ledig

Ausbildung:

1987 – 1991 Volksschule, 1040 Wien

1991 – 1999 Bundesgymnasium Geringergasse, 1110 Wien

seit 1999 Studium der Technischen Chemie, TU Wien

04.2002 Abschluss der 1. Diplomprüfung
Spezialisierung auf Technische Biochemie und
Naturstofftechnologie

10.2002 – 06.2003 Auslandsstudium an der Universidad de Córdoba,
Spanien

03.2004 Teilnahme am „Athens-Programm“, einem
einwöchigen Austausch zwischen verschiedenen
technischen Universitäten, an den Grandes Ecoles
d’Ingenieurs de Paris

11.2004 Teilnahme am „Athens-Programm“ an der Universidad
Politecnica de Madrid

seit 08.2005 Arbeit an der Diplomarbeit „Identifizierung
extrazellulärer Proteine des biologischen
Pflanzenschutzmittels *Trichoderma atroviride* mittels
Elektrophorese und Massenspektrometrie“