Proteomic analysis of the biological control fungus *Trichoderma*

Jasmine Grinyer

Bachelor of Technology (Biotechnology), Macquarie University

Bachelor of Technology (Biotechnology) Hons, Macquarie University

A thesis presented to Macquarie University in partial fulfillment for the degree of Doctor of Philosophy.

Proteome Systems Ltd.,

The Department of Biological Sciences, Macquarie University and

Department of Chemistry and Biomolecular Sciences, Macquarie University,

Sydney, NSW, Australia

August 2006

Table of contents

Abstract

Trichoderma harzianum and *T. atroviride* are filamentous fungi commonly found in soil. Both display biocontrol capabilities against a range of phytopathogenic fungi including *Rhizoctonia solani* and *Botrytis cinerea* which are known pests of hundreds of commercially important crops including tomatoes, potatoes, beans, cucumber, strawberries, cotton and grapes. These *Trichoderma* species secrete a combination of enzymes degrading cell walls and antibiotics to overgrow and kill fungal phytopathogens. They are seen as an environmentally friendly alternative to chemical fungicides currently used on crops.

A proteomic approach was taken to separate and identify proteins from a strain of *T. harzianum* with well established biocontrol properties. Several methods were developed in this thesis to display the whole proteome content and several subcellular proteome fractions from *T. harzianum*. Proteins were separated by two-dimensional electrophoresis and identified by mass spectrometric methods. The resulting proteomic maps represent the first extensive array of cellular and sub-cellular proteomes for *T. harzianum*.

Cellular protein patterns of *T. atroviride* (*T. harzianum* P1) grown on media containing either glucose or *R. solani* cell walls were compared by differential gel electrophoresis to identify a suite of new proteins involved in the biological control response. Twenty four *T. atroviride* protein spots up-regulated in the presence of the *R. solani* cell walls were identified by mass spectrometry and N-terminal sequencing. Proteins identified from this study included previously implicated enzymes degrading cell walls and three novel proteases, vacuolar serine protease, vacuolar protease A and trypsin-like

protease. The genes encoding two of these proteases, vacuolar protease A and vacuolar serine protease have been cloned by degenerate primer PCR and genomic walking PCR and sequenced. The gene sequences and protein sequences derived from these genes have been partially characterised.

Declaration

This thesis contains original work, which was performed by me. Several aspects of this work have been performed in collaboration with others; these people have been acknowledged and their contributions recognised in the section in which their assistance was received. No part of this thesis has been presented to any other institution for any other award. I consent to this thesis being made available for photocopy or for loan.

Jasmine Grinyer

August 2006

Publications

The following publications are presented in their published form in this thesis and are referred to from this point onwards as listed in roman numerals.

Publication I:

Herbert BR, Grinyer J, McCarthy JT, Isaacs M, Harry EJ, Nevalainen H, Traini MD, Hunt S, Schulz B, Laver M, Goodall AR, Packer J, Harry JL, Williams KL (2006) Improved 2-DE of microorganisms after acidic extraction. Electrophoresis 27:1630-1640

Publication II:

Grinyer J, McKay M, Nevalainen H, Herbert BR (2004) Fungal proteomics: initial mapping of biological control strain *Trichoderma harzianum.* Curr Genet 45:163-169

Publication III:

Grinyer J, McKay M, Herbert BR, Nevalainen H (2004) Fungal proteomics: mapping the mitochondrial proteins of a *Trichoderma harzianum* strain applied for biological control. Curr Genet 45:170-175

Publication IV:

Pedersen SK, Harry JL, Sebastion L, Baker J, Traini MD, McCarthy JT, Manoharan A, Wilkins MR, Gooley AA, Righetti PG, Packer NH, Williams KL, Herbert BR (2003) Unseen proteome: Mining below the tip of the iceberg to find low abundance and membrane proteins. J Proteome Res 2:303-311

Publication V:

Grinyer J, Hunt S, McKay M, Herbert BR, Nevalainen H (2005) Proteomic response of biological control *Trichoderma atroviride* to growth on the cell walls of *Rhizoctonia solani*. Curr Genet 47:381-388

Publication VI:

Herbert BR, Hopwood F, Oxley D, McCarthy J, Laver M, Grinyer J, Goodall A, Williams K, Castagna A, Righetti PG (2003) β-elimination: an unexpected artefact in proteome analysis. Proteomics 3:826-831

Please Note: my maiden name of Baker was used in Publication IV.

Abbreviations

2D	Two-dimensional
2DE	Two-dimensional electrophoresis
ASB-14	Tetradecanoylamido propyl dimethyl ammonio propane sulfonate
bp	Base pairs
C7BzO	3-(4-Heptyl)phenyl 3-hydroxy propyl dimethyl ammonio propane sulfonate
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propane sulfonate
CID	Collision-induced dissociation
CSI	Cross species identification
CWDE	Cell wall-degrading enzymes
DIGE	Differential gel electrophoresis
DTT	Dithiothreitol
GlcNAc	1,4-β-linked N-acetyl-D-glucosamine
GRAVY	Grand average hydropathy
GWPCR	Genomic walking polymerase chain reaction
IEF	Isoelectric focusing
IPG	Immobilised pH gradient
kDa	Kilodaltons
LC-MS/MS	Liquid chromatography mass spectrometry
MALDI	Matrix assisted laser desorption ionisation
MCE	Multi-compartment electrolyser
MS	Mass spectrometry
MudPIT	Multi-dimensional protein identification technology
NP-40	Nonidet P-40
PAGE	Polyacrylamide gel electrophoresis
PAM	Propionamide
PCR	Polymerase chain reaction
PMF	Peptide mass fingerprinting
PTM	Post-translational modification
SDS	Sodium dodecyl sulphate
TBP	Tributylphosphine
TOF	Time of flight

Acknowledgements

This thesis comprises approximately three years of full-time research conducted in the laboratories of Proteome Systems Ltd., North Ryde, Sydney Australia (http://www.proteomesystems.com), the Department of Biological Sciences, Macquarie University over the period 2002-2005 and the Department of Chemistry and Biomolecular Sciences, Macquarie University during 2006.

I would like to thank both of my supervisors Professor Helena Nevalainen and Associate Professor Ben Herbert for contributing in such different ways in this thesis. Thanks Ben, for initiating this PhD project through Proteome Systems – I know it was not easy to convince the powers that I was deserving enough. Thanks also Ben, for making me laugh when I needed a positive outlook. I would also like to thank Helena Nevalainen for keeping me motivated, and on-track to finish this thesis. I know Helena had to put up with many of my grammatical errors over the years (with dread), but hopefully they have become more infrequent as time passed.

I would also like to thank Professor Peter Bergquist for reviewing my thesis prior to submission. I appreciate the time you have spent assisting me and your comments on thesis writing have been extremely valuable.

Completing my thesis has been harder than I expected with a young, growing family. My husband, Scott has dealt with many of my frustrations (and the occasional good moments) and he especially deserves a heartfelt thanks. Without his support, this work

would have remained incomplete for many more years. For my daughter, Emily, you will now get some well-deserved attention from mum, along with baby number 2.

I would like to thank many people who I have worked with over the years, for their endless support, motivation and advice. Many thanks must go to Sanne Pedersen, Matt McKay, Sybille Hunt, Janice Joss and Mathew Traini from Proteome Systems. Thanks to Ron Bradner, Junior Te'o, Moreland Gibbs, Natalie Curach, Caiyan Wu, Niru Chikballapur and Liisa Kautto for helping to keep me sane at Macquarie University. The work presented here was much easier to complete with your help and support.

1. Introduction

The biological control response of *Trichoderma* spp. to fungal pathogens of plants has been investigated for more than twenty years. However, limited information is available on the role played by *Trichoderma* proteins during this response. Traditionally, genes and proteins involved in biocontrol have been characterised addressing one gene or protein at a time. The advent of gene transcription profiling and proteomics has enabled a holistic approach to gene expression and protein research where multiple gene products involved in a biological process can be simultaneously identified. On proteomics, where the current thesis will be concentrating, the key for success is efficient extraction of proteins from the organism of interest for their analysis by two-dimensional electrophoresis (2DE). Proteins are particularly difficult to extract from fungi as fungal cell walls are very thick and possess a high carbohydrate content. The two main skeletal carbohydrates in the fungal cell wall are chitin (polymer of 1,4-β-linked N-acetyl-D-glucosamine) and β-glucan (polymer of D-glucose). High carbohydrate content can make protein separation by 2DE difficult and new extraction procedures need to be developed to eliminate the contaminating carbohydrate material. Multiple protein identification strategies are required to maximise the number of separated fungal proteins as only a relatively small number of fungal genome sequences are available (http://www.biologie.uni-hamburg.de/b-online/library/genomeweb/GenomeWeb/fungal-gen-db.html). The *T. harzianum* genome sequence is still unavailable although the *T. reesei* sequence is in final stages of round two sequencing. Considering all of the above, the starting point for this thesis was the development of methodologies for efficient extraction of proteins from filamentous fungi and separation of the proteins by 2DE.

1.1 Proteomics and two-dimensional electrophoresis

Proteomics is the large-scale study of protein structure and function. The proteome is defined as the entire protein content of an organism through its life cycle, or on a smaller scale, the entire protein content of a particular cell type or cellular compartment, under a particular type of stimulation at any one time (Monti et al. 2005). 2DE remains the most widely accepted method of studying an organisms' proteome and related protein expression, despite the recent development of alternative methods including multi-dimensional protein identification technology (Aebersold and Mann 2003; Washburn et al. 2001) and stable isotope labeling (Gygi et al. 1999). 2DE is the best available method for analysing large, complex mixtures of proteins from samples such as whole cell lysates. The large-scale separation ability of intact proteins achieved by 2DE has not yet been surpassed by any other technique, which is an important factor when considering the size of the proteome of an organism. The human genome, for example, has been estimated to contain between 25,000 – 30,000 genes. However, the proteome complexity is greatly increased if protein post-translational modifications, such as glycosylation, phosphorylation and splice variants are considered, bringing the original estimation of 30,000 proteins up to approximately 300,000 (Righetti et al. 2005). Similar increases in proteome size are expected for any organism where post-translational modifications and splice variants occur. Post-translational modifications are most often manifested as protein isoforms, i.e. proteins with similar molecular weights but different isoelectric points, which are resolved on 2D gels.

2DE is an electrophoretic technique that separates proteins by two independent fundamental parameters. The first dimension of 2DE separates a mixture of proteins

according to their isoelectric point or pI, by isoelectric focusing (IEF). During IEF, proteins migrate along a pH gradient until they reach their pI, or net zero charge. For IEF, immobilised pH gradient (IPG) strips are commonly used to reproducibly separate μg - mg quantities of protein mixtures. IPG strips replace the less-reproducible technology based on carrier ampholyte IEF (Righetti and Drysdale 1973). IPG technology involves polymerising acrylamide with a pH gradient to a plastic backing sheet. This results in a highly reproducible, stable medium for user-handling while providing maximum protein loading capabilities (Bjellqvist et al. 1982). The second dimension of 2DE is the separation of proteins according to their size by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS PAGE). Prior to SDS PAGE during an IPG equilibration step, sodium dodecyl sulphate (SDS) binds to focused proteins giving them a negative charge, enabling the protein mixture to migrate into the second dimension gel with identical electrophoretic mobilities (Gorg et al. 2004). A range of SDS PAGE gels are manufactured commercially with various acrylamide gradients, eliminating the need for users to pour their own gels, resulting in high levels of reproducibility. The combination of IEF and SDS PAGE can result in the visualisation of up to 5000 protein spots on a large format (25 x 25 cm) single gel. However, separation of 1000-2000 spots is more realistic (Gorg et al. 2004).

While the separation of thousands of proteins on a single acrylamide gel is a significant accomplishment, it is clearly not enough to tackle the complexity of, for example, the human proteome with all of its expected isoforms. Over the past decade, much research has been conducted into simplifying an organism's proteome. This has been done by replacing a single whole cell protein extract of an organism with two or more cellular extracts, thereby greatly increasing the number of proteins that can be

displayed by 2DE from a single sample. One approach has been to isolate different cellular compartments or sub-cellular fractionation (Huber et al. 2003). This works by displaying a smaller subset of proteins on a 2D gel, such as a cytoplasmic fraction, cell wall fraction, secreted fraction, or organelle fractions such as the nuclear fraction and mitochondrial fraction. One of the main advantages of subcellular fractionation is the visualisation of proteins that may have been hidden by more abundant proteins, such as highly abundant cytoplasmic proteins. Other approaches include simplifying the proteome by sample prefractionation (Righetti et al. 2003), or the use of narrow-range IPG strips (Wildgruber et al. 2000). With all of the above-mentioned cellular fractions, the first, most crucial step is to extract the proteins of interest in a contaminant-free, high-yielding manner appropriate for 2DE.

1.1.1 Recent advances in the preparation of proteomic samples

Recent advances in the preparation of proteomic samples have yielded many beneficial improvements to the proteomic field, in particular to 2DE. Ultimately, these advances have resulted in higher protein yields extracted from samples, better resolved 2DE protein maps, more membrane proteins observed on two-dimensional (2D) maps and a higher level of reproducibility obtained. These improvements are a direct result of improvements in cell lysis equipment such as high powered ultrasonic probes, bead mills or pressure cycling technology (PCT; Smejkal et al. 2006), the combined use of 2D-compatible chemicals that aid in protein extraction and solubilisation with stronger chaotropic detergents to solubilise previously insoluble membrane proteins. The use of uncharged denaturing agents and new zwitterionic surfactants (Santoni et al. 1999) assists in maintaining a denatured state, in-sample alkylation (Galvani et al. 2001a; Galvani et al. 2001b; Herbert et al. 2001; Righetti 2006) allows retention of proteins in

their denatured state during electrophoresis, and the use of compatible protease inhibitor cocktails to prevent protein degradation has minimised protein artefacts on 2D gels (Grinyer et al. 2004a). The availability of more sensitive staining and detection methodology (Berggren et al. 2002) has led to the visualisation of a larger number of proteins by 2DE. All of these improvements have helped to create higher levels of resolution on 2D gels with the end result that more protein is separated and visualised from an organism, thereby leading to a more complete picture of an organism's proteome.

Cell lysis

The availability of more efficient cell lysing equipment including bead mills (Retsch, Germany) and ultrasonic probes with microtips (Branson Ultrasonic, USA) has been a major breakthrough for extracting proteins from organisms containing a thick cell wall. Proteome research of tough samples such as plant material (i.e. leaf), fungal mycelia and gram-positive bacteria have all benefited from the development of efficient cell-lysing equipment. Bead mills function to assist cell lysis by combining appropriately sized glass beads with a cellular sample. The tubes are shaken vigorously and the beads smash against the sample with great force causing the cells to rupture. Rapid agitation of *Aspergillus oryzae* mycelia in a mini-bead mill with glass beads resulted in the highest level of protein extraction when compared to boiling in strong alkali, boiling in SDS and chemical lysis with Y-PER reagent (Nandakumar and Marten 2002). Ultrasonicators function to assist cell lysis by passing high frequency sound waves into a cellular sample through a tapered microtip. The energy produced from the sound waves is high enough to rupture the majority of cell types in less than one minute. Ultrasonic cell lysis yielded an equal amount of protein extracted as obtained

using the French Press and has far exceeded traditional methods of cell lysis including boiling in sodium hydroxide, SDS and mutanolysin treatment of *Clostridium perfringens* (Guerlava et al. 1998). Pressure cycling technology (PCT) is a highly efficient cell lysis technique that uses alternating cycles of high and low pressure to induce cell lysis. For *Escherichia coli*, 14.2 % more total protein could be extracted by PCT than using a standard bead mill (Smejkal et al. 2006).

Chaotropes

Cell lysis, and the ability to retain maximal protein solubility during cell lysis, is important when discussing ways of improving protein yield on 2D gels and visualising a greater proportion of the lower abundance proteins including membrane proteins. The benefits of the addition of chaotropes, such as urea and thiourea, to protein extraction solutions have been known for some time (Rabilloud 1998, Molloy et al. 1998, Molloy et al. 1999, Musante et al. 1998). Urea acts as a denaturant by disrupting hydrogen bonds leading to protein unfolding (Gorg et al. 2004) and thiourea acts to break hydrophobic interactions, such as interactions between membrane proteins and the cell wall or membrane (Rabilloud 1998). Together, urea and thiourea appear to be an ideal combination for protein extraction for 2DE (Rabilloud et al. 1997; Rabilloud 1998). The combination of urea and thiourea results in a higher level of protein solubility obtained which leads to better resolution on 2D gels (Rabilloud 1998). However, there is one major limitation to the use of urea for protein extraction. When urea is heated to 37 ºC or above, it forms an equilibrium with ammonium isocyanate which can bind to and modify proteins. Protein modification of this type is called carbamylation (McCarthy et al. 2003). The isocyanate ions can cause extensive modifications by forming blockages at the N-terminus of the protein and modifications

to lysine, arginine and cysteine amino acids which in turn create false pI isoforms and untrue peptide masses for peptide mass fingerprinting (PMF; McCarthy et al. 2003). Carbamylation of proteins does not occur under electrophoretic conditions as isocyanate ions are charged and migrate out of the system (McCarthy et al. 2003), but can occur under poorly controlled sample preparation and storage conditions. Carbamylation of proteins can be eliminated by ensuring the sample is maintained at temperatures ranging from 15-20 ºC during sample preparation.

Detergents

The addition of detergents to protein extraction solutions greatly enhances protein solubility, particularly of those proteins containing hydrophobic regions. SDS is a very efficient detergent for solubilising proteins. However, it is almost completely incompatible with 2DE (Luche et al. 2003). It is an anionic detergent that interferes with the low salt, low charge conditions required for IEF. If SDS is to be used as a detergent for 2DE, its concentration must be diluted to below 0.2 % and should be quenched four-fold with a nonionic or zwitterionic detergent in the presence of urea and thiourea to retain protein solubility (Gorg et al. 2004). The development of new nonionic and zwitterionic detergents is being pursued in an attempt to find a 2DE compatible detergent that has protein solubility properties equal to those of SDS. Some nonionic detergents commonly used for 2DE analysis include Triton X-100, NP-40 and dodecyl maltoside. The latter appears to be the most promising nonionic detergent in terms of its ability to solubilise membrane proteins (Luche et al. 2003). Zwitterionic detergents such as CHAPS and the sulfobetaines ASB-14 and C7BzO in combination with urea and thiourea have proven to be most effective at solubilising membrane proteins, but are quite expensive (Chevallet et al. 1998; Tastet et al. 2003).

Unfortunately there is not a single extraction solution that will yield optimal protein extraction for every type of biological system and sample. Therefore, a range of detergents in combination with urea and thiorea-containing extraction solutions need to be tested before proteomic studies can begin (Luche et al. 2003).

Reducing and alkylating agents and protease inhibitors

Other important reagents for 2DE include reducing agents, alkylating agents and protease inhibitors. Reducing agents cleave disulfide bonds within a protein and between protein complexes. Two commonly-cited reducing agents used for 2DE are dithiothreitol (DTT) and tributylphosphine (TBP). While DTT is charged, TBP is uncharged and thus better suited to the electrophoretic conditions of 2DE. TBP is used in very low concentrations (approximately 5 mM) and is very efficient at reducing proteins that contain a high cysteine content (Herbert et al. 1998). For extraction solutions containing TBP, acrylamide should be added (10 mM final concentration) to alkylate the reduced proteins. The alkylation reaction covalently attaches a propionamide (PAM) group to all cysteine amino acids (Cys_PAM), preventing disulfide bonds from reforming and in turn, preventing a beta-elimination artefact (loss of a H_2S group from cysteine residues) in 2DE (Herbert et al. 2003, Publication VI, Section 5.1). However, the efficiency of both the reduction and alkylation reactions using TBP and acrylamide hinge on the pH of the extraction solution and ultimately require a basic pH (at least pH 8.3, the pK of cysteine) for complete reaction. Small amounts of Tris-base (40 mM) are added to the sample solubilisation solution to raise the pH to approximately pH 9-10 to achieve rapid and complete reduction and alkylation reactions. Protease inhibitors have been found to play a crucial role in preventing the degradation of proteins in filamentous fungi (Grinyer et al. 2004a,

Publication II, Section 3.1.2). However, they are not typically required for protein extractions from the yeast *Saccharomyces cerevisiae* or the Gram-negative bacterium, *E. coli*. It is believed that the denaturing conditions in the protein extraction solution are strong enough to denature proteases. This is true to a limited extent but is not always the case. Regardless of the sample for 2DE, protease inhibitors should be added to protein extraction solutions to prevent protein degradation by proteolysis (Grinyer et al. 2004a, Publication II, Section 3.1.2).

1.1.2 Simplification to ultimately resolve more of the proteome

Due to the limited number of proteins that can be displayed on a single 2D gel (typically 1000-2000 per gel), it is necessary to split the proteome of an organism into different compartments to see a greater proportion. This can be done in a number of ways: to display proteins based on their solubility in different sample solutions; to display proteins according to their cellular location; to prefractionate a whole cell extract into narrower pH ranges; to view the proteome over narrow, overlapping pH ranges using narrow-range IPG strips; to remove highly abundant proteins to reveal the less-abundant proteins, or to use a combination of the methods outlined above.

Protein solubility

An effective way of simplifying the proteome to see a wider range of proteins is to differentially extract proteins based on their solubility properties. For instance, soluble, cytoplasmic proteins can be extracted initially in a solution containing only 40 mM Tris-base and then the remaining cellular pellet is solubilised in a solution containing urea and a detergent such as CHAPS. The final pellet then can be extracted in a solution containing a stronger protein extraction solution such as a combination of

urea, thiourea and zwitterionic detergents such as ASB-14 or C7BzO (Molloy et al. 1998). Protein extractions of this type have worked well for sequentially extracting proteins from *E. coli*, where 11 membrane proteins were purified in the final extract, five of which never had been identified before on 2D gels (Molloy et al. 1998). Almost 90 % of proteins were solubilised in the first two extraction steps, indicating that a very small membrane-enriched protein fraction would remain for the final extraction step. Sequential extraction of proteins based on solubility has the added benefit of being conducted in a single tube, further minimising protein loss (Molloy et al. 1998). More recently, a similar sequential extraction procedure was used to separate proteins from a meningococcal strain, *Neisseria meningitidis.* As a result, thirty three protein products were identified which previously had never been identified on 2D gels (Mignogna et al. 2005).

Organelle isolation

A logical way of simplifying the proteome of an organism is to isolate organelles from eukaryotic cells prior to protein extraction, thereby obtaining one or more proteomic maps or 2D gels from each organelle separated. By considering the cell membrane as an organelle, this type of proteome simplification can also work for prokaryotic organisms. Organelles are typically purified by sucrose density ultracentrifugation (Huber et al. 2003). Enriched mitochondrial fractions have also been obtained from filamentous fungi and yeasts using a differential centrifugation method (Grinyer et al. 2004b, Publication III, Section 3.1.3). Proteins linked to cell membranes can be purified by sodium carbonate extraction (Molloy et al. 2000), where ultracentrifugation is used to pellet cellular membranes under basic pH conditions. These membranes can then be treated with urea, thiourea and detergents for protein solubilisation and 2DE.

Membrane proteins have also been extracted from micro-organisms with organic solvents such as a one to one mixture of chloroform and methanol (Molloy et al. 1999). Organic solvent extraction of this type yielded eight proteins that had not been reported previously from 2D gels. Five of these eight proteins contained a positive grand average hydropathy (GRAVY) value indicating these proteins are highly hydrophobic. Commercial kits are available from Sigma-Aldrich, Pierce and Qiagen for the purification of organelles and extraction of proteins for 2DE.

Protein prefractionation

It is becoming apparent that the majority of proteins identified by 2DE are highly abundant, hydrophilic proteins, whereas up to 80 % of the yeast proteome, for example, consists of low abundance proteins. A protein's cellular abundance can be predicted from the codon usage bias of the encoding gene (Kurland 1991). A technique that enhances the chance of visualising low abundance proteins in a proteome involves prefractionation of a protein sample. Sample prefractionation involves separating proteins, typically by charge, prior to IEF and 2DE (Herbert and Righetti 2000) and works equally well with whole cell or subcellular protein fractions. The development of multi-compartment electrolysers (MCE) and similar devices (Rotofor, BIORAD; Gradiflow, Gradipore; reviewed in Stasyk and Huber 2004) allows a mixture of proteins to be electrophoretically separated into pH range compartments as narrow as 0.1-0.5 pH units. These prefractionated samples can be applied to narrow range IPG strips specifically suited to the pH range of the sample, eliminating interference from proteins that do not have a pI specific for the pH range of the IPG. Prefractionation of a sample can remove abundant proteins, i.e. albumin from human plasma (Herbert and Righetti 2000), and can assist in the isolation of low abundance proteins (Pedersen et

al. 2003, Publication IV, Section 3.1.4). Sample prefractionation of the yeast membrane proteome followed by mass spectrometry aided in the identification of 780 protein isoforms from 323 gene products. From these, 49 % were confirmed to be membrane bound or associated and 28 % were low abundance proteins. Prefractionation of protein samples is the future of protein analysis by 2DE, but relies upon the requirement for high protein concentrations in starting material.

Removal of abundant proteins

Removal of abundant proteins is an important aspect for many proteome projects. While not considered further in this thesis, it is relevant to disease-based proteomic studies using human serum or plasma to identify disease biomarkers. Six highly abundant proteins represent 85 % of the total protein content in human plasma including albumin, immunoglobulin (Ig) G , IgA, haptoglobin, alpha-1-antitrypsin and transferrin. It is essential to remove these proteins to visualise the proteome of the less-abundant proteins, which are likely to comprise disease biomarkers. This can be undertaken with a range of methods including affinity purification by Cibacron Blue (Coomassie blue dye-based method), monoclonal antibody purification, or IgG and IgY depletion (Zolotarjova et al. 2005). A limitation of these methods is the co-depletion of low abundance proteins, as observed when cytokines were co-depleted with the Montage Albumin Deplete Kit from Millipore (Granger et al. 2005). The MCE has been used to capture albumin in a narrow pH range, allowing 2DE studies of acidic and alkaline proteins in human plasma (Herbert and Righetti 2000).

Narrow-range IPGs

The simplest method of resolving a greater number of proteins from a cellular protein extract is to display the proteome across numerous narrow-range IPG strips, but this procedure can be problematic. Narrow range IPG strips work well if they fall into a pH range where the majority of proteins have their pI. If the characteristics of the yeast proteome are taken as an example, more than 90 % of the proteome lies within the 4-10 pH range. This observation creates a problem if a whole cell lysate is taken and displayed across a narrow 6-7 pH gradient, where the pI of less than 20 % of proteins fall (Pedersen et al. 2003, Publication IV, Section 3.1.4). When IEF is applied to the narrow-range IPG, the proteins that have a pI outside of the pH range will stack at either end of the IPG strip. This phenomenon leads to the accumulation of high protein concentrations and can cause protein precipitation at the edges of the IPG strip. Reduced electrophoretic capabilities result, causing protein smearing and incomplete focusing on the 2D gel. However, narrow-range IPG strips can work well if the protein samples are appropriate for the pH range chosen or if the total protein loads are reduced to overcome protein precipitation. The benefits of narrow-range IPG usage were recently demonstrated when almost 2300 *S. cerevisiae* proteins were separated and visualised across numerous narrow-range IPGs, compared to only 755 protein spots visualised across a single 3-10 pH gradient (Wildgruber et al. 2000). Narrow-range IPG strips are very effective at displaying prefractionated pH samples such as those generated by the MCE (Pedersen et al. 2003, Publication IV, Section 3.1.4).

1.1.3 Visualisation of the proteome

After 2DE, protein location needs to be visualised before identification can occur. A broad range of staining methods is used. However, the most common include staining

with colloidal Coomassie G250, silver nitrate or silver diamine silver staining. In addition, fluorescent staining using dyes such as Sypro® Ruby (Molecular Probes) is used and the incorporation of small, uncharged fluorophores to protein samples prior to 2DE. For example, CyDye™ has been introduced.

Coomassie stains

The most common dye for protein staining in 2DE is Coomassie Brilliant Blue, either Coomassie R-250 or Coomassie G-250. Coomassie attaches to proteins by electrostatic and hydrophobic interactions with the amino groups of proteins. Recently, the use of colloidal Coomassie staining has yielded highly sensitive, reproducible staining results (Smejkal 2004). Protein spots with concentrations between 3 ng – 15 ng of protein are detectable (Westermeier and Marouga 2005) which is typically enough protein to obtain mass spectrometry results using MALDI-TOF MS (matrix-assisted laser desorption ionisation time of flight mass spectrometry). Coomassie staining is cheap, quantitative and reproducible.

Silver stains

Silver staining of proteins is also a relatively cheap method of visualising proteins with sensitivity levels reaching below 1 ng of protein per spot (Westermeier and Marouga 2005). Silver staining protocols exist for both silver nitrate and silver diamine stains. Typically, silver staining requires more hands-on time, but, results can be obtained in a shorter overall time than with most Coomassie based stains. On the downside, silver staining can lead to large variations in gel to gel staining making it difficult to be used for differential gel electrophoresis (DIGE). Most sensitive silver stains are in general, incompatible with mass spectrometry. However, there is atleast one mass

spectrometric compatible silver staining method that removes glutaraldehyde from the fixing solution enabling the identification of proteins after silver staining by mass spectrometry (Shevchenko et al. 1996).

Fluorescent stains

Fluorescent dyes were introduced in 1994 (Nile Red; Bermudez et al. 1994) and dyes such as Sypro® Ruby are sensitive enough to detect 1 ng of protein per spot (Steinberg et al. 1996). Other dyes have recently been introduced including Deep Purple™ which can detect protein concentrations as low as 300 pg per spot (Mackintosh et al. 2003). Fluorescent staining provides greater sensitivity than Coomassie staining and can cover a greater dynamic range i.e. it can be used to quantitate very low to high levels of proteins reproducibly, but is more expensive.

The latest advance in detecting proteins on SDS PAGE gels is the introduction of pre-labelled proteins (in a protein mixture) with cyanine dyes (CyDye™ DIGE Fluors, Amersham Biosciences) prior to 2DE. Each CyDye™ is charge and size-matched but has a different excitation and emission wavelength. Two or more samples can be labelled with different dyes and applied to a single 2D gel where co-migration allows for highly reproducible DIGE (Unlu et al. 1997). This technology is currently quite expensive, but greatly reduces the number of 2D gels required for differential analysis and provides high statistical confidence for protein volume comparisons. No doubt, this type of staining will become widely adopted in the 2DE and proteomics field over the next few years. Sensitivity of the CyDyes is at least equal to that of the most sensitive silver stains available.

Recently, stains able to detect protein modifications and characteristics have been coupled with fluorescence. There are two products on the market for protein modifications, one for detecting protein glycosylation (Pro-Q®-Emerald; Hart et al. 2003) and second for detecting protein phosphorylation (Pro-Q®-Diamond; Steinberg et al. 2003). Both dyes allow counterstaining with either Sypro® Ruby or Coomassie to visualise the whole protein content in a gel after detection of post-translational modifications (PTM). A third fluorescent stain has been developed to detect hydrophobic integral membrane proteins. Pro-Q® Amber selectively stains integral membrane proteins in denaturing SDS PAGE gels which has been demonstrated with multiple transmembrane proteins including bacteriorhodopsin and the a and c subunits of F_1F_0 ATP synthase (Hart et al. 2004).

The 2D gel patterns require conversion into digital format for image analysis. Images are typically captured by CCD camera in a light-proof cabinet especially for fluorescently-stained gels, or on high resolution scanners or densitometers for gels stained with silver or Coomassie stains.

1.1.4 Protein identification by mass spectrometry and alternative techniques

Mass spectrometry is the most common, high throughput method for identifying proteins from 2D gels. It accurately weighs peptides after digestion with a suitable enzyme and the masses of these peptides can then be matched against the theoretical peptide masses of known proteins for identification. The two types of mass spectrometry used throughout the work presented here include MALDI-TOF MS and liquid chromatography mass spectrometry (LC-MS/MS).

MALDI-TOF MS

For MALDI-TOF MS analysis, proteins are cut from 2D gels, destained, dried and digested with trypsin (typically about 200 pg trypsin at a concentration of 20 μg mL^{-1}) overnight (Figure 1 shows a schematic of MALDI-TOF-MS; Karas et al. 1987). Tryptic peptides can be applied directly to the MALDI target plate, or can undergo a recommended clean-up and concentration step in C18 zip-tip columns (Millipore). A C18 column desalts and concentrates the peptides, enhancing peptide signal in the mass spectrometer (Wa et al. 2006). Co-elution of peptides from the C18 column with a MALDI matrix (an ultraviolet-absorbing compound), such as α-cyano-4-hydroxy cinnamic acid is typically used for peptide analysis (Mano and Goto 2003). Matrices are used for mass spectrometry to ionise the peptides when fired on by a laser, enabling them to move from the MALDI target plate into the mass spectrometer through this energy transfer (Mano and Goto 2003). Once samples have been inserted into the mass spectrometer, a vacuum is applied to remove air from the system to prevent interference during mass spectrometry. A laser is then fired at the sample, peptides are ionised and move through the flight chamber until they reach a detector at the far end of the chamber; smaller peptides move quickly, larger peptides move more slowly. The detector records the time taken for a peptide to move from the target plate down the length of the flight chamber and the flight time of each peptide is proportional to the square root of the peptides mass to charge (*m/z*) ratio. The resulting spectrum can be calibrated using trypsin auto-digest peptides, spiked with commercially available peptides, or a combination of both used to give peptide masses with an error of less than 0.1 Da (Chernushevich et al. 2001). Peptides with masses between 700 Da and 3300 Da can be detected accurately using MALDI-TOF. The list

of *m/z* values can be compared to lists generated from the theoretical digestion of known proteins for identification (Wysocki et al. 2005).

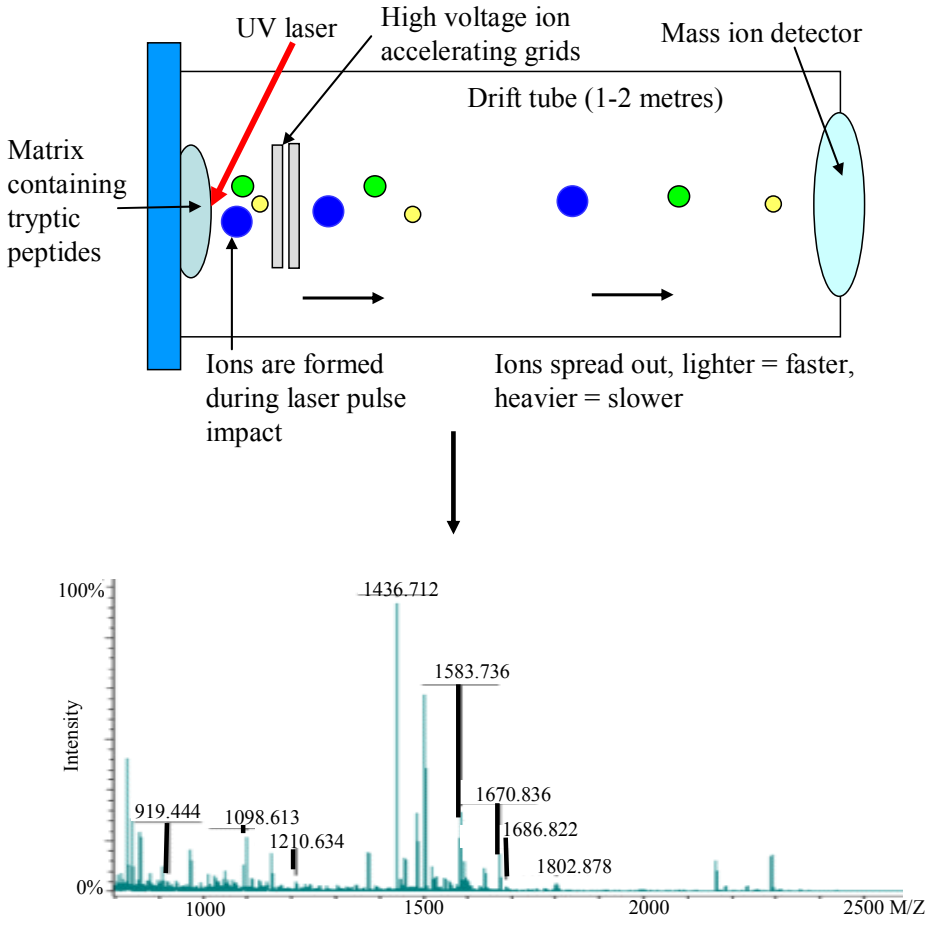


Figure 1: Schematic of MALDI-TOF mass spectrometry and the resulting peptide mass fingerprint. Proteins are initially separated by 2DE and individual spots are excised from the gel. Protein spots are then digested with trypsin and the resulting peptides are ionised and resolved on the basis of mass to charge ratio (M/Z).

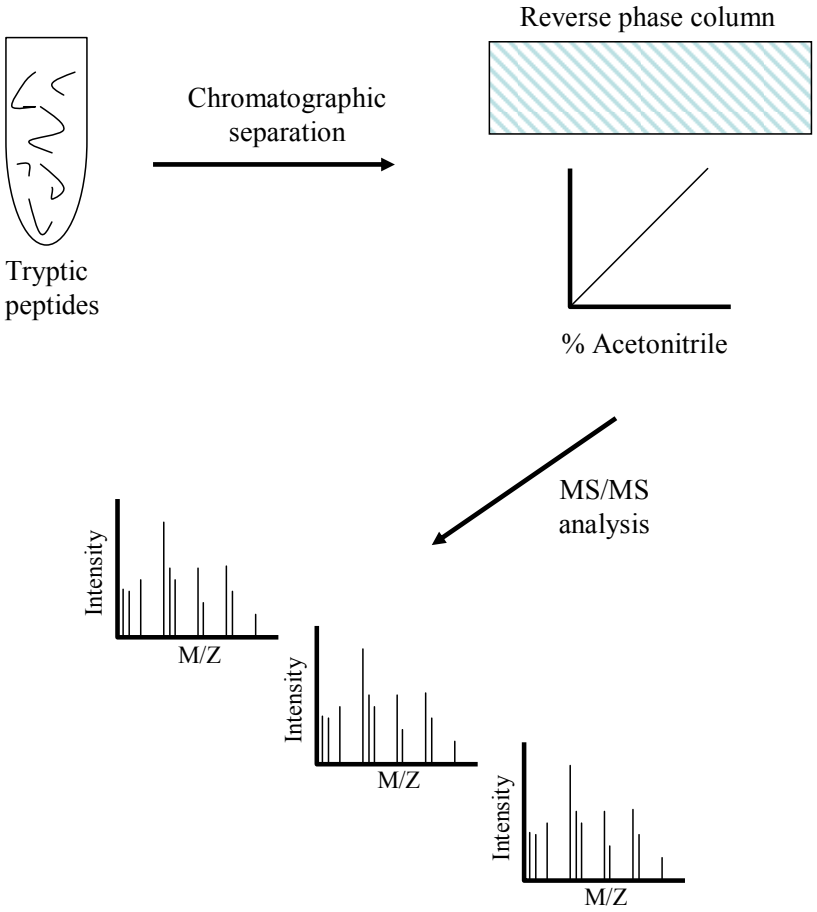


Figure 2: Schematic of liquid chromatography mass spectrometry. Proteins are initially separated by 2DE and individual spots are excised from the gel and digested with trypsin. The resulting peptides are separated on a reverse phase C18 column and eluted based on increasing hydrophobicity with an acetonitrile gradient elution. The MS/MS spectra are used to identify the protein.

LC-MS/MS

LC-MS/MS combines the separation capabilities of liquid chromatography with the accurate mass analysis of mass spectrometry (Figure 2 shows a schematic of LC-MS/MS). With 2DE as the initial protein separation method, individual protein spots are digested with enzymes such as trypsin and an acid added to protonate the peptides.

The digest is applied to an alkyl-containing C8 or C18 reverse phase column. All digested peptides bind to this column based on their hydrophobic interaction with the column's alkyl-groups and the peptides are eluted with a linear concentration gradient of 0-60 % organic solvent, such as acetonitrile (Shi et al. 2004). Thus, reverse phase chromatography provides a phase for peptide separation allowing more peptides to be sequenced as they elute sequentially from the column compared to other MS/MS procedures. Once eluted from the column, peptides enter the mass spectrometer by nanoflow or electrospray ionisation for mass detection, and can be further fragmented for MS/MS or amino acid sequence analysis. A peptide survey scan is typically conducted (MS), and selected peptides (based on intensity, *m/z* or charge) are chosen for MS/MS fragmentation (Delahunty and Yates 2005). Fragmentation of peptides typically takes place by collision-induced dissociation (CID) where an inert gas is inserted into the chamber to provide obstacles for the peptides, forcing them to fragment when collisions occur. Peptide ions fragment in predictable patterns, which ensures spectra can be compared to theoretical fragmentation for protein identification (Delahunty and Yates 2005) using software such as SEQUEST (Qian et al. 2005).

2D-LC-MS/MS

More recently, 2D-LC-MS/MS has gained popularity as it avoids the limitations of 2DE such as the lack of solubilisation and visualisation of low abundance and membrane proteins. Two chromatography columns (multi-dimensional protein identification technology; MudPIT) are used to separate peptide mixtures for 2D-LC-MS/MS. Elution of peptides from the first chromatography column (typically a strong cation exchange column) is undertaken in step gradients before peptides are applied to a C18 reverse phase column, thereby increasing the separation abilities of peptides and

enhancing the number of peptide sequences obtained. The benefits of this technique include the absence of protein separation by SDS PAGE (a mixture of unknown proteins is digested and inserted into the system directly) and enhancement of the likelihood of obtaining protein identifications for a large variety of protein types. Proteins identified by 2D-LC-MS/MS include hydrophobic membrane proteins, low abundance proteins and modified proteins, and has provided valuable information on protein topology and orientation in membranes (Wu and Yates 2003). The major limitation of this technique is that protein characteristics such as molecular weight, pI and post-translational modification status are unknown because the starting sample for 2D-LC-MS/MS is a mixture of peptides from a mixture of unknown proteins. However, it is likely that in the near future, combinations of 2DE with MS and 2D-LC-MS/MS approaches will yield better results than each individual attempt as recently demonstrated by Breci et al. (2005).

Protein databases for identification

Any proteomic project is limited by the protein database available for the organism under study. It is expected that close to 100 % of the protein spots will be identified when MS analysis is conducted from 2D gels of an organism of which the genome has been sequenced. When a protein cannot be identified, and changing the PMF search parameters has not helped, it is generally assumed that the protein is heavily modified e.g. glycosylated. Therefore, further PTM characterisation methods such as enzymatically removing sugars and analysing them by mass spectrometry are required to determine these modifications. However, if proteomic analysis is conducted on an organism where DNA sequencing is incomplete or not publicly available, protein identification frequencies are significantly reduced to less than 20 % as was observed

in a study of the filamentous fungus *Trichoderma harzianum* (Grinyer et al. 2004a, Publication II, Section 3.1.2; Grinyer et al. 2004b, Publication III, Section 3.1.3). Proteomic projects of organisms where no genome sequence data exists, rely heavily on identifying proteins from databases of closely related species (cross species identification; CSI) with peptide mass fingerprint (PMF) data where minimal DNA sequence changes have occurred between species. In a bioinformatic study, it was found that at least 70 % protein sequence identity and 70 % amino acid composition similarity were required to obtain identification from PMF analysis (Wilkins and Williams 1997). Once similarities between the species under study and the sequenced species fall below 70 %, it becomes highly unlikely that protein identifications from PMF data will result (Lester and Hubbard 2002). Tryptic peptide masses of both proteins must be conserved across species to make reliable protein identifications. However, as sequence coverage typically falls in the 20 – 50 % amino acid coverage range, it is unlikely that one or two amino acid changes will greatly affect PMF results. Generally, it has been found that protein mass and pI are not well conserved across species boundaries when protein identity drops below 80 %. This observation was found to particularly be true for protein pI, which was much less stable than molecular mass in a study of bacterial genomes (Lester and Hubbard 2002; Wilkins and Williams 1997) as several amino acid changes can greatly alter the pI of a protein, particularly in smaller proteins.

1.1.5 Differential gel expression (DIGE) and image analysis

Proteomics can begin to provide answers to protein function when protein expression profiles in an organism or cell are compared under two different physiological states (diseased and healthy cells or cells grown under different physiological conditions).

The comparison method is known as differential gel expression (DIGE). Protein up-regulation can be predicted from an increase in protein spot intensity (i.e. doubling) or appearance of a protein spot that was previously absent under alternative growth conditions. Similarly, protein down-regulation can be predicted when a previously seen protein spot appears smaller or disappears under alternative growth conditions. Highly specialised image analysis software programs such as Melanie (GeneBio), PDQuest (BioRAD), and PSLImage (Proteome Systems Ltd.) have been developed for the accurate determination of up or down-regulated proteins. Replicate gel images are entered into these programs where spot detection, spot editing and spot land-marking functions are undertaken to align gel images. Spot volumes can then be compared to determine if a protein is up or down-regulated (Grinyer et al. 2005, Publication V, Section 3.2.1). Traditionally, it was believed that triplicate or quadruplicate gels were required to establish the variation observed between samples. For example, triplicate gels were run for growth on one type of medium and triplicate gels for growth on a second type of medium for each pH range of interest. The minimum number of replicates has been statistically determined but ultimately depends on the number of sample groups being investigated (Hunt et al. 2005). No matter how effective image analysis software becomes, the quality of the gel images will always play a role in determining the usefulness of the data generated. The introduction of the CyDye technology (Unlu et al. 1997) has greatly assisted in increasing reproducibility as two or more samples can be applied to a single 2D gel, thereby limiting the total number of gels required.

1.2 A proteomic approach to study the filamentous fungus - *Trichoderma*

Species of the genus *Trichoderma* have several biotechnological applications. *T. reesei* is well known for its ability to secrete very large amounts of proteins, up to 100 g of protein per litre of medium, making it an ideal tool for the production of large amounts of industrial enzymes and recombinant proteins (Peberdy 1994; Cherry and Fidantsef 2003). *T. harzianum* and *T. atroviride* have biological control properties against phytopathogens, providing a chemical-free alternative against fungal pathogens infecting plant crops (Papavizas 1985; Chet 1987). The mycoparasitic property of *Trichoderma* makes it possible to reduce the amount of chemically-based fungal pesticides used in the environment whilst naturally enhancing crop yields worldwide. Understanding the mechanism of biocontrol that *Trichoderma* spp. employ to provide protection against plant pathogenic fungi will ultimately assist in the development of a strain that has biological control activity against a wider range of fungal pathogens. This in turn leads to environmentally sound protection of a wider range of plant crops under different environmental conditions. A proteomics approach was used during this work as it enables the simultaneous identification of multiple proteins involved in a biological process.

Some proteomic based publications relating to the yeasts *S. cerevisiae* and *Candida albicans* existed at the onset of this project. This provided a blank canvas for the development of methods for protein extraction from filamentous fungi. At that point, only a few genome sequences were available for fungal species, including the yeast *S. cerevisiae*, the dimorphic yeast *Schizosaccharomyces pombe* and filamentous fungus *Neurospora crassa*. The genome sequence for the genus *Trichoderma* was unavailable and very little proteomic work had been completed on filamentous fungi in general; in

fact, only one publication was available on *Trichoderma* proteomics (Lim et al. 2001). The applied aspect of this work was on biological control of particular mycopathogenic fungi and for that purpose, *Trichoderma spp*. was the best choice as a model organism due to its well-established biocontrol capabilities against mycoparasitic fungi. *Trichoderma* species contain the largest number of gene and protein sequences in public databases, an essential component for successful proteomics research.

1.2.1 Overview of biological control mediated by *Trichoderma* spp.

T. harzianum (A3091) and *T. atroviride* ATCC 74058 (formerly *T. harzianum* P1) are filamentous fungi commonly found in soil. Both display biocontrol capabilities against a range of phytopathogenic fungi including *Rhizoctonia solani* and *Botrytis cinerea* (Chet 1987; Papavizas 1985) which are known pests of hundreds of plant crops, including tomatoes, beans, cucumber, strawberries, cotton and grapes (Prins et al. 2000). The mycoparasitic properties of *Trichoderma* enable it to protect plant crops by releasing, for example, cell wall-degrading enzymes (CWDE; Chet et al. 1998; de la Cruz et al. 1992; Geremia et al. 1993; Harman et al. 1993; Lora et al. 1995; Lorito et al. 1994a) and antibiotics (Lorito et al. 1996; Schirmbock et al. 1994) when challenged with pathogenic species of fungi. Table 1 highlights species within the *Trichoderma* genus displaying biocontrol abilities.

Table 1: Species of *Trichoderma* with known biological control functions.

Trichoderma species that have a biocontrol role	Phytopathogens that induce biocontrol	References
T. harzianum	*B. cinerea*	Lorito et al. 1994b; Lorito et al. 1994c
T. atroviride	*B. cinerea* and *R. solani* cell walls	Mach et al.1999; Kullnig et al. 2000
T. asperellum	*R. solani* cell walls	Bara et al. 2003; Ramot et al. 2004; Viterbo et al. 2004
T. longibrachiatum	*Pythium ultimum*	Migheli et al. 1998
T. virens	*R. solani* hyphae	Mukherjee et al. 2003; Pozo et al. 2004
T. viride	*R. solani* and *R. solani* toxin	Shanmugam et al. 2001

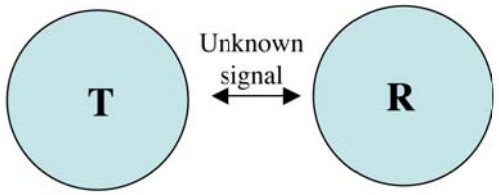
The exact mechanism for how biocontrol species of *Trichoderma* recognise and attack phytopathogenic fungi is unknown, although some determinants have been identified. An unknown diffusable substance (less than 90 kDa in size) triggers the expression of the *ech42* gene of *T. harzianum* P1 (Kullnig et al. 2000). As fungi secrete low levels of exochitinases constitutively, they degrade the fungal cell wall to produce oligomers that are thought to trigger the fungal attack (Benetiz et al. 2004). The *ech42* gene encodes a chitinase that is expressed before *Trichoderma* comes into physical contact with a phytopathogen such as *R. solani*. Other CWDEs including other chitinases, proteases and β-glucanases are released in an attack to weaken the cell wall structure of the fungal pathogen (Goldman et al. 1994). Lectins are believed to be involved in the early stages of phytopathogen recognition, particularly in the formation of appresorium-like (hook-like) structures and the coiling of *Trichoderma* around the phytopathogen (Chet and Inbar 1994; Inbar and Chet 1992; Inbar and Chet 1995). These hook-like structures are able to penetrate the weakened fungal cell wall and the infection ultimately leads to the death of the fungus. The sequence of events leading to a mycoparasitic attack by *Trichoderma* is represented in Figure 3.

1.2.2 Cell wall-degrading enzymes of *Trichoderma*

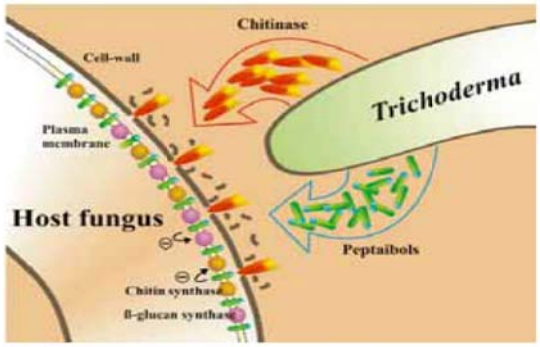
The cell wall of the majority of filamentous fungi consists primarily of chitin, β-glucans and proteins (Gooday 1995). The CWDEs of *Trichoderma* fall into three classes of enzymes: chitinases, glucanases, and proteases, which have been summarised in Table 2.

Chitinases

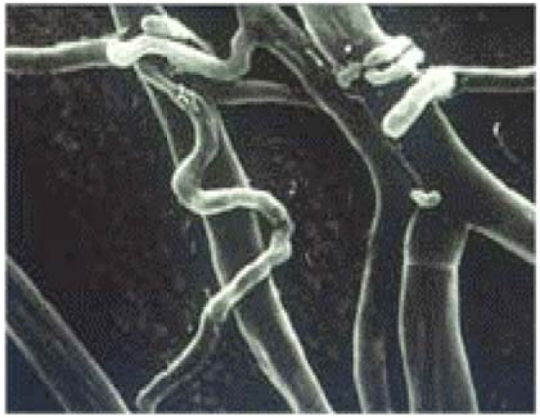
Chitinases degrade chitin, the second most abundant organic source in the natural environment. Chitin is a homopolymer composed of 1,4-β-linked N-acetyl-D-glucosamine (GlcNAc) units. Chitinases fall into three classes depending on the mechanism of chitin degradation: (1) endochitinases cleave randomly along the chitin fibril, (2) 1,4-β-N-acetylglucosaminidases cut the chitin polymer in an exo-type fashion into GlcNAc monomers and (3) exochitinases or chitobiosidases release only diacetylchitobiose units (reviewed in Viterbo et al. 2002a). Chitinases are grouped into glycosyl hydrolase families 18, 19 or 20 depending on amino acid sequence similarities. Most fungal endochitinases belong to family 18 whereas all N-acetylglucosaminidases belong to family 20 (Viterbo et al. 2002a). In a recent bioinformatic study of the *T. reesei* genome database, no members of family 19 chitinases were found when the *T. reesei* genome was compared to family 19 chitinases from *Hordeum vulgare* and *Encephalitozoon cuniculi* (Seidl et al. 2005). This may or may not be indicative of the presence of family 19 chitinases in other *Trichoderma* species as *T. reesei* is not known for its biocontrol abilities.



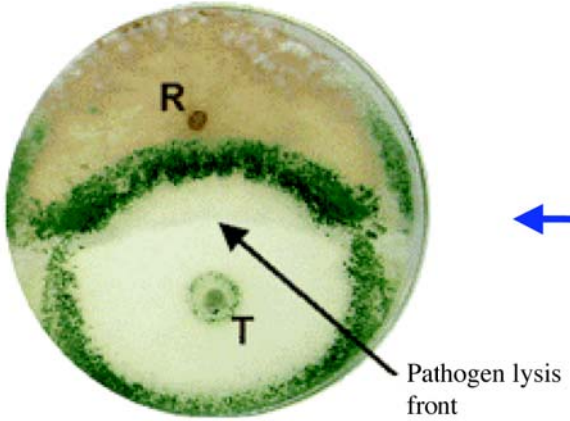
1) Recognition. An unknown signal is passed between *Trichoderma* spp. (T) and fungal pathogen, *Rhizoctonia* (R) to initiate the biocontrol response



2) Attack. *Trichoderma* releases CWDEs and antibiotics to weaken the pathogen cell wall structure (Photo: courtesy of International Subcommision of *Trichoderma* and *Hypocrea* Taxonomy)



3) Penetration. Lectin interactions are involved in the coiling, attachment and penetration of *Trichoderma* to the pathogen (Photo: courtesy of Chet et al. Weizmann Institute of Science)



4) Killing of the host. *Trichoderma* (T) overgrows and kills the pathogen (R) as indicated by the green spores at the edges of the plate (Photo: courtesy of Brunner et al. 2005)

Figure 3: Schematic of the mechanisms involved in the mycoparasitism of biological control mediated by *Trichoderma* spp.

Ech42 (endochitinase 42 kDa) has received most attention of the above-mentioned chitinases. Support for the role of Ech42 in the biocontrol response of *T. harzianum* began with research on its inhibitory effect on spore germination of the phytopathogen *B. cinerea*. Endochitinase alone could reduce the spore germination frequency up to 86-fold and even further when mixed with other chemical antifungal agents such as gliotoxin, flusilazole or miconazole (Lorito et al. 1994c). When grown under biological control inducing conditions containing 0.75 % colloidal chitin, Ech42 activity was increased 42-fold in strains containing multiple copies of the gene (Carsolio et al. 1999), compared to an Ech42 gene disruptant where no activity was observed. The expression of Ech42 was later thought to be triggered by an unknown diffusible substance from *R. solani* that was below 90 kDa in size (Kullnig et al. 2000). In a proteomic analysis of upregulated proteins when *T. atroviride* P1 was grown on the cell wall of *R. solani*, Ech42 was also identified as an up-regulated and secreted protein (Grinyer et al. 2005, Publication V, Section 3.2.1).

Table 2: Cell wall-degrading enzymes of *Trichoderma* spp.

Enzyme	**Properties that induce protein expression**	**Species / strain of *Trichoderma***	**Is the role of the enzyme confirmed during biocontrol?**	**References**
Chitinases				
Ech42 (Endochitinase 42 kDa)	Carbon and nitrogen starvation, physiological stresses, autolytic products	*T. atroviride* P1 and other *Trichoderma* spp.	Yes, expressed prior to confrontation	Zeilinger et al. 1999
Chit33 (Chitinase 33 kDa)	Carbon and nitrogen starvation, temperature stress	*T. harzianum* CECT 2413	Yes, expressed during pathogen overgrowth only	de la Mercedes et al. 2001
Chit36 (Chitinase 36 kDa)	Physiological stresses, colloidal chitin, N-acetyl glucosamine	*T. harzianum* *T. asperellum*	Yes, expressed prior to confrontation	Viterbo et al. 2001; Viterbo et al. 2002b
Nag1 (N-acetylglucosaminidase 1)	Chitin, N-acetyl glucosamine, fungal cell walls, nitrogen starvation with chitin	*T. atroviride* P1	Yes, expression after physical contact	Zeilinger et al. 1999
Exc2 (N-acetylglucosaminidase)	Glucosamine	*T. asperellum* *T. harzianum*	No	Ramot et al. 2004; Draborg et al. 1995
40 kDa chitobiosidase	Crab shell chitin	*T. atroviride* P1	No	Harman et al. 1993
Glucanases				
BGN13.1 (78 kDa β-1,3-glucosidase)	Fungal cell wall polymers, Autoclaved yeast cells or mycelia	*T. atroviride* P1 *T. harzianum*	Yes	de la Cruz et al. 1995a; Donzelli et al. 2001b
110 kDa β-1,3-exoglucanase (lam1.3)	Laminarin, co-culturing with fungal hosts	*T. harzianum*	No	Cohen-Kupiec et al. 1999

83 kDa β-1,3-exoglucanase	Laminarin, *R. solani* cell walls	*T. asperellum*	No	Bara et al. 2003
BGN16.1 (51 kDa β-1,6-endoglucanase	Yeast cell walls	*T. harzianum*	No	de la Cruz and Llobell 1999
BGN16.2 (43 kDa β-1,6-endoglucanase	Yeast cell walls	*T. harzianum*	No	de la Cruz et al. 1995b
Egl1 (β-1,4-endoglucanase)	*Pythium ultimum*	*T. longibrachiatum*	Yes	Migheli et al. 1998
Proteases				
Prb1 (basic serine protease 1)	*R. solani* cell wall	*T. harzianum*	Yes	Geremia et al. 1993; Flores et al. 1997
TVSP1 (*T. virens* serine protease 1)	*R. solani*	*T. virens*	Yes	Pozo et al. 2004
PRA1 (acidic serine protease)	*Meloidogyne incognita*, nitrogen and carbon starvation	*T. harzianum* CECT 2413	No	Suarez et al. 2004
PapA (extracellular aspartyl protease)	Organic nitrogen	*T. harzianum* *T. asperellum*	Yes	Delgado-Jarana et al. 2002
AP1 (aspartyl protease 1)	Cucumber seedling roots	*T. asperellum*	No	Viterbo et al. 2004
Vacuolar protease A	*R. solani* cell wall	*T. atroviride* P1	No	Grinyer et al. 2005
Vacuolar serine protease	*R. solani* cell wall	*T. atroviride* P1	No	Grinyer et al. 2005
Trypsin-like protease	*R. solani* cell wall	*T. atroviride* P1	No	Grinyer et al. 2005

Chitinase *chit33* encodes an endochitinase from *T. harzianum* involved in mycoparasitism. In a study into gene regulation, *chit33* was found to be weakly expressed during growth on chitin and *R. solani* cell wall material (de las Mercedes et al. 2001). Addition of N-acetylglucosamine also induced *chit33* expression for short periods of time. Expression of *chit33* was induced by either carbon or nitrogen starvation and to a lesser degree, under temperature stress (de las Mercedes et al. 2001). In a separate study, a cellulose binding domain (CBD) from *T. reesei* was engineered onto the *chit33* gene, producing a CBD-Chit33 hybrid. This modification resulted in the CBD-Chit33 protein having a stronger chitin binding capacity than the native Chit33 protein. Overexpression of CBD-Chit33 resulted in higher levels of chitinase and antifungal activities against *R. solani* when compared to overexpression of the native chitinase and the wild type strain (Limon et al. 2004). These results demonstrate the effectiveness of modifying a chitinase with a CBD for increasing hydrolytic activity towards chitin and fungal cell walls and enhancing the biocontrol capacity of the strain (Limon et al. 2004).

Chitinase Chit36 has recently been isolated and characterised from *T. harzianum* (Viterbo et al. 2001). The gene encoding *chit36* from *T. asperellum* is regulated by both glucose and nitrogen and is induced upon stressful conditions, as well as on colloidal chitin and N-acetylglucosamine (Viterbo et al. 2002b). Chit36 has a clear and efficient mycoparasitic effect: culture filtrates containing secreted Chit36 completely inhibited germination of *B. cinerea* spores (Viterbo et al. 2001). Direct contact with *R. solani* is not necessary for *chit36* expression. Instead, an unidentified *R. solani* soluble molecule, less than 12 kDa in size that was shown to be capable of diffusing through dialysis membrane is capable of inducing *chit36* expression (Viterbo et al. 2002b).

Three genes encoding N-acetylglucosaminidase have been isolated from *T. harzianum* and *T. atroviride* strains, including *nag1*, *exc1* and *exc2* (Peterbauer et al. 1996; Draborg et al. 1995). The Nag1 protein is 580 amino acids in length and the *nag1* gene is present as a single copy in the fungal genome. Nag1 protein expression is induced upon growth on chitin, N-acetylglucosamine and the cell walls of *B. cinerea* (Peterbauer et al. 1996). More recently, *nag1* induction was found to require a combination of both nitrogen starvation and the presence of chitin (Donzelli and Harman 2001a). Gene expression of *nag1* occurs after physical contact had been made between *R. solani* and *T. harzianum* P1 (Zeilinger et al. 1999). When a *nag1* disruption mutant was created that contained only 4 % N-acetylglucosaminidase activity compared to the wild type, it had 30 % reduced ability to protect beans against infection by *R. solani* and *Sclerotinia sclerotiorum* (Brunner et al. 2003).

Much less research has been conducted on *exc1* and *exc2* genes, although both N-acetylglucosaminidase-encoding genes originate from the same strain of *T. harzianum* (Draborg et al. 1995), indicating that more than one copy of the gene exists in some *Trichoderma* spp. The gene *exc1* encodes a 568 amino acid polypeptide (similar to Nag1), whereas Exc2 is 602 amino acids in length. When amino acid sequences of both Exc1 and Exc2 were aligned, 72 % and 59 % amino acid similarity was found between *T. harzianum* and *T. asperellum* protein homologs respectively (Draborg et al. 1995; Ramot et al. 2004). Both *exc1* and *exc2* genes from *T. asperellum* T203 are induced in the presence of glucosamine (Ramot et al. 2004).

While the expression of both the Ech42 and Nag1 chitinase proteins appears to have a clear role in the biological control mechanism of *Trichoderma* spp., there are conflicting reports highlighting an alternative role for Ech42. These reports suggest that Ech42 expression is triggered by stress-related factors rather than biological control-inducing factors such as fungal cell wall presence. Some stress-related factors that induce *ech42* expression include nitrogen starvation in combination with the presence of chitin (Donzelli and Harman 2001a), soluble chitooligosaccharides, other chitinolytic enzymes (Zeilinger et al. 1999), carbon starvation and physiological stresses (Mach et al. 1999). Mach et al. (1999) report that several copies of the putative stress response element, encoded by a nucleotide string of CCCCT, are found in the promoter of *ech42* and provide further evidence that stress is an expression trigger. It has been suggested that *ech42* may be induced by autolysis products of *Trichoderma* spp. (Seidl et al. 2005). It is clear that further research is required to uncover the true role of Ech42 for *Trichoderma* biocontrol.

The third class of chitinases produced by *Trichoderma* spp. are the chitobiosidases. A single chitobiosidase has been purified from *T. harzianum* P1 (Harman et al. 1993) that was expressed by growth on crab shell chitin and is 40 kDa in molecular mass. The protein contains N-linked glycosylation that accounts for approximately 5 kDa. When combined with the 42 kDa endochitinase, increased biocontrol activity was observed (Lorito et al. 1993).

Glucanases

The glucan structure of the fungal cell wall provides a substrate for fungal glucanases which play an integral role in cell wall hydrolysis. Glucanases are classified as either

β-1,3-glucanases if they hydrolyse the β-1,3-glucan (laminarin) structure of the fungal cell wall, β-1,4-glucanases for hydrolysing the 1,4-β-D glucan (cellulose) or β-1,6-glucanases for hydrolysing minor structural polymers of the cell wall. Beta-1,3-glucanases can be classified further depending on the mechanism of hydrolysis: endo-β-glucanases release oligosaccharides whereas exo-β-glucanases release monosaccharides.

A 78 kDa β-1,3-glucosidase has been isolated from many *Trichoderma* spp., including *T. atroviride* and *T. harzianum* (de la Cruz et al. 1995a; Donzelli et al. 2001b). Also termed BGN13.1, this glucanase has an endolytic mode of action. BGN13.1 is induced by fungal cell wall polymers or autoclaved yeast cells and mycelia but repressed by glucose (de la Cruz et al. 1995a; Donzelli et al. 2001b). A synergistic effect was observed when combinations of β-1,3-glucosidase, N-acetyl-β-glucosaminidase, endochitinase and 1,4-β-chitobiosidase were used to inhibit *B. cinerea* spore germination and germ tube elongation, with the highest level of antifungal activity being observed with a combination of all four enzymes (Lorito et al. 1994a). Beta-1,3-glucosidase expression was up-regulated and the enzyme secreted into the culture medium as shown in a proteomic study of biocontrol-related proteins when *T. atroviride* was grown on the cell wall of *R. solani* (Grinyer et al. 2005, Publication V, Section 3.2.1).

A β-1,3-exoglucanase has been isolated from *T. harzianum*. The glucanase is 110 kDa in molecular mass and its expression is induced in the presence of laminarin (β-1,3-glucan) or in dual cultures with *R. solani* and *Sclerotium rolfsii*, but it is repressed by glucose (Cohen-Kupiec et al. 1999). Another β-1,3-exoglucanase was purified from *T.*

asperellum (recently distinguished as a different species from *T. viride*) with a molecular mass of 83 kDa. This exoglucanase was active against laminarin and *R. solani* cell walls (Bara et al. 2003). Other β-1,3-exoglucanases have been found in *Trichoderma* including a 75 kDa enzyme from *T. harzianum* T-Y (Ramot et al. 2000). Smaller exoglucanases have also been isolated with masses of 29 kDa, 31 kDa and 40 kDa from different *T. harzianum* strains (Noronha and Ulhoa 2000; Kitamoto et al. 1987; Dubourdieu et al.1985).

Two β-1,6-endoglucanases, BGN16.1 and BGN16.2, have been isolated from *T. harzianum* CECT 2413. BGN16.1 is a 51 kDa basic enzyme that is un-glycosylated (de la Cruz J and Llobell 1999). BGN 16.2 has a mass of 43 kDa (de la Cruz et al. 1995b). No cross-reactivity was observed when antibodies raised against BGN16.2 were probed with BGN16.1, indicating these two enzymes originate from two separate genes. Both individual enzymes, BGN16.1 and BGN16.2 have activity against yeast cell walls, but no antifungal activity was observed when each enzyme was used alone. In combination with BGN13.1, BGN16.1 and BGN16.2, antifungal activity was observed (de la Cruz et al. 1995a; de la Cruz and Llobell 1999; de la Cruz et al. 1995b).

The β-1,4-glucanase family hydrolyses the homopolymer of 1,4-β-glucan (cellulose) and includes the 1,4-β-D-glucan cellobiohydrolases, endo-1,4-D-glucanases and 1,4-β-D-glucosidases. Transformants of *T. longibrachiatum* CECT2606 containing extra copies of the β-1,4-endoglucanase gene *egl1* provided significant protection to cucumber seedlings emerging in *Pythium ultimum*-infested soil. Emergence of

cucumber seedlings after treatment with transformant spore solution was significantly increased when compared to the wild-type strain (Migheli et al. 1998).

Proteases

Proteases cleave proteins into peptides from either the C- or N-terminus of the protein, classifying them either as exopeptidases or endopeptidases. Four classes of proteases exist, including serine proteases, cysteine proteases, aspartic proteases and metallo proteases (Rawlings and Barrett 1993). However, only two classes, serine proteases and aspartic proteases, appear to play a role in fungal biocontrol. Serine proteases consist of two families of enzymes, the chymotrypsin family including enzymes such as chymotrypsin, trypsin or elastase and the subtilisin family. The aspartic proteases typically include digestive enzymes and fungal proteases. Both families of enzymes contain similar active site geometry.

An alkaline serine protease was first linked to biocontrol mediated by *T. harzianum* when its expression was up-regulated on growth of autoclaved mycelia of *B. cinerea*, *T. viride*, *F. oxysporum* and *Sclerotina minor* or cell wall material of *R. solani*, and repressed when grown on glucose (Geremia et al. 1993). The purified serine protease, PRB1, has a molecular mass of 31 kDa and a pI of 9.2 (Geremia et al. 1993). When the copy number of *prb1* was increased in *T. harzianum* transformants, high expression levels of *prb1* were detected during interactions between *T. harzianum* and *R. solani* (Flores et al. 1997). *Prb1* was induced by *R. solani* cell walls and transformants containing extra *prb1* copies significantly reduced the disease caused by *R. solani* in cotton plants during greenhouse trials (Flores et al. 1997).

Recently, an acidic protease PRA1 of *T. harzianum* CECT 2413 was isolated and characterised. PRA1 has a mass of 28 kDa and an isoelectric point of 4.7-4.9. PRA1 is a serine protease containing trypsin-like activity. The *pra1* gene is induced by conditions simulating antagonism to the nematode *Meloidogyne incognita* and by nitrogen and carbon starvation at 4 h (Suarez et al. 2004). PRA1 has N-terminal sequence homology with another trypsin-like protease from *T. viride* (Uchikoba et al. 2001) and to a trypsin-like protease identified from *T. atroviride* (Grinyer et al. 2005, Publication V, Section 3.2.1).

Other serine proteases found to play a role in fungal biocontrol include the extracellular serine protease TVSP1 from *T. virens*. Conidiation, growth rate and extracellular protein secretion were unaffected when the *tvsp1* gene was overexpressed or knocked out from *T. virens*. However, when *tvsp1* was overexpressed, an increased ability to protect cotton seedlings from *R. solani* was observed (Pozo et al. 2004). High activities of two serine proteases from *T. harzianum* were observed after 12 h of induction on *Agaricus bisporus* (mushroom) cell walls (Williams et al. 2003). Chymoelastase activity was between 9.6 and 12.2-fold higher and trypsin-like activity was 4.9 and 6.3-fold higher in aggressive isolates compared to non-aggressive isolates. No difference in protease activity was observed between aggressive and non-aggressive isolates after 24 h of induction on *A. bisporus* cell walls (Williams et al. 2003). Two more serine proteases, a vacuolar serine protease and trypsin-like protease, were shown to be up-regulated in a proteomic study when *T. atroviride* was grown on the cell wall of *R. solani* (Grinyer et al. 2005, Publication V, Section 3.2.1). However, further work is necessary to determine if these proteases are secreted from *Trichoderma* during the biocontrol response.

More recently, aspartic proteases have been linked to *Trichoderma's* biocontrol capabilities. The gene *papA* encoding an extracellular aspartyl protease from *T. harzianum* was isolated. PapA protein has a predicted mass of 36.7 kDa after signal peptide cleavage and a pI of 4.35. The corresponding *papA* gene is regulated by pH, repressed by ammonium, glucose and glycerol and induced by organic nitrogen (Delgado-Jarana et al. 2002). Transformants overexpressing *papA*, had a four-fold increase in protease activity compared to the wild type strain (Delgado-Jarana et al. 2002). Two aspartyl proteases were identified when cucumber seedlings were inoculated with *T. asperellum* in a study of plant root colonisation. One aspartyl protease was found to be secreted into hydroponic medium during plant root colonisation and was homologous to PapA from *T. harzianum*. The second was identified through mRNA expression analysis and was suggested to be similar to AP1 from *Botryotinia fuckeliana* (Viterbo et al. 2004). Only *papA* was found to play a mycoparasitic role, as induction was observed when confronted with *R. solani* (Viterbo et al. 2004). Recently, the secretion proteome of *T. harzianum* CECT 2413 when grown on the cell walls of *R. solani*, *P. ultimum* and *B. cinerea* was shown to contain an aspartyl protease (P6281) as the predominant secreted protein. The gene sequence encoding this aspartyl protease has been characterised. Vacuolar Protease A was identified as an up-regulated protein by differential gel electrophoresis analysis (Grinyer et al. 2005, Publication V, Section 3.2.1) when *T. atroviride* was grown on the cell wall of *R. solani*. When the gene sequence of this protease was translated, it was shown that the predicted protease contained a conserved domain shared by other aspartyl proteases (Section 3.2.2). However, confirmation is required to determine if this protease is secreted during growth on the cell wall of *R. solani*.

1.2.3 Synergism of fungal biocontrol agents and enzymes

Synergism of biocontrol agents was studied when a combination of *Gliocladium virens* (sexual form of *T. virens*) and *Burkholderia cepacia* biomass was used to treat infected tomato and pepper seeds. The combination of fungi provided protection to pepper seeds against *R. solani*, *P. ultimum*, *S. rolfsii* and *F. oxysporum* that was similar to controls lacking the presence of phytopathogens (Mao et al. 1998). The combined application of *G. virens* and *B. cepacia* provided greater resistance to disease than when the seeds were treated with either fungus alone. This combination was also found to be efficient in small field trials, providing greater fresh weight yields for pepper plants and greater fruit yield for tomatoes (Mao et al. 1998).

The positive effects of the combined action of biocontrol-related enzymes and antibiotics have been known for some time. A synergistic inhibitory effect was observed on spore germination and germ tube elongation of *B. cinerea* when four enzymes from *T. harzianum* P1 were purified to homogeneity and combined, including 1,3-β-glucosidase, N-acetyl-β-glucosaminidase, an endochitinase and a chitin 1,4-β-chitobiosidase (Lorito et al. 1994a). Transgenic apple plants expressing a combination of fungal endochitinase and exochitinase exhibited better resistance against *Venturia inaequalis* than transgenic apple plants expressing only one of these enzymes, suggesting that fungal enzyme synergism also takes place *in planta* (Bolar et al. 2001). Synergistic interactions between fungal CWDEs and antifungal compounds have also been observed. When the endochitinase, chitobiosidase and glucanase from *T. harzianum* P1 were combined with chemical antifungal agents including gliotoxin, flusilazole, miconazole, captan and benomyl, a synergistic inhibitory effect was

observed to spore inhibition of *B. cinerea* (Lorito et al. 1994c). In addition, parallel formation of CWDEs including endochitinase, chitobiosidase, β-1,3-glucanase and antibiotics including trichorzianines A_1 and B_1 from *T. harzianum* were secreted into the culture medium during growth on *B. cinerea* cell walls. When the CWDEs and antibiotics were applied to *B. cinerea* spores in combination, a synergistic action inhibited spore germination at a greater rate than when compared to individual applications of CWDEs and antibiotics (Schirmbock et al. 1994).

1.2.4 Enhancing the biocontrol capabilities of *Trichoderma* spp.

There are a number of different possibilities for the use of *Trichoderma* spp. as biological control agents. First, *T. atroviride* can be applied to crops and / or soils during planting, as CWDEs produced by the fungus are known inhibitors of spore germination in phytopathogenic fungi (Lorito et al. 1994b). The application of *T. harzianum* spores to fruiting clusters of grape crops provided significant protection from *Botrytis* grape rot (Harman et al. 1996). Significant protection was also obtained when mixtures of the chemical fungicide iprodione and *T. harzianum* conidia were applied together (Harman et al. 1996). Biocontrol capabilities can be increased by transforming extra copies of genes encoding CWDEs into *Trichoderma* (Baek et al. 1999; Flores et al. 1997; Pozo et al. 2004; Viterbo et al. 2001).

Secondly, genes coding for proteins contributing to biological control can be cloned into crops, providing them with an inbuilt genetic resistance to particular pathogenic fungi. Towards this end, the *T. harzianum ech42* gene was inserted into tobacco and potato plants; it was shown that expression of the fungal *ech42* in different parts of the

plant provided protection from foliar and soil-borne fungal pathogens (Lorito et al. 1998). Expression of a combination of CWDEs in plants has been trialled with transgenic apple plants. Both the endochitinase *ech42* and exochitinase *nag70* genes were transformed into the plant separately and in combination and it was found that plants expressing both enzymes were more resistant to *V. inaequalis* (causal agent of apple scab) than plants expressing each enzyme alone (Bolar et al. 2001). In a separate study, three CWDE-encoding genes from *T. atroviride ech42*, *nag70* and *glu78* were transformed into apple plants individually or in combination. The highest amount of protection against the pathogens *R. solani* and *Magnaporthe grisea* was observed when endochitinase and exochitinase were co-expressed in transgenic plants, indicating that a synergistic effect was taking place (Liu et al. 2004). Current approaches include over-expression of genes encoding CWDEs in *Trichoderma* and transforming plants with these genes to provide environmentally friendly alternatives to the use of existing chemical fungicides. These two approaches would be considered in continuation of this work with the genes encoding proteins of relevance for biological control, found during the research.

However, either of the above approaches have not been able to out-compete or eliminate the use of chemical fungal pesticides. A recent approach to enhance the biocontrol ability of *T. atroviride* involves a completely different train of thought, which may prove to be more efficient than chemical fungicides. Glucose oxidase was found to be responsible for the biocontrol capabilities of *Talaromyces flavus* against phytopathogens including *Verticillium dahliae*, *S. sclerotiorum* and *R. solani* by inhibiting the growth of microsclerotia and hyphae (Murray et al. 1997; Stosz et al. 1996). Glucose oxidase catalyses the oxidation of D-glucose to D-glucono-1,5-lactone

and hydrogen peroxide. The formation of hydrogen peroxide greatly inhibits phytopathogen growth rate, thereby providing an alternative biocontrol route (Kim et al. 1988). While *T. atroviride* does not contain a glucose oxidase ortholog, the *goxA* gene (glucose oxidase) from *Aspergillus niger* was transformed into the former species under the *T. atroviride nag1* promoter in an attempt to increase the biocontrol efficiency of *T. atroviride*. The resulting transgenic strain of *T. atroviride*, called SJ3-4 was able to overgrow fungal pathogens more quickly than wild-type strains. It was noted that *goxA* expression occurred directly after phytopathogen contact and the glucose oxidase protein was secreted from the fungus (Brunner et al. 2005). Culture supernatant containing secreted glucose oxidase was able to inhibit *B. cinerea* spores to a three-fold greater extent than the *T. atroviride* P1 wild type strain. In plate confrontations, the transgenic strain overgrew phytopathogens *R. solani* and *P. ultimum* more quickly than the wild type strain (Brunner et al. 2005). This strain was also more effective at inducing systematic resistance in plants providing them with an inbuilt ability to protect themselves against foliar phytopathogens (Brunner et al. 2005).

There is little doubt that the application of transformant varieties of *Trichoderma* (containing multiple copies of genes encoding CWDEs or *goxA* varieties), along with recombinant plant varieties containing fungal genes encoding CWDEs will provide higher levels of control over fungal phytopathogens than current existing fungal biocontrol strains. Whether application of *Trichoderma* as a biocontrol agent has the ability to completely eliminate the need for chemical fungicides will only be determined as further research proceeds into engineering more efficient fungal biocontrol strains. Education of the general public will have an impact on the

acceptance of genetically engineered microorganisms for agriculture to achieve chemical free farming, the ultimate and environmentally beneficial goal.

Aims of the Thesis

1) Method development for extracting proteins from fungi, in particular *Trichoderma*, for their display by two-dimensional gel electrophoresis.
2) Identification of proteins from two-dimensional gels to produce reference maps of the *Trichoderma* proteome.
3) Application of differential gel electrophoresis to identify novel proteins involved in the biocontrol response of *Trichoderma atroviride*.
4) Isolation of genes encoding novel biocontrol-related proteins.

The Results and Discussion section of this thesis features work described in five publications (Sections 3.1.1, 3.1.2, 3.1.3, 3.1.4 and 3.2.1) supported by additional, unpublished information (Sections 3.1.5 and 3.2.2). The first four publications and the related additional section describe the development of proteomic methods for fungi, particularly *Trichoderma harzianum,* and include protein mapping and identification results. In the final publication the developed methods are applied to discover proteins involved in the biological control response of *T. atroviride* when grown on the cell wall of phytopathogen, *R. solani*. Two genes encoding proteins related to biological control were isolated from *T. atroviride* and are outlined in the final section discussing additional, unpublished information. Each publication will be briefly introduced and an overall summary presented at the conclusion of the Results and Discussion, in Section 3.3.

2. Materials and Methods

Materials and methods used in this work are summarised in Table 2a. Detailed descriptions describing each method can be found in the corresponding publication or section.

Table 2a: A list of the materials and methods used during this work.

Method	**Refer to publication / section**
Fungal strains and cultivation – *Trichoderma* spp.	II, III, V
Fungal strains and cultivation – *A. fumigatus* and *O. floccosum*	3.1.5
R. solani cell wall preparation	V
Whole cell protein isolation	I, II, V, 3.1.5
Enriched mitochondrial protein preparation	III
Preparation of membrane proteins	IV, 3.1.5
MCE fractionation and isolation of alkaline proteins	IV, VI
Precipitation of proteins from the culture supernatant	V
Sequential protein extraction	3.1.5
Isoelectric focusing and 2D PAGE	I, II, III, IV, V, 3.1.5
Matrix-assisted laser desorption / ionisation time-of-flight mass spectrometry	II, III, IV, V, VI, 3.1.5
Liquid chromatography mass spectrometry	II, III, V
Image analysis of 2DE gels	V, 3.1.5
Protein assay	V
N-terminal sequencing	V
Genomic DNA isolation	3.2.2
Designing consensus primers	3.2.2
PCR and genomic walking PCR	3.2.2
Transformation into TA cloning vector	3.2.2
DNA sequencing	3.2.2

3. Results and Discussion

3.1 Method development for the display and identification of fungal proteins by 2DE and mass spectrometry

This section comprises four publications which address the problems of displaying the fungal proteome by 2DE and highlights the benefits of simplifying the proteome for 2DE analysis. Identification of fungal proteins was carried out by mass spectrometry. Publication I addresses the development of a novel acid-extraction method for the extraction of yeast and fungal proteins for 2DE analysis. Acid-extraction is then used in mapping the proteome of the whole cell protein in *Trichoderma harzianum* as outlined in Publication II. This work also highlights the need for different combinations of mass spectrometry techniques to maximise protein identification from an organism of which a genome sequence has not been resolved. Publication III shows a process for simplifying the proteome of *T. harzianum* with an enriched mitochondrial fraction as an example. It also makes use of the knowledge obtained in Publication II showing that more than one identification method should be employed to maximise protein identification. Publication IV addresses several issues surrounding 2DE including the widely-reported lack of identifications typical for membrane proteins and low abundance proteins. The yeast, *Saccharomyces cerevisiae*, is used as an example in a process developed for the display and identification of membrane proteins and low abundance cellular proteins. The success of the procedure depends greatly on sufficient simplifications prior to 2DE. An additional chapter describing further subcellular protein extraction for 2DE and identification work carried out with *T. harzianum* and other filamentous fungal species follows Publication IV.

3.1.1 Improved 2-DE of microorganisms after acidic extraction (Publication I)

Optimisation of the protein extraction methodology prior to the start of a proteomic project revolving around 2DE will help to ensure that maximum information will be obtained from the study. Publication I presents a method for the extraction of proteins from a range of micro-organisms based on the use of an extraction solution with an acidic pH. Organisms from which proteins have successfully been extracted include gram negative and positive bacteria, yeast and fungi.

Publication I is largely the product of a working collaboration between myself, Ben Herbert and an Honours student, Mitchell Isaacs. I was predominantly involved with the work addressing protein extraction from *Saccharomyces cerevisiae* and *Trichoderma harzianum*. Mitchell produced the data relevant to the gram-positive bacterium *Bacillus subtilis*. I acknowledge the work of other co-authors involved in the extraction of proteins from *Escherichia coli* under the improved conditions described in this paper.

Publication I

Due to copyright laws, the following article has been omitted from this thesis. Please refer to the following link for the abstract details.

Herbert BR, Grinyer J, McCarthy JT, Isaacs M, Harry EJ, Nevalainen H, Traini MD, Hunt S, Schulz B, Laver M, Goodall AR, Packer J, Harry JL, Williams KL (2006) Improved 2-DE of microorganisms after acidic extraction. Electrophoresis 27:1630-1640

http://dx.doi.org/10.1002/elps.200500753

3.1.2 Fungal proteomics: initial mapping of biological control strain *Trichoderma harzianum* (Publication II)

This publication represents the first 2DE mapping effort of the proteome of the biological control fungus *Trichoderma harzianum*. The whole cell proteome of *T. harzianum* was displayed by 2DE and a combination of MALDI-TOF-MS and LC-MS/MS used for protein identifications. The production of a reference map was essential to determine the feasibility of a proteomics approach to identify proteins from an organism with limited protein and DNA sequences in existing databases.

I would like to acknowledge Matthew McKay for his skilful assistance in performing the LC-MS/MS and interpretation of MS/MS data. I also wish to acknowledge my supervisors Helena Nevalainen and Ben Herbert for their helpful advice and assistance in the preparation of the paper.

Publication II

Due to copyright laws, the following article has been omitted from this thesis. Please refer to the following link for the abstract details.

Grinyer J, McKay M, Nevalainen H, Herbert BR (2004a) Fungal proteomics: initial mapping of biological control strain *Trichoderma harzianum*. Curr Genet 45:163-169

http://dx.doi.org/10.1007/s00294-003-0474-4

3.1.3 Fungal proteomics: mapping the mitochondrial proteins of a *Trichoderma harzianum* strain applied for biological control (Publication III)

The work described in this publication begins to break down the proteome of *T. harzianum* into sub-cellular compartments, presenting a proteome of an enriched mitochondrial preparation. The mitochondrial preparation was produced by differential centrifugation of the cellular material. The resulting 2D map was produced and protein identifications were made using the combination of identification methods devised in Publication II. The outcomes begin to show the usefulness of sub-cellular fractionation for targeted protein display, which is further highlighted in Publication IV and in an additional chapter of sub-cellular proteomic results, Section 3.1.5.

I would like to acknowledge the skilful assistance of Matthew McKay in performing the LC-MS/MS and analysis of MS/MS data. I also wish to acknowledge my supervisors Helena Nevalainen and Ben Herbert for their helpful guidance and useful advice in the preparation of the publication.

Publication III

Due to copyright laws, the following article has been removed from this thesis. Please refer to the following link for abstract details.

Grinyer J, McKay M, Herbert B, Nevalainen H (2004) Fungal proteomics: mapping the mitochondrial proteins of a *Trichoderma harzianum* strain applied for biological control. Curr Genet 45:170-175

http://dx.doi.org/10.1007/s00294-003-0475-3

3.1.4 Unseen proteome: Mining below the tip of the iceberg to find low abundance and membrane proteins (Publication IV)

The quest to find elusive membrane proteins and low abundance proteins on 2D gels primed our interest in pursuing these proteins from *Saccharomyces cerevisiae* for which the genome sequence was available at the time of publication. The advantages of fractionating the proteome of an organism before 2DE analysis are discussed in Publications II and III and are highlighted again in this publication. Sub-cellular fractionation was carried out to isolate membrane proteins from *S. cerevisiae*. A pH fractionation on a multi-compartment electrolyser was undertaken to separate alkaline proteins, with a pI range of 7-10.5, from acidic proteins. Proteins in the alkaline fraction were then mapped. A large number of membrane-bound and membrane-associated proteins and a high number of low abundance proteins were identified. This publication highlights the importance of concentrating on a particular area of a proteome to minimise the effect of cellular contaminants and to obtain more specific information on the proteome of interest.

This publication consists of work predominantly carried out by Susanne Pedersen. My specific contribution involved analysis of the large amount of mass spectrometry data for protein identification displayed on the 2D maps shown in Figures 3 and 6, and more specifically in finding the splice variants described in Figure 7. Please note that my maiden name Jasmine Baker appears on this publication.

Publication IV

Due to copyright laws, the following article has been removed from this thesis. Please refer to the following link for abstract details.

Pedersen SK, Harry JL, Sebastian L, Baker J, Traini MD, McCarthy JT, Manoharan A, Wilkins MR, Gooley AA, Righetti PG, Packer NH, Williams KL, Herbert BR (2003) Unseen proteome: mining below the tip of the iceberg to find low abundance and membrane proteins. J Proteome Res 2:303-311

http://dx.doi.org/10.1021/pr025588i

3.1.5 Additional, unpublished information – Further method development and mapping of sub-proteome extracts of *Trichoderma harzianum*

Additional work not reported in Publications I-IV was carried out on sub-cellular proteomes obtained from the biocontrol fungus *T. harzianum*. Protein fractions investigated include Tris-soluble or cytoplasmic protein fractions, Tris-insoluble or cell wall-containing protein fractions and cell membrane protein fractions. The acid extraction methodology described in Publication I was also trialled on other filamentous fungi of general interest such as *Aspergillus fumigatus* and *Ophiostoma floccosum*.

The work described in this chapter is solely my own. I would like to thank the contributions made by my supervisors Helena Nevalainen and Ben Herbert for their ideas and approaches in reviewing this chapter.

Introduction

Trichoderma harzianum

Subcellular proteomes from *Trichoderma harzianum* were examined in addition to the whole cell proteome and mitochondrial protein extracts mapped and analysed in detail in previous work (Publication II and Publication III, Sections 3.1.2 and 3.1.3). Protocols were developed for extraction of Tris-soluble or cytoplasmic protein fractions, Tris-insoluble protein fractions, and cell envelope protein fractions. Tris-soluble protein fractions are expected to contain cytoplasmic proteins that are easily solubilised in a low salt environment whereas Tris-insoluble protein fractions should contain proteins that are more difficult to extract, such as those associated or bound to cellular membranes. The removal of cytoplasmic proteins from a cellular sample enables the Tris-insoluble protein fraction to be visualised without interference from these highly abundant proteins. The resulting two proteomic maps potentially double the number of proteins visualised by 2DE. Protein extraction methodologies will be outlined in this chapter and, where available, protein identifications from mass spectrometric analysis are also shown. The methods described in this chapter are new for fungal proteomic studies.

Aspergillus fumigatus and *Ophiostoma floccosum*

The acid extraction method developed in this work (Herbert et al. 2006, Publication I, Section 3.1.1) was tested with *Aspergillus fumigatus* and *Ophiostoma floccosum* to ensure that the methodology for extracting proteins from the fungal mycelia (whole cell) would work across different fungal species. Both fungal strains are of particular interest. *A. fumigatus* is an opportunistic human pathogen and is of medical relevance to patients with asthma and cystic fibrosis (Nikolaizik et al. 2002). *A. fumigatus* can

colonise the lungs and nasal tissue of patients with low immune systems. *O. floccosum* has been used for biocontrol purposes and is currently being developed as a host for expression of recombinant gene products (Gutierrez et al. 1999).

Methodology

Fungal strains and cultivation conditions

Trichoderma harzianum A3091 (IMI 206040) and *Aspergillus fumigatus* (NCPF 7367) were sporulated and maintained on Potato Dextrose agar (PDA; Difco Laboratories, Detroit, MI, USA). Spores were collected in 5-10 mL of 0.9 % sodium chloride (w/v), 0.01 % Tween-80. A volume containing 8 x 10^8 spores was used to inoculate 50 mL potato dextrose broth (Difco Laboratories, Detroit, MI, USA) in a 250 mL Erlenmeyer flask. The cultures were grown at 28 °C on a shaker at 250 rpm for 54 h or 96 h for *T. harzianum* and *A. fumigatus* respectively. Fungal and yeast protease inhibitor cocktail (0.05 % v/v, Sigma, St Louis, MO, USA) was added to the culture after completion and incubated at RT for 20 min. Fungal mycelia were collected by filtration of the cultures through Whatman 3MM paper (Whatman, Maidstone, England) and washed with 150-200 mL of Milli-Q (18.2 M Ohms) water. Mycelia were either used immediately or stored at –20 °C until use.

Ophiostoma floccosum J2026 5.1 (Wu et al. 2006) was maintained on PDA. Spores were collected in 5 mL of 0.9 % (w/v) sodium chloride, 0.01 % (v/v) Tween-80. The amount of 1-2 x 10^8 conidia was used to inoculate 50 mL minimal medium (MM; 110 mM potassium phosphate, 38 mM ammonium sulphate, 2.4 mM magnesium sulphate, 4.1 mM calcium chloride, 2.9 mM manganese suphate, 7.2 mM iron sulphate, 0.35 mM zinc suphate, 0.71 mM cobalt sulphate, pH 6.5) with 3 % (w/v) starch and 1.5 % (w/v) soybean flour in a 250 mL Erlenmeyer flask. Cultures were grown at 23-28 °C on a shaker at 250 rpm for 1-5 days. The mycelia were washed three times with 50 mL Milli-Q water by inverting and centrifugation at 4000 g for 10 min at 15 °C and stored at -20 °C until use.

Preparation of protein extracts

Sequential extraction – Tris-soluble/cytoplasmic protein fraction

Approximately 0.5 g of mycelia (wet weight) was resuspended in 10 mL of 40 mM Tris-HCl, pH 7.4, 1 mM PMSF, 0.1 % (v/v) protease inhibitor cocktail. Mycelia were lysed by sonicating at 60 % intensity for 8 cycles of 15 s on ice. The suspension was spun at 14000 g for 20 min at 4 °C. The supernatant was collected and the pellet was re-extracted as above. The supernatants from the two extractions were pooled and the pellet was stored and used for preparation of the Tris-insoluble protein fraction. The combined supernatants were precipitated with 9 volumes of RT acetone for 15 min at RT and the precipitated protein was pelleted by centrifuging at 2500 g for 15 min at 20 °C. The protein pellet was resuspended in 4 mL of sample solution (7 M urea, 2 M thiourea, 1 % C7BzO, 40 mM Tris, 5 mM tributylphosphine, 10 mM acrylamide, 1 mM PMSF and 0.1 % (v/v) protease inhibitor cocktail) and incubated at RT for 90 min to allow complete reduction and alkylation of proteins. These reactions were quenched with 10 mM dithiothreitol before insoluble material was removed by spinning at 21000 g for 10 min. Samples were prepared at RT to prevent urea from precipitating out of solution, which occurs at temperatures below 15 °C. The sample was used directly to rehydrate IPG strips or stored at –20 °C until required.

Sequential extraction – Tris-insoluble protein fraction

The pellet from the soluble protein fraction was re-extracted three times by sonicating at 60 % intensity for 8 cycles of 15 s on ice with 10 mL of 40 mM Tris-HCl, pH 7.4, 1 mM PMSF, 0.1 % (v/v) protease inhibitor cocktail. The sample was spun at 14000 g for 20 min at 4 °C to minimise any soluble protein contaminants. The remaining pellet was resuspended in 2 mL of sample solution (7 M urea, 2 M thiourea, 1 % C7BzO, 40

mM Tris, 5 mM tributylphosphine, 10 mM acrylamide, 1 mM PMSF and 0.1 % (v/v) protease inhibitor cocktail) and the sample was reduced and alkylated as described above.

Cell membrane protein fraction

Approximately 1 g of mycelia (wet weight) was resuspended in 10 mL of extraction solution (100 mM sodium carbonate, 1 mM PMSF, 0.1 % (v/v) protease inhibitor cocktail). The sample was ultrasonicated at 60 % intensity for 6 cycles of 30 s and kept cool on ice before adding another 90 mL of the extraction solution. The sample was incubated on ice with stirring for 90 min before the cell debris was pelleted by spinning at 2500 g for 10 min at 4 °C. The supernatant was taken and spun at 115000 g for 70 min at 4 °C to pellet cellular membranes. The pellet was washed twice in 40 mM Tris-HCl, pH 7.4, 1 mM PMSF, 0.1 % (v/v) protease inhibitor cocktail before solubilising in 2.5 mL of sample solution (7 M urea, 2 M thiourea, 1 % C7BzO, 40 mM Tris, 5 mM tributylphosphine, 10 mM acrylamide, 1 mM PMSF and 0.1 % (v/v) protease inhibitor cocktail). The sample was reduced and alkylated as described above.

Whole cell protein preparation (*Aspergillus* and *Ophiostoma*)

Approximately 0.5-1 g of mycelia (wet weight) was resuspended in 10 mL of protein extraction solution (7 M urea, 2 M thiourea, 1 % C7BzO (w/v), 80 mM citric acid, 5 mM tributylphosphine, 1 mM PMSF and 0.1 % (v/v) protease inhibitor cocktail). The mycelia were lysed by sonicating at 60 % intensity for six cycles of 30 s on a Branson Sonifier 250 (Branson Ultrasonic, CT, USA). The samples were kept on ice throughout the sonication cycles. Cellular debris was pelleted by centrifuging at 38000 g for 20 min at 20 °C and the supernatant collected and precipitated with 9 volumes of RT

acetone for 15 min at RT. Precipitated proteins were pelleted by centrifuging at 2500 g for 15 min at 20 °C. They were resuspended in 4 mL of sample solution (7 M urea, 2 M thiourea, 1% C7BzO (w/v), 40 mM Tris, 5 mM tributylphosphine, 10 mM acrylamide, 1 mM PMSF and 0.1 % (v/v) protease inhibitor cocktail) and the sample was reduced and alkylated as defined above.

Isoelectric focusing and 2D PAGE

The samples were used directly to passively rehydrate 3-10 and 4-7, 11 cm IPG strips (Amersham Pharmacia, Uppsala, Sweden) by applying 180 μL of each sample to them. The soluble extract of the sequential extraction was diluted one to one with sample buffer before IPG rehydration. IPG's were focused to 90,000 Volt hours using a three-step focusing program. The program included a rapid ramp to 300 Volts for 4 h, a linear ramp to 10,000 Volts over 8 h, and a 10,000 Volt step until the Volt hours were reached. IPGs were equilibrated for 20 min in 6 M urea, 2 % (w/v) SDS, 50 mM Tris-acetate buffer, pH 7.0, 0.1 % (w/v) bromophenol blue and then were placed on top of a Proteome Systems 6-15 % SDS PAGE gelchips (Proteome Systems, Sydney, Australia) and run at 50 milliAmps constant until the blue dye reached the bottom of the gel. Gels were fixed in 10 % (v/v) methanol, 7 % (v/v) acetic acid solution for 30 min, then stained with Sypro Ruby solution (Molecular Probes) for 16 h. They were destained in the fixing solution before scanning on a fluorescence scanner (Alpha Innotech Corporation, California). Gels were restained with Coomassie colloidal blue G250 (17 % (w/v) ammonium sulphate, 34 % (v/v) methanol, 3.6 % (v/v) orthophosphoric acid, 0.1 % (w/v) Coomassie G-250) for further analysis.

Image analysis

Spot detection and spot counting were conducted using the Proteome Systems ImagepIQ software package. Spots were detected using default parameters. However, changing spot intensity and contrast levels to a value of 4 was undertaken to intensify spot appearance.

MALDI-TOF-MS

Protein spots were detected and excised using the Xcise^{TM} apparatus (Shimadzu Biotech, Japan). Excised gel pieces were washed three times with 100 µL 50 mM ammonium bicarbonate, pH 8.2, 50 % (v/v) acetonitrile and dried using a SpeedVac centrifuge for 20 min. Trypsin in 50 mM ammonium bicarbonate (20 µg mL^{-1}) was added to each gel piece and the samples incubated at 30 °C for 16 h. The peptides were extracted by sonication and automatically desalted and concentrated using ZipTips^{TM} from Millipore (Bedford, MA) in a Map II/8 (Bruker-Daltonik GmbH, Germany) liquid handling unit and spotted onto the Axima MALDI target plate with 1.5 µL alpha-cyano-4-hydroxy-cinnamic acid. Peptide mass fingerprints of tryptic peptides were generated by MALDI-TOF-MS using an Axima CFR (Kratos, Manchester, UK). Peptide mass fingerprints were searched against proteins from all fungal species including yeasts using Mascot Peptide Mass Fingerprint program where a modified MOWSE scoring algorithm was used to rank results (http://www.matrixscience.com/help/scoring_help.html).

Results and discussion

Sequential extraction of *T. harzianum* proteins

Soluble proteins were extracted using a solution containing Tris-base as described earlier. Contamination of soluble proteins in further sequential extractions was minimised by repeating the extraction procedure on the remaining pellet. The Tris-insoluble proteins in the pellet were extracted using a combination of urea, thiourea and chaotrophic detergents to give a hydrophobic protein-containing fraction as described earlier. The Tris-soluble extract of *T. harzianum* resulted in 620 spots being resolved across a 3-10, 11 cm pH gradient (Figure 4a) and 818 spots on a 4-7, 11 cm pH gradient (Figure 4b). The Tris-insoluble extract (containing cell envelope-enriched proteins) resulted in the detection of 580 protein spots on a 3-10 (Figure 4c), 11 cm pH gradient, and 815 protein spots on a 4-7, 11 cm pH gradient (Figure 4d).

From the 2D gel shown in Figure 4b, 95 of the darkest spots were excised and prepared for mass spectrometry. Eighteen of these 95 spots were confidently identified and are labelled in Figure 5, giving an identification frequency of approximately 19 %. Many of the identified proteins are known to be located in the cytoplasm of the fungal mycelia, for example, enolase, eukaryotic initiation factor 4a, actin, ubiquitin, phosphoglycerate kinase, glyceraldehyde-3-phosphate dehydrogenase and pyruvate kinase. Several mitochondrial proteins were also identified, indicating that during cell lysis, the mitochondrial wall was damaged and the mitochondrial contents were released and mixed with the cytoplasm. The damage inflicted upon organelles including the mitochondria was predictable as the ultrasonic probe is a highly efficient device that is designed to cause cell lysis and membrane rupture. The abundant mitochondrial proteins were identified as ATP synthase beta chain, probable heat

shock protein 60 and 6-phosphogluconate dehydrogenase. Further information in regard to each of these protein identifications is shown in Table 3, which also features the number of peptides matched to each of the proteins and the amino acid coverage obtained by mass spectrometry. As determined in earlier chapters (Publications II and III, Sections 3.1.2 and 3.1.3), more information could have been obtained if an MS/MS approach (peptide sequencing) was combined with the above MALDI-TOF-MS (peptide mapping) approach. However, these resources were not available at the time of analysis.

From the 2D gel shown in Figure 4d, 95 of the darkest spots were excised and prepared for mass spectrometry. Thirteen of the spots were confidently identified by mass spectrometry, giving an identification frequency of approximately 14 %. Proteins identified from the gel shown in Figure 6 include mitochondrial proteins, proteins from the nucleus and some highly abundant cytoplasmic proteins that still remain after several attempts of washing with 40 mM Tris buffer to minimise their presence. The mitochondrial proteins identified include several heat shock proteins and ATP synthase beta-chain. The small G-protein Gsp1p is a nuclear protein and the activation mediator subcomplex of RNA polymerase II is a transcription regulator that is also involved in glucose repression (Myer and Young 1998). The cytoplasmic proteins identified are all highly abundant proteins and include actin, glyceraldehyde-3-phosphate dehydrogenase and phosphoglycerate kinase. Some of these proteins are known to co-localise to cellular membranes, as discussed in the section on membrane protein extract of *T. harzianum* below. The absence of a genome database and hence restricted opportunity to identify protein sequences for *Trichoderma spp*. was a major limiting factor to the success of identification of Tris-soluble and Tris-insoluble proteins as

discussed previously in Section 1.1.2 and 1.1.3. Table 4 provides more information in regards to the each of the proteins identified.

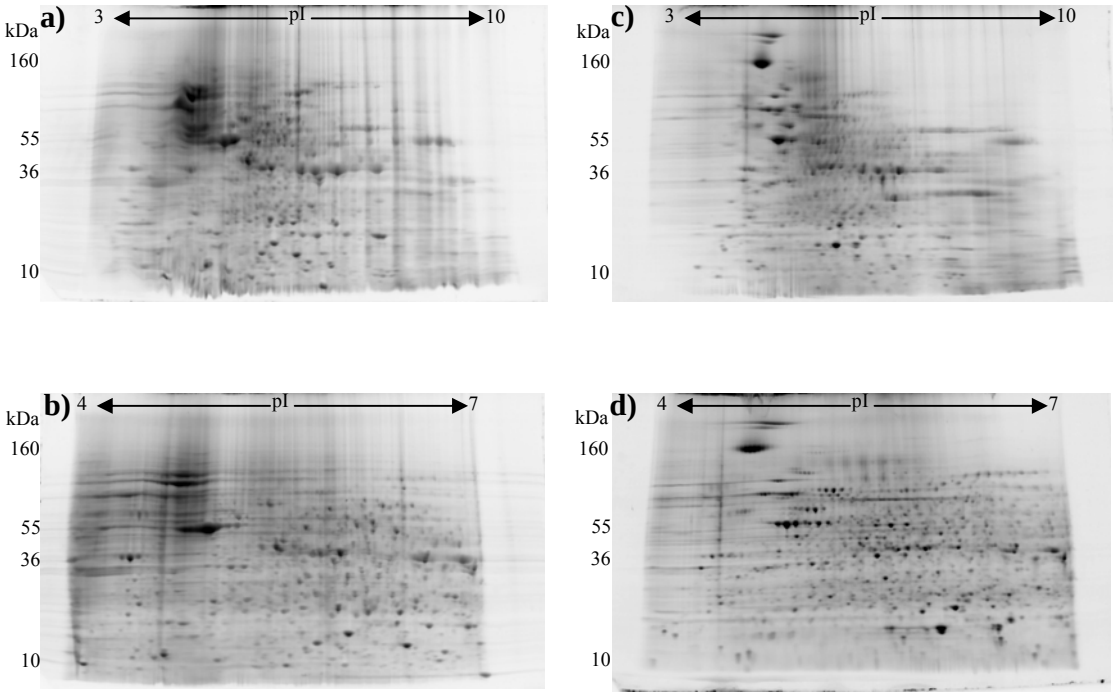


Figure 4

Sypro Ruby-stained 2D maps of proteins from the sequential extraction of *T. harzianum*. The soluble extract is displayed on 11 cm 3-10 (a) and 4-7 (b) IPG strips in the first dimension and the less-soluble extract is displayed on 11 cm 3-10 (c) and 4-7 (d) IPG strips in the first dimension.

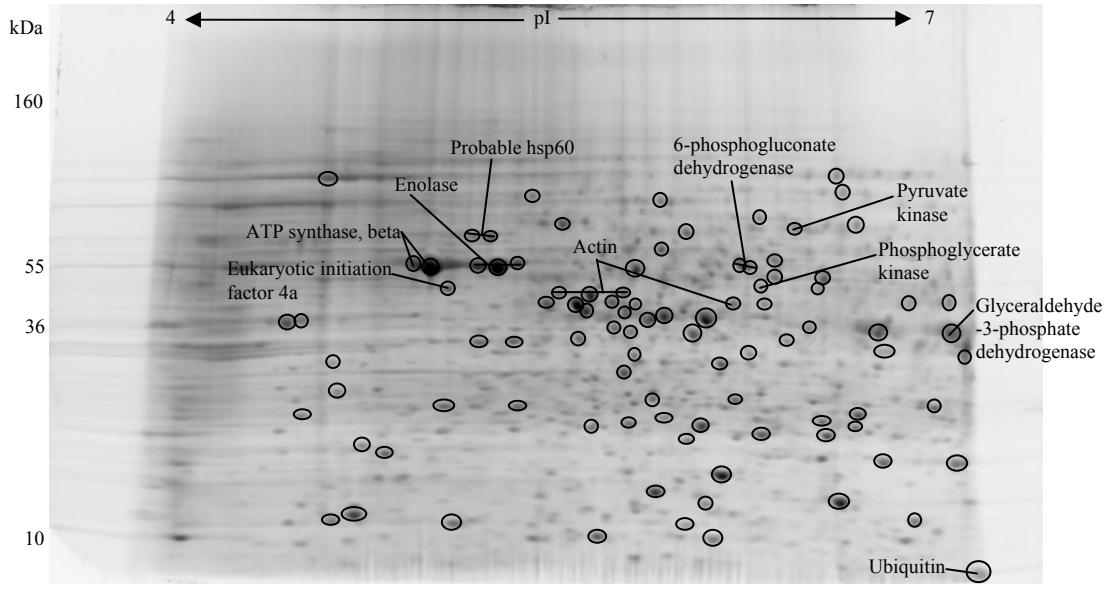


Figure 5

Protein identifications made by MALDI-TOF-MS from the Tris-soluble extract of the sequential protein extraction procedure (gel b from Figure 4) from *T. harzianum*.

Table 3: Identified proteins from the Tris-soluble fraction of a sequential protein extraction of *Trichoderma harzianum* by MALDI-TOF mass spectrometry.

Protein name	Species	Accession no.	MW (Da)	pI	No. of peptides matched	Coverage
ATP synthase β-chain	*Neurospora crassa*	gil114555	55499	5.15	12-15	24-27 %
Enolase	*Aspergillus oryzae*	gil3023683	47377	5.46	10-13	31-38 %
Probable heat shock protein hsp60	*N. crassa*	gil11277955	60452	5.6	15-19	25-30 %
Eukaryotic initiation factor 4a	*Schizosaccharomyces pombe*	gil19115766	44408	4.86	12	23 %
Actin	*Botryotinia fluckeliana, Humicola grisea var thermoidea, N. crassa*	gil3182891 gil4249564 gil2072156	41574-41622	5.38-5.56	12-18	36-49 %
Ubiquitin	*Saccharomyces cerevisiae*	gil136678	8552	6.56	13	89 %
6-phosphogluconate dehydrogenase	*A. oryzae*	gil21954534	54182	5.82	8-14	19-35 %
Phosphoglycerate kinase	*Trichoderma viride*	gil400757	44353	6.16	14	43 %
Glyceraldehyde-3-phosphate dehydrogenase	*T.harzianum*	gil2494643	36177	6.97	15	49 %
Pyruvate kinase	*Hypocrea jecorina*	gil400142	58851	6.2	16	21 %

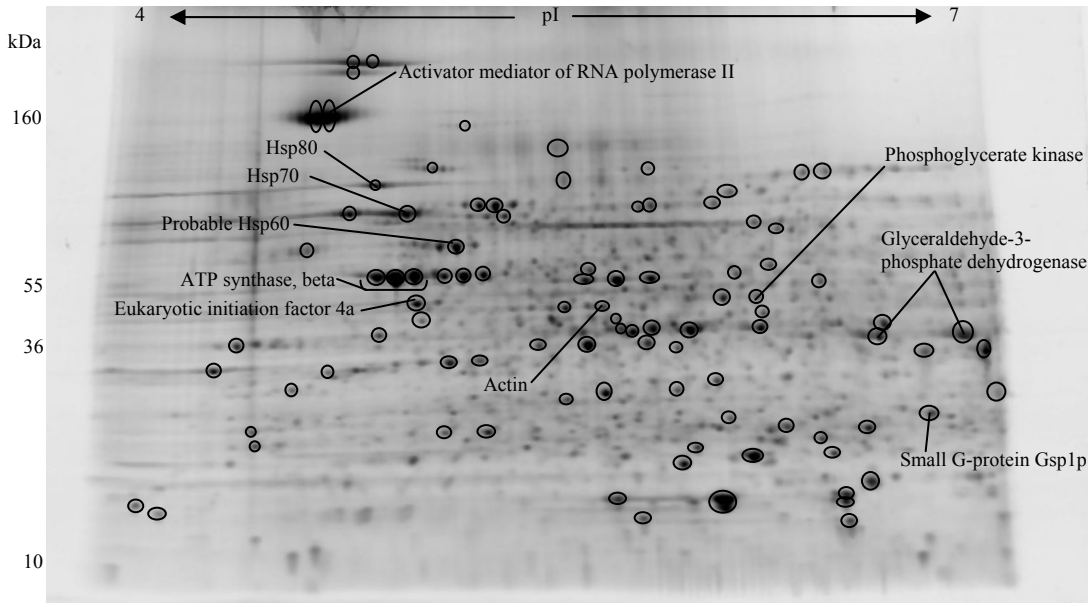


Figure 6

Protein identifications made by MALDI-TOF-MS from the Tris-insoluble extract of the sequential protein extraction procedure (gel d from Figure 4) from *T. harzianum*.

Table 4: Identified proteins from the Tris-insoluble fraction of a sequential protein extraction of *Trichoderma harzianum* by MALDI-TOF mass spectrometry.

Protein name	**Species**	**Accession no.**	**MW (Da)**	**pI**	**No. of peptides matched**	**Coverage**
Activation mediator subcomplex of RNA polymerase II	*Saccharomyces cerevisiae*	gil6319923	166754	6.06	15	10 %
Heat shock protein 80	*Neurospora crassa*	gil12718221	80139	5	12	16 %
Heat shock protein 70	*N. crassa*	gil3929355	70558	5.03	19	29 %
Probable heat shock protein hsp60	*N. crassa*	gil11277955	60452	5.6	21	29 %
ATP synthase β-chain	*N. crassa*	gil114555	55499	5.15	12-15	24-32 %
Eukaryotic initiation factor 4a	*Schizosaccharomyces pombe*	gil19115766	44408	4.86	9	22 %
Actin	*N. crassa*	gil2072156	41622	5.56	7	21 %
Small G-protein Gsp1p	*Candida albicans*	gil8698689	24275	6.44	11	40 %
Glyceraldehyde-3-phosphate dehydrogenase	*Trichoderma harzianum*	gil2494643	36177	6.97	14	45-46 %
Phosphoglycerate kinase	*T. viride*	gil400757	44353	6.16	10	35 %

Membrane protein extract of *T. harzianum*

Cellular membranes were first isolated using a sodium carbonate solution before extracting the proteins in a hydrophobic solution. The high molarity of urea (7 M) mixed with the more hydrophobic thiourea (2 M) and strong detergents used in our extraction solutions helped to solubilise cell envelope proteins from *T. harzianum*. A total of 587 protein spots were identified on an 11 cm 3-10 pH gradient (Figure 7a) and 788 protein spots on an 11 cm, 4-7 gradient (Figure 7b) from the cell envelopes of *T. harzianum*.

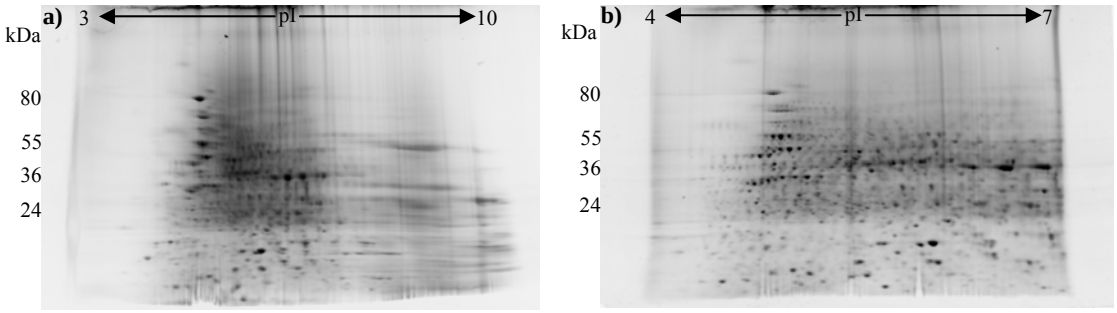


Figure 7

Sypro Ruby-stained 2D maps of the cell envelope fraction of *T. harzianum* run on 3-10 (a) and 4-7 (b), 11 cm IPG strips in the first dimension.

From the 2D gel shown in Figure 7b, 95 of the darkest spots were excised from the gel and prepared for mass spectrometry. Twenty nine spots were identified, giving an identification frequency of 30.5 % (Table 5). The absence of a genome database and hence protein sequences for *Trichoderma spp*. was a major contributing factor to the limited success of protein identifications from the membrane fraction, as discussed previously in Section 1.1.2 and 1.1.3. Many individual spots were identified as fragments of ATP synthase beta chain (Figure 8), indicating that some proteolysis had occurred during the protein preparation even with the presence of fungal protease inhibitors and PMSF throughout the procedure. Proteins identified from the membrane preparation include several mitochondrial proteins including heat shock proteins and ATP synthase beta chain, several cytoplasmic proteins including eukaryotic initiation factor 4a, 14-3-3-like protein, actin and glyceraldehyde-3-phosphate dehydrogenase, one nuclear protein named small G-protein Gsp1p and one ribosomal protein named 60S ribosomal protein P0. In the past, the sodium carbonate extraction procedure has been used successfully to isolate membrane bound proteins from *E. coli* (Molloy et al. 2000). More recently, the method has been applied successfully in purification of an enriched membrane-associated or membrane-bound protein fraction from *S. cerevisiae* (Publication IV, Section 3.1.4). In *S. cerevisiae*, 49 % of the 780 spots analysed by mass spectrometry were identified as membrane-bound or membrane-associated. The above results indicate that membrane proteins should be present on the 2D gel shown in Figure 8.

When a more thorough literature search of the proteins identified from this membrane extract was conducted, many of the so-called “cytoplasmic” proteins were found to co-localise with the plasma or organelle membranes, explaining their appearance on the

2D gel of the membrane preparation. Glyceraldehyde-3-phosphate dehydrogenase does associate with plasma membranes in many organisms (Laschett et al. 2004; Daubenberger et al. 2003; Villamon et al. 2003), indicating more than one function for this protein. Similarly, 14-3-3-like protein has also been found to associate with plasma and organelle membranes in plants (Pertl et al. 2005), bacteria (Assossou et al. 2001) and in a green alga (Voigt et al. 2001). Actin interacts with annexin 2 which belongs to a family of peripheral membrane-binding proteins involved in membrane organisation and membrane traffic in humans (Rescher et al. 2004; Gerke and Moss 2002). Hence actin can be re-classified as a membrane-associating protein. In a separate study using immunoelectron microscopy, actin from yeast was shown to associate on invaginations of the plasma membrane (Mulholland et al. 1994). ATP synthase beta-chain is located in the mitochondria, and when complexed to F(1)F(0) ATPase, associates with the mitochondrial inner membrane to convert ATP to ADP in the presence of a proton gradient across a membrane (Gavin et al. 2004). Gsp1p governs the interaction of importins and exportins with nucleoporins, aiding transport of materials from the nucleus of *S. cerevisiae* and has been shown to directly interact with Nup60p, which is a component of the nucleoporin complex (Denning et al. 2001). The cited evidence shows that many "cytoplasmic" and organellar proteins indeed can have more than one function and can co-localise in different parts of the cell at any one time. If the above evidence is taken into account, of the nine gene products identified from the membrane extract shown in Figure 8 and Table 5, six (or 67 % of the gene products) can also co-localise with a cellular membrane.

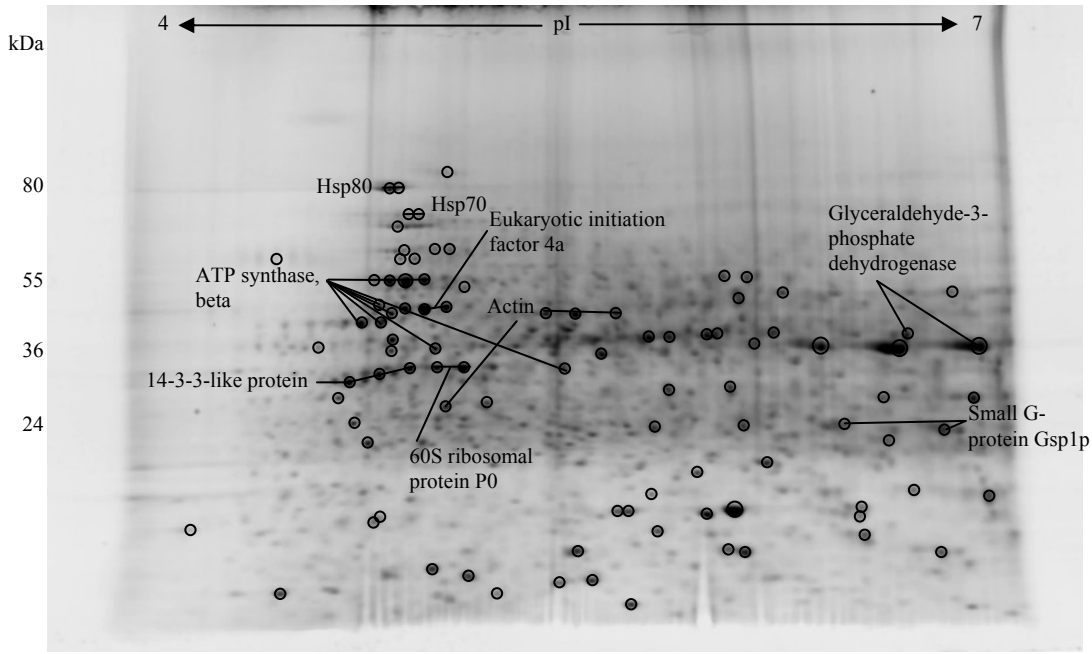


Figure 8

Protein identifications made by MALDI-TOF-MS from the membrane extract of *T. harzianum*.

Table 5: Identified proteins from the membrane preparation of *Trichoderma harzianum* by MALDI-TOF mass spectrometry.

Protein name	**Species**	**Accession no.**	**MW (Da)**	**pI**	**No. of peptides matched**	**Coverage**
Heat shock protein 80	*Neurospora crassa*	gil6979704	78860	5.12	10-12	15-16 %
Heat shock 70 kDa protein	*N. crassa*	gil3929355	70558	5.03	10-13	20 %
ATP synthase β-chain	*N. crassa*	gil114555	55499	5.15	8-16	17-37 %
Eukaryotic initiation factor 4a	*Schizosaccharomyces pombe*	gil19115766	44408	4.86	9-11	20-30 %
60S ribosomal protein	*N. crassa*	gil13899020	33513	4.85	8-10	25-41 %
14-3-3-like protein	*Hypocrea jecorina*	gil12054276	30404	4.89	7-12	28-45 %
Actin	*H. jecorina, Humicola grisea var thermoidea, N. crassa*	gil2492674, gil4249564, gil2072156	41718	5.43	12-19	36-55 %
Glyceraldehyde-3-phosphate dehydrogenase	*Trichoderma harzianum*	gil2494643	36177	6.97	10-20	56 %
Small G-protein Gsp1p	*Candida albicans*	gil8698689	24275	6.44	7-10	24-33 %

Whole cell protein extract of *Aspergillus* and *Ophiostoma*

The acid extraction protocol developed for *T. harzianum* was shown to work for the bakers' yeast, *S. cerevisiae* and also for bacteria including *E. coli* and *B. subtilis* as outlined in Publication I (Section 3.1.1). However, this method was trialled using *A. fumigatus* and *O. floccosum* to ensure these results were consistent across other filamentous fungal species. The results are shown in Figures 9 and 10. The acid extraction procedure worked well for both fungal organisms with minimal streaking appearing in the acidic portion of the 2D gels. The acidic pH of the protein extraction solution minimised the solubility of chitin and β-1,3-glucan from the cell wall, thereby reducing the streaking commonly caused by these polysaccharides (Herbert et al. 2006, Publication I, Section 3.1.1). The resulting clean 2D gels are suitable for further analysis of the separated proteins.

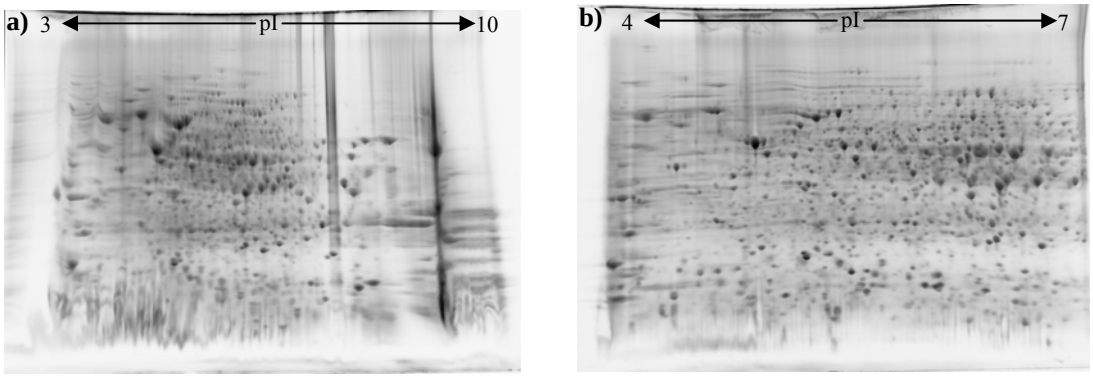


Figure 9

Sypro Ruby-stained 2D gels of the whole cell extract of *A. fumigatus* run on 3-10 (a) and 4-7 (b), 11 cm IPG strips in the first dimension.

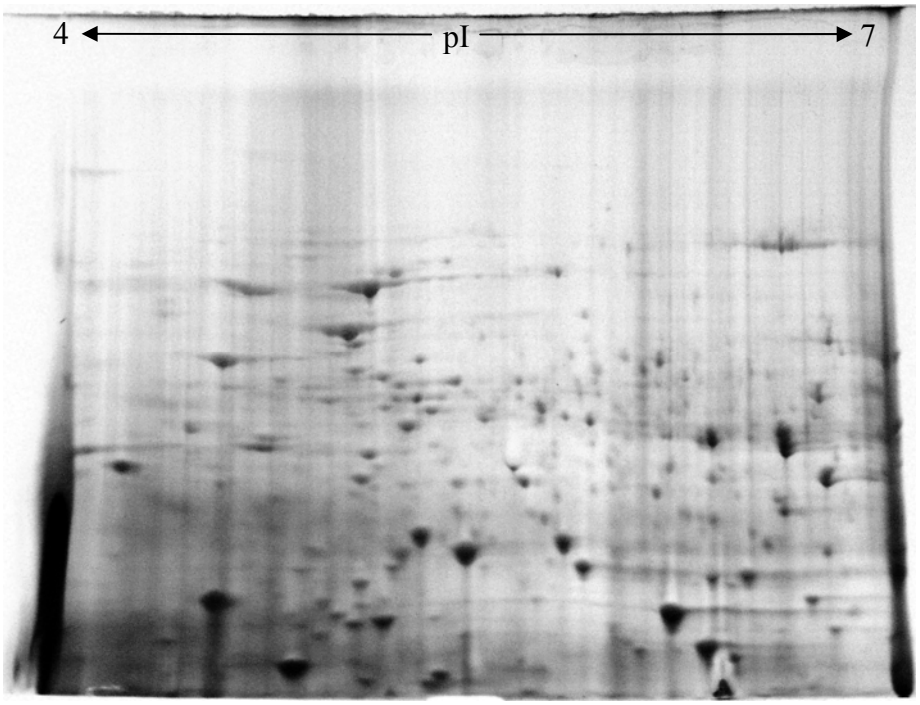


Figure 10

Sypro Ruby-stained 2D gels of the whole cell extract of *O. floccosum* run on a 4-7, 11 cm IPG strip in the first dimension.

Protein redundancies

Identification of the same protein from different sub-cellular proteome maps, protein redundancies, is a common occurrence. For example, proteins found in the whole cell preparation of *T. harzianum* should also be observed in at least one of the extracts from the sequential extraction (Tris-soluble or Tris-insoluble extracts). Some protein overlap is also expected from the gels representing the Tris-insoluble fraction and the membrane fractions. Protein redundancies have been highlighted in Table 6 below. It should be noted that MALDI-TOF-MS was used alone to identify the selected proteins from Tris-soluble, Tris-insoluble and membrane extracts. However, in the whole-cell and enriched mitochondrial extracts, a combination of MALDI-TOF-MS and LC-MS/MS was used to provide a more thorough sub-proteome map. From all of the protein extracts shown in Table 6, only two proteins, glyceraldehyde-3-phosphate dehydrogenase and ATP synthase beta chain are present across all cellular extracts from *T. harzianum*. Both are highly abundant, highly soluble proteins that are generally located in the cytoplasm but also have some association with the cellular membrane. This fact is particularly true for both glyceraldehyde-3-phosphate dehydrogenase and ATP synthase beta chain, which have dual cellular locations, as discussed in the section on membrane extracts from *T. harzianum*. Several proteins were identified in four of the five different protein extracts and include some of the heat shock proteins and actin. Again, the heat shock proteins are primarily located in cellular organelles, but are highly soluble and are most likely released from the organelles during ultrasonic treatment. Therefore, their presence in cytoplasmic preparations is explained. Actin has also been found to co-localise with cellular membranes, explaining its presence in membrane-enriched fractions.

Table 6: Spot identifications compared across protein extraction techniques applied to in this work.

Protein name	WC	EM	TS	TI	M
ATP synthase β-chain	Y	Y	Y	Y	Y
Probable heat shock protein 60	Y	Y	Y	Y	-
Heat shock protein 60	Y	-	-	-	-
Heat shock protein 80	Y	Y	-	Y	Y
Heat shock protein 70 kDa	Y	-	-	Y	Y
ORF YJR045c	Y	-	-	-	-
Heat shock protein 88 kDa	Y	-	-	-	-
14-3-3 homolog protein	Y	-	-	-	Y
Eukaryotic initiation factor 4a	Y	-	Y	-	Y
Actin	Y	-	Y	Y	Y
Phosphoglycerate kinase	Y	-	Y	Y	-
Glyceraldehyde-3-phosphate dehydrogenase	Y	Y	Y	Y	Y
Small G-protein Gsp1p	Y	-	-	Y	Y
Probable cytosceletal binding protein	Y	-	-	-	-
Enolase	Y	Y	Y	-	-
Mitochondrial processing peptidase β-subunit	Y	-	-	-	-
6-phosphogluconate dehydrogenase	Y	-	Y	-	-
ORF YOR185c	Y	-	-	-	-
GTP binding protein SAR1	-	Y	-	-	-
C-terminal fragment of cytochrome c oxidase polypeptide II	-	Y	-	-	-
Calmodulin	-	Y	-	-	-
Probable GTP binding protein DRAB11		Y			
NADH-ubiquinone oxidoreductase 24 kDa subunit, mitochondrial	-	Y	-	-	-
Guanidine nucleotide-binding protein β-subunit-like protein	-	Y	-	-	-
Ketol-acid reductoisomerase	-	Y	-	-	-
Citrate synthase	-	Y	-	-	-
ATP synthase α-chain, mitochondrial	-	Y	-	-	-
Probable translation elongation factor EF-Tu	-	Y	-	-	-
Similar to yeast YNL313c	-	Y	-	-	-
Tubulin β-chain	-	Y	-	-	-
COX1/OXI3 intron 1 protein	-	Y	-	-	-
Elongation factor 2	-	Y	-	-	-
Probable oxoglutarate dehydrogenase	-	Y	-	-	-
Yeast chaperonin Hsp 78	-	Y	-	-	-
Vacuolar ATP synthase subunit B	-	Y	-	-	-
78 kDa glucose regulated protein	-	Y	-	-	-
Elongation factor 1α	-	Y	-	-	-
ADP/ATP carrier protein	-	Y	-	-	-
Ubiquitin	-	-	Y	-	-
Pyruvate kinase	-	-	Y	-	-
Activation mediator subcomplex of RNA polymerase II	-	-	-	Y	-
60S ribosomal protein	-	-	-	-	Y

WC – whole cell extract, EM – enriched mitochondrial extract, TS – Tris-soluble extract, TI – Tris-insoluble extract, M – membrane extract, Y – present in indicated extracts.

Summary

Five procedures have been developed and optimised for the display of different sub-cellular protein extracts of *T. harzianum* on 2D gels. These include the whole cell extract (Publications I and II, Sections 3.1.1 and 3.1.2), an enriched mitochondrial extract (Publication III, Section 3.1.3), a Tris-soluble extract, Tris-insoluble extract, and a membrane extract (Section 3.1.5). A combination of mass spectrometry techniques was used to identify some of these proteins with the primary method employed being MALDI-TOF-MS. Further protein identifications were obtained by LC-MS/MS in the case of the whole cell extract and enriched mitochondrial extract. In total, 573 proteins were analysed for mass spectrometry and 116 individual protein spots were identified giving an overall identification frequency of approximately 20 %. Of the 116 protein spots identified, 42 gene products were found (shown in Table 6). The low overall protein identification frequency proves that cross-species protein identification by mass spectrometry has limitations and highlights the importance of having access to a completed genome sequence to increase the identification frequency. Only 43 % of the identified proteins were matched to proteins from the *Trichoderma* or *Hypocrea* (sexual form of *Trichoderma*) protein databases as these databases contain limited information. Despite its limitations, cross-species identification has made a major contribution to this work by more than doubling the number of protein identifications that were made when compared to using the *Trichoderma* databases alone. Electrospray ionisation mass spectrometry (ESI-MS/MS) and other sensitive mass spectrometry techniques may have yielded slightly better identification frequencies than those achieved by MALDI-TOF MS or LC-MS/MS. However, these improved technologies would be insignificant without a species-specific protein database for *Trichoderma spp*.

Twelve proteins (or 29 %) were identified across more than one protein extract and represent highly abundant and highly soluble proteins. Epitope labelling work with the yeast *S. cerevisiae* indicated that, of the 1200 proteins localised to a subcellular organelle, over 25 % showed mixed compartmentalisation with a particular organelle and the cytoplasm (Karlberg et al. 2000). Thirty of the identified gene products (or 71 %) were identified from a single protein extract. This highlights the importance of displaying sub-cellular proteomes in maximising the information that can be obtained using a 2D approach.

3.2 Discovery of novel determinants in the biological control of phytopathogens by *Trichoderma atroviride*

3.2.1 Proteomic response of the biological control fungus *Trichoderma atroviride* to growth on the cell walls of *Rhizoctonia solani* (Publication V)

This publication represents the first 2D proteomics approach to attempt to identify determinants playing a role in the biocontrol response of *Trichoderma atroviride* when grown on cell wall material of *Rhizoctonia solani*. 2DE combined with differential gel electrophoresis and image analysis has helped to confirm known biocontrol-related proteins and identify three novel proteases from *Trichoderma* that may play a role in biocontrol.

I would like to acknowledge the skilful assistance of Matthew McKay in assisting with the LC-MS/MS and interpretation of MS/MS data and to Sybille Hunt for her assistance with PSLImage, the image analysis software package used for differential gel electrophoresis. I also wish to acknowledge my supervisors Helena Nevalainen and Ben Herbert for their helpful ideas in the experimental approach for this work and for discussions in the preparation of the final manuscript.

Publication V

Due to copyright laws, the following article has been removed from this thesis. Please refer to the following link for abstract details.

Grinyer J, Hunt S, McKay M, Herbert BR, Nevalainen H (2005) Proteomic response of the biological control fungus *Trichoderma atroviride* to growth on the cell walls of *Rhizoctonia solani*. Curr Genet 47:381-388

http://dx.doi.org/10.1007/s00294-005-0575-3

3.2.2 Additional, unpublished information - Isolation of the sequences of biologically relevant protease-encoding genes in the biological control response of *Trichoderma atroviride*

This section describes the isolation of two gene sequences encoding two proteases that were overexpressed when *Trichoderma atroviride* was grown on the cell wall material of the phytopathogen *Rhizoctonia solani.*

Work described in this chapter is my own. I would like to thank Junior Te'o and Moreland Gibbs for introducing the techniques for genomic walking PCR. I would like to thank my supervisors Helena Nevalainen and Ben Herbert, particularly Helena for giving me the time to complete the gene isolations. Many thanks go to you both for your contributions made in reviewing this section.

Introduction

Three proteases, vacuolar protease A (VPA), vacuolar serine protease (VSP) and trypsin-like protease (TLP), were up-regulated when *T. atroviride* was grown on the cell wall material of the phytopathogen *R. solani* when compared to growth on glucose (Grinyer et al. 2005, Publication V, Section 3.2.1). Consequently, gene sequences encoding two of the three proteases (VPA and VSP) were isolated. The gene sequences, *vpa1* and *vsp1*, were isolated using a combination of various molecular methods. First, consensus primers were designed to obtain a large portion of each gene sequence based on the recently available genome sequence of *T. reesei*, and then genomic walking PCR was used to isolate the N- and C-terminal sequences. The gene sequences for both *vpa1* and *vsp1* will be presented in this chapter, including bioinformatic analysis providing further gene and protein information.

A homolog for the TLP protein termed PRA1 was identified from *T. harzianum* CECT 2413 in a proteomic display of the secretome. The *T. harzianum* protease PRA1 (acidic protease) was found to have a molecular mass of 28 kDa and pI of between 4.7-4.9 (Suarez et al. 2004). The pI range differs from that observed from the trypsin-like protease from *T. atroviride* (which has a pI of approximately 6). The N-terminal sequences of both trypsin-like proteases are well conserved with only one amino acid substitution in the first 10 amino acids, occurring at position nine where the alanine from *T. atroviride* is replaced by a leucine in *T. harzianum*. The gene encoding PRA1 has previously been isolated from *T. harzianum* (Suarez et al. 2004) and accordingly, it was not pursued further from *T. atroviride*. A trypsin-like protease has also been isolated from *T. viride* with a molecular mass of 25 kDa and a pI of 7.3. However, the corresponding gene has not yet been isolated. This trypsin-like protease also displays

N-terminal amino acid homology with the *T. atroviride* protease. The only amino acid substitution occurs at amino acid 6 where a threonine from *T. atroviride* is replaced by an arginine in *T. viride* (Uchikoba et al. 2001). PRA1 is secreted when *Trichoderma* was grown on fungal cell wall material (Suarez et al. 2004), implying that the TLP identified from *T. atroviride* in this work also should be secreted under these conditions.

Methodology

Fungal strain and cultivation conditions

Trichoderma atroviride P1 (ATCC 74058) was maintained on Potato Dextrose agar (PDA) plates. The fungus was grown on cellophane discs placed on PDA plates incubated at 28 ºC for 24 h.

Isolation of Genomic DNA

Mycelium from two PDA plates carrying cellophane discs were removed and freeze-dried overnight. Once dried, 2.5 mL of Lysis buffer (50 mM Tris-HCl, pH 7.5, 50 mM EDTA, 3 % sodium dodecyl sulphate (SDS), 1 % mercaptoethanol) was added to the mycelium and incubated at 65 ºC for 1 h. The material was moved to an Eppendorf Light Gel Separation tube and 3 mL of 50 % phenol, 50 % chloroform mixture added. The tube was inverted gently before spinning at 3000 g for 5 min at RT. The phenol-chloroform step was repeated in a fresh Eppendorf Light Gel separation tube. Genomic DNA was precipitated by adding 100 µL of 3 M sodium acetate, pH 8 and an equal volume of isopropanol and incubated for 1-2 min. Genomic DNA was removed using a glass Pasteur pipette and collected. The DNA was washed with 500 µL of 70 % ethanol, dried and resuspended in 200 µL TE buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA, pH 8).

Designing consensus primers for the isolation of *vpa1* and *vsp1*

The design of consensus primers was based on the cDNA sequence of ortholog genes from the recently sequenced *T. reesei* genome and closely related gene sequences from *Gibberella zeae* PH-1. Primers were designed at the N- and C-terminus of *vpa1* and

vsp1. When differences existed between gene sequences of these organisms, degenerate primers were designed. See Table 7 for primer sequences.

Table 7: Primer design for obtaining *vpa1* and *vsp1* sequences, including the primers used for genomic walking PCR (contain a GW prefix).

Gene	**N-terminal primer 5'-3'**	**C-terminal primer 5'-3'**
vpa1	ATG AAG AGC GCK TTA CTT RCC GC	TT GGC CTT GGC MAG ACC RAC RGC
vsp1	ATG CGG TCC GYT GTC GCY CTC TC	TTA CTC GGA RAG CTC AGA GAT GAA CT
GW PCR primers		
GWVPA-F2		GGC GAT GCT TTC CTG CGC CGA TAC TAC
GWVPA-R2	TCT TGT GGA CGC CGG CTT GGG CGG A	
GWVSP-F2		CGC CAA GAA GAT CCA CGA CCT CGT CGA
GWVSP-R2	GGA GCA GAG CTC TCG TGG ATG GTG CC	

R represents an A or G; M represents an A or C; Y represents a C or T; K represents a G or T

PCR for partial isolation of *vpa1* and *vsp1*

PCR reactions were performed in 50 μL reaction mixtures containing 10 pmol forward and reverse primers, 0.25 mM deoxynucleoside triphosphates (dNTPs), 1 U of *Taq* DNA polymerase (unless otherwise specified), 0.25 mM magnesium chloride, 10 mM Tris-hydrochloride (pH 8.8), 50 mM potassium chloride, 0.001 % gelatin, and 1 to 10 ng template DNA.

PCR reactions for *vpa1* and *vsp1* gene isolation

Stepdown PCR was used to obtain PCR fragments of both *vpa1* and *vsp1* genes: step 1, 94 ºC for 10 min; step 2, 20 cycles of 94 ºC for 30 sec, 65-45 ºC for 30 sec and 72 ºC for 2 min; step 3, 20 cycles of 94 ºC for 30 sec, 55 ºC for 30 sec, 72 ºC for 2 min; step 4, 72 ºC for 2 min; step 5, 4 ºC soak. AmpliTaq Gold polymerase (Applied Biosystems) was used for PCR reactions for the amplification of *vpa1* and TripleMaster Polymerase (Eppendorf) was used for the amplification of *vsp1*.

Genomic walking PCR (GWPCR)

Genomic walking primers were designed approximately 60-100 bases from the N- and C-terminus of each protease gene (Table 7). Genomic DNA was digested with an array of 17 restriction enzymes (11 providing sticky-ends and 6 blunt-ends). Linker assembly, linker library construction and GWPCR were performed according to Morris *et al*. (1995; 1998). Reaction mixtures were cycled with the following profile: step 1, 94 ºC for 15 min; step 2, 35 cycles of 94 ºC for 30 sec, 65 ºC for 15 sec, 72 ºC for 3 min; step 3, 72 ºC for 1 min; step 4, 4 ºC soak. AmpliTaq Gold polymerase (Applied Biosystems, Australia) was used for GWPCR reactions.

Transformation of *vpa1* and *vsp1* genes

PCR products were extracted from a 1 % (w/v) agarose gel using Qiagen (Australia) Gel Extraction Kit following manufacturers' instructions and ligated into a pCR 2.1 Vector from the TA Cloning Kit (Invitrogen, Australia). The vector was then transformed into competent DH5α *E. coli* cells and blue / white selection used to observe tranformants. Plasmid DNA was purified from these transformant cells using Qiagen QIAprep Spin Miniprep Kit following the manufacturers' instructions.

Sequencing of *vpa1* and *vsp1* genes

DNA was sequenced by the dideoxy chain termination method (Sanger *et al*. 1977) using the ABI PRISM BigDye terminator cycle sequencing ready reaction kit (Applied Biosystems) and was analysed on an ABI 377 automated sequencer.

Database searching and analysis

Sequence alignments were performed using the BLASTP service available from the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/BLAST/). Introns were predicted using Augustus Gene Prediction from the University of Göttingen (http://augustus.gobics.de/submission). Signal peptides were predicted using SignalP version 3.0 web server (http://www.cbs.dtu.dk/services/SignalP/). O-linked glycosylation sites were predicted using NetOGlyc version 3.1 server (http://www.cbs.dtu.dk/services/NetOGlyc/) and N-linked glycosylation sites were predicted using NetNGlyc version 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/).

Results and Discussion

Isolation of *vpa1* and *vsp1* genes

Consensus primers were initially designed around conserved regions of *vpa1* and *vsp1* gene sequences from several closely related fungal species. Over 50 % of each gene sequence was obtained using these degenerative primers. The original approach was to use genomic walking PCR (GWPCR) to obtain the remaining portion of each gene sequence. However, a different approach was taken once the *T. reesei* genome sequence was publicly made available on 25th July 2003. The partial sequences obtained from both *vpa1* and *vsp1* were subjected to a BLASTN search against the *T. reesei* cDNA library, resulting in identification of the full length cDNA sequence of both corresponding genes. From here, two degenerate primers were designed based on N- and C-terminal regions of *T. reesei* and *G. zeae* PH-1 (see Table 7 in Methodology). The almost complete gene sequence was obtained by PCR with these primers. However, GWPCR was required to obtain the true N- and C-terminal sequences. One round of GWPCR resulted in the amplification of specific *T. atroviride* sequences for both *vpa1* and *vsp1* (shown in Figures 11 and 13). The corresponding genes were named *vpa1* and *vsp1* and were sequenced in both directions to obtain accurate gene sequences. The gene sequences for both *vpa1* and *vsp1* and the corresponding predicted protein sequences have been submitted to Genbank on NCBI. The Genbank accession numbers are DQ661006 and DQ661007 for *vpa1* and *vsp1* respectively and ABG57251 and ABG57252 for the predicted protein sequences VPA and VSP respectively. The Genbank entries for *vpa1* and *vsp1* are shown in Appendix 5.2 and Appendix 5.3 respectively.

Features of *vpa1*

The *vpa1* gene is 1321 bp long (Figure 11). The first of the predicted introns starts at position 99 and is 66 bp in length; the second intron starts at position 319 and is 68 bp in length. Excision of the introns and translation of the predicted amino acid sequence from the gene sequence results in an open reading frame of 395 amino acid residues (VPA; Figure 12). This protein has a predicted pI of 4.81 and a molecular mass of 42.2 kDa, matching the position of the corresponding protein from the 2D gel in the previous publication (Grinyer et al. 2005, Publication V, Section 3.2.1).

The translated protein sequence of *vpa1* was submitted to a BLASTP search and matched most closely to a hypothetical protein from *G. zeae* PH-1 (FG10782.1) with 81 % sequence identity. This hypothetical protein was shown to contain a conserved aspartic proteinase domain. Other fungal vacuolar proteases had lower levels of sequence homology to VPA. A signal peptide 17 amino acids in length was predicted at the N-terminus of the protein, indicating that this protease was secreted. VPA contains one predicted O-linked glycosylation site at threonine 55 and one predicted N-linked glycosylation site at asparagine - any amino acid - serine (NXS motif) from amino acids 138-140. The presence of a single sulphate on the O-linked sugar, phosphorylation on the N-glycan structure or a combination of both may explain the train of spots identified as VPA from 2D gel analysis undertaken in Grinyer et al. (2005), Publication V, Section 3.2.1. (Harrison et al. 2002).

Figure 11

Sequence of the *vpa1* gene encoding the vacuolar protease A of *T. atroviride* (Genbank accession number DQ661006). The gene sequence is highlighted in blue and predicted introns in pink. Up- and down-stream gene sequences are shown in black.

```
   1  CTGCTCGGGC GGTACTCAAA GGGCCCCCTT GCAGGGCCGC CGTGCGTCAT
  51  CAGCCCGTCA GCACAGGCAC CAATAGGCGC GACAGCACGC AAACGGCGCG
 101  AGCACGAGGC CTAACCCCCC GTTCGCCCTG TTTGATTAGC CCAGGTCCAG
 151  TGACCTCTTG GTCTGTAGCA GCGGTAGCAT CACCGCCAGC CACCGATGCT
 201  ACACCAGAGG CCGACCAGTA CTAAGCTGCC TGCCGCCTTC TTATCTAAGT
 251  TCCCGCTACT GCTGCTGGCG CCACTACAAA GGCAGCTTCG TCATCCCTTC
 301  TCCTCGGGCT GCTCTTCCAA TCCATCAACT CACCATCCTC CATCCATCAC
 351  CTCTCACCCC AAAAACACCC ATCATCAAGT TCCCAAGTGT CGCCCAGTAT
 401  GAAGAGCGCA TTGATTGCCG CTGCGGCGCT TGTCGGCTCC GCCCAAGCCG
 451  GCGTCCACAA GATGAAGCTG CAAAAGGTTT CCCTGGAGCA GCAGCTGGTA
 501  TGAACCCGCC CCAATCTGCC TCTGCACCGT CGCACGAGCA CTAACGCTTC
 551  TCTTTTTCAA CAGGAGGGCT CATCCATCGA GGCCCAAGTC CAGCAGCTCG
 601  GCCAGAAGTA CATGGGCGTG CGCCCTACTA GCCGTGTCGA TGTCATGTTC
 651  AACGACAACG TGCCCAAGGT CAAGGGCGGC CACCCGGTCC CCGTCACCAA
 701  CTTCATGAAT GCCCAATGTG AGTCAACGTG CAATGGCACT ACTCCAGCAA
 751  GCGAGTGTAG CGCTTTACTG ACTTGAATGT GACAGACTTC TCCGAGATCA
 801  CCATCGGCTC TCCTCCCCAG ACTTTCAAGG TCGTCCTTGA CACGGGAAGC
 851  TCCAACCTCT GGGTTCCCTC GCAGTCCTGC AACAGCATTG CCTGCTTCCT
 901  GCACTCCACG TACGATTCGT CTTCCTCGTC ATCGTACAAG AAGAATGGAT
 951  CCGATTTCGA AATCCACTAC GGATCGGGTA GCTTGACTGG ATTCATCTCC
1001  AACGATGTTG TCACTATCGG TGACCTCCAG ATCAAGGGCC AGGACTTTGC
1051  CGAGGCTACC AGCGAGCCCG GCCTTGCCTT TGCCTTTGGC CGCTTTGACG
1101  GCATTCTTGG CCTTGGCTAC GATACCATTT CCGTCAATGG TATTGTTCCT
1151  CCCTTTTACC AGATGGTGAA CCAGAAGCTT CTGGACGAGC CTGTTTTCGC
1201  CTTCTACCTT GGAAGCGGCG ACGAAGGTTC TGTGGCTACC TTTGGTGGCG
```

1251 TTGATGAGTC CCACTACTCG GGCAAGATTG AGTACATTCC TCTCCGCCGC

1301 AAGGCTTATT GGGAGGTTGA CCTTGACTCC ATTGCCTTTG GCGATGAGGT

1351 CGCCGAGCTT GAGAACACTG GTGCCATCCT CGACACTGGT ACCTCGCTCA

1401 ACGTTCTCCC CTCTGGCATC GCCGAGCTGC TCAATGCTGA GATTGGCGCC

1451 AAGAAGGGCT ACGGTGGCCA GTACACCATC GACTGTGCCA AGCGCGACTC

1501 CCTCCCCGAC ATTACCTTCA GCCTTGCCGG CTCCAAGTAC AGCCTCCCTG

1551 CTTCTGACTA CATCCTTGAG GTGTCTGGCA GCTGCATCTC TACCTTCCAG

1601 GGCATGGACT TCCCCGAGCC CGTGGGCCCC CTGGTCATCC TCGGCGATGC

1651 TTTCCTGCGC CGATACTACT CAGTCTATGA CCTCGGCAAG GGAGCTGTCG

1701 GTCTTGCCAA GGCCAAATAA AGGAATAATA TAGTACACAG TGGAATGAAT

1751 GTGTTGCAAG GATGGGCACA GCTTATATGT TGCTAGAAGG CAGTTTTGGT

1801 TAATTGAGCG CCTTTGCCTA TATAGTACTG CGAGGTGTAC GGATCAAGAT

1851 TGCAGCGCCG TTTATGTAAT ATAATGGAAC ACGGGTACTT GTTTTGCTGC

1901 TCTAATTTTA GTTGATATTC ACAAATTGAT CCTGTTTTCA ATCTTTTGCA

1951 TTTATTTGGT GAATGGCTTG ACGCGTAAAC TTGAGGTGGC TACAGCGTTT

2001 GTGGCTGCTT TGTGTGTATT TGGTTTCCAG TAGGTATTCC TTTTTGGTGG

2051 TGTCTGGTAC AACAATCTAC CCATCGCCTA CCTGGTAGGT ACTCAAAAGA

2101 ATATGGTTAA TACAGGTTTG TATCACGGTA AGATCTGGTA ATCAGTGAGA

2151 TTTCGTGCAT TAACATTGGC ACCCTGCCTT GGTACAGAGC TGAAAGTCAG

2201 CGTGGGATC

Figure 12

The translated amino acid sequence of *T. atroviride* vacuolar protease A (VPA; accession number ABG57251). A predicted signal peptide encompassing amino acids 1-17 is highlighted in green, a predicted N-linked glycosylation site is shown in italics and is underlined and a predicted O-linked glycosylation site at threonine 55 is highlighted in red. The amino acid sequence was found to contain a conserved aspartyl domain from amino acids 78-394 that is highlighted in blue.

MKSALIAAAALVGSAQAGVHKMKLQKVSLEQQLEGSSIEAQVQQLGQKYMGVRPTSRVDV
MFNDNVPKVKGGHPVPVTNFMNAQYFSEITIGSPPQTFKVVLDTGSSNLWVPSQSCNSIA
CFLHSTYDSSSSSSYKK*NGS*DFEIHYGSGSLTGFISNDVVTIGDLQIKGQDFAEATSEPG
LAFAFGRFDGILGLGYDTISVNGIVPPFYQMVNQKLLDEPVFAFYLGSGDEGSVATFGGV
DESHYSGKIEYIPLRRKAYWEVDLDSIAFGDEVAELENTGAILDTGTSLNVLPSGIAELL
NAEIGAKKGYGGQYTIDCAKRDSLPDITFSLAGSKYSLPASDYILEVSGSCISTFQGMDF
PEPVGPLVILGDAFLRRYYSVYDLGKGAVGLAKAK

Features of *vsp1*

The *vsp1* gene is 1678 bp long. The only predicted intron starts at position 366 and is 65 bp in length (Figure 13). Excision of this intron and translation of the gene sequence results in an open reading frame of 537 amino acid residues (VSP; Figure 14). This protein has a predicted pI of 5.28 and a molecular mass of 57.4 kDa. While this molecular mass does not match the location on the 2D gel where the protein was originally observed, it is likely that downstream processing has occurred resulting in the smaller fragment visible on the 2D gel. This assumption is supported by the lack of mass spectrometry data obtained for the first 250 amino acids of the protein (results taken from Grinyer et al. 2005, Publication V, Section 3.2.1), indicating that a correct protein fragment is observed.

The entire predicted protein sequence of VSP was submitted to BLASTP and it matched most closely to a hypothetical protein from *G. zeae* PH-1 (FG00192.1) showing 72 % sequence identity. The amino acid sequence was found to contain both a peptidase S8-conserved domain (185-448 amino acids) and a subtilisin-conserved domain (41-82 amino acids). However, we can predict that the subtilisin-conserved domain is absent from the protein isoform identified in this work due to the missing mass spectrometry data from the N-terminal 250 amino acids. Other fungal vacuolar serine proteases had lower levels of sequence homology to VSP. A signal peptide 15 amino acids in length was predicted at the N-terminus of the protein, indicating that this protein is secreted. The protein sequence of VSP contains two N-linked glycosylation sites. However no O-linked glycosylation site could be predicted. The presence of a phosphorylation site with an N-linked glycan could explain the pI

isoforms observed from the original 2D analysis (Grinyer et al. 2005, Publication V, Section 3.2.1; Harrison et al. 2002).

Figure 13

Sequence of the *vsp1* gene encoding a vacuolar serine protease (Genbank accession number DQ661007). The gene sequence is highlighted in blue and predicted intron in pink. Up- and down-stream gene sequences are shown in black.

```
   1  GCTGAGGCAG CACTGGAGCC TCGTCCCAGC CGCCCTTATC TACTGCGCCA
  51  TCGCTCGTCT CTTCGTCCTC TTTCTTTCTC TTCTTCCATC ATCCTTCTCC
 101  CCATCCATCC ATCGCCGTCT CTTCTCCCAT CTATCTGCCC ATCTCGCAGG
 151  AGGCCGTTGG AGTTTTCTTG ATTTCCGTCA CCTTTGTTTT CGTCTCTTCT
 201  GGCCGCGTGT CGCAGGTTGA CCAACTCCCA TATTCCCTCA TTCATCACAT
 251  CTATCAGGAC CATCATTATG CGGTCCGTCA TCGCTCTCTC TGTGGCGGCC
 301  GTCGCCCAGG CCAGCTCGTT CCAGGTTGGC ACCATCCACG AGAGCTCTGC
 351  TCCCGTCCTC AGCAATATTG AGGCCAACGC CATTCCGGAC AACTACATCA
 401  TCAAATTCAA GGATCACGTT GATGAGGCTG GTGCCGACAA GCACCAGAAC
 451  TGGATCCAGA GCATCCACGA TGAAGGCGAG CAGCAGCGCC TTGAGCTCCG
 501  CAAGCGAAGC AGCATCTTCG GCGCGGACGA GGCCTTTGAC GGCCTGAAGC
 551  ACACCTTCAA GATTGGTGAT TTCAAGGGCT ACGCTGGTCA CTTTACCGAG
 601  GACATCATCG AGCAGGTCCG GAACCACCCG GATGTAAGTT CACGCTACCA
 651  TTTCCATCGC CACTGGCAGG TCAATTGAGG CTAATCCATC GTTTCCAGGT
 701  CGAGTATATT GAGCGCGACA CCATTGTCCA CACTATGCTT CCCCTCGACT
 751  CACAAGACAG CGTTGTTGTC GAGGACTCGT GCAACCCCGA GACCGAGAAG
 801  CAGGCTCCCT GGGGCTTAGC CCGTATCTCT CACCGAGACA CCCTGGGCTT
 851  CGGCACATTT AACAAGTACC TCTACACCGC TGACGGTGGC GAGGGTGTTG
 901  ATGCCTATGT CATTGACACT GGTACCAACA TTGAGCACGT CGACTTCGAG
 951  GGTCGTGCCA AGTGGGGCAA GACCATCCCT GCCGGCGATG AGGATGAGGA
1001  TGGCAACGGC CACGGCACTC ACTGCTCTGG TACCGTTGCT GGTAAGAAGT
```

1051 ACGGAGTTGC CAAGAAGGCC CATGTCTACG CCGTCAAGGT CCTCCGATCC

1101 AACGGCTCTG GCACCATGAG CGATGTCGTC AAGGGCGTCG AGTACGCCGC

1151 TCTCTCCCAC CAGGAGCAGG TTAAGCAGGC CAAGGCCGGC AAGCGAAAGG

1201 GCTTCAAGGG CTCAGTCGCC AACATGTCTC TTGGTGGTGG CAAGACCTCT

1251 GCTCTCGACG CCGCTGTCAA CGCTGCCGTC AAGGCCGGTG TTCACTTCGC

1301 CGTCGCTGCT GGCAACGACA ACGCCGATGC CTGCAACTAC TCCCCCGCCG

1351 CCGCCACCGA GCCCGTCACC GTCGGTGCCT CTGCTCTGGA CGACAGCCGT

1401 GCTTACTTCT CCAACTACGG CAAGTGCACT GACATCTTCG CCCCTGGCTT

1451 GAGCATCCAG TCCACCTGGA TCGGCTCCAA GTATGCCGTC AACACCATCT

1501 CTGGTACCTC CATGGCGTCT CCCCACATCG CCGGTCTCCT GGCCTACTAC

1551 CTGTCTCTCC AGCCTGCTGG TGACTCTGAG TTTGCTGTTG CCCCCATCAC

1601 CCCCAACAAG CTCAAGAAGG ACCTCATCAC CATTGCTACC CAGGGTACTC

1651 TGTCTGATAT CCCCGCCGAT ACCCCCAACC TGCTTGCCTG GAACGGTGCT

1701 GGCTGCAGCA ACTACTCCCA GATCGTCAAG GCCGGTGGCT ACAAGGCCAA

1751 GGCCCAGAAG CAGACCAAGC TCCCCGCTAC CATTGAGGAG CTCGAGGACG

1801 CCATCGAGGG TGACTTTGAG GTTGTCTCTG GCAAGATTGT CAAGGGTGCC

1851 AAGTCCTTTG GCTCCAAGGC TGAGAAGTTC GCCAAGAAGA TCCACGACCT

1901 CGTCGAGGAG GAGATTGACG AGTTCATGTC CGAGCTCTCC GAGTAAGAAT

1951 CTTTATTAAG CATATATTTC CTTTTTATTT GTAATGGCCA TTATGCGCCG

2001 ATGAGTACGT CGAGTACGAC GGGTTATGGA ACGGGACCAG GAGAGACTAT

2051 TTGTTTTGGG TAATTGGCAT G

Figure 14

The translated amino acid sequence of the vacuolar serine protease (VSP; accession number ABG57252). A predicted signal peptide encompassing amino acids 1-15 is highlighted in green, two predicted N-linked glycosylation sites are highlighted in italics and underlined. The amino acid sequence was found to contain two conserved domains, a peptidase S8 subtilase domain between amino acids 185-448 and a subtilisin-conserved domain from amino acids 41-82. Both conserved domains have been highlighted in blue.

MRSVIALSVAAVAQASSFQVGTIHESSAPVLSNIEANAIPDNYIIKFKDHVDEAGADKHQ
NWIQSIHDEGEQQRLELRKRSSIFGADEAFDGLKHTFKIGDFKGYAGHFTEDIIEQVRNH
PDVEYIERDTIVHTMLPLDSQDSVVVEDSCNPETEKQAPWGLARISHRDTLGFGTFNKYL
YTADGGEGVDAYVIDTGTNIEHVDFEGRAKWGKTIPAGDEDEDGNGHGTHCSGTVADKKY
GVAKKAHVYAVKVLRS*NGS*GTMSDVVKGVEYAALSHQEQVKQAKAGKRKGFKGSVANMSL
GGGKTSALDAAVNAAVKAGVHFAVAAGNDNADACNYSPAAATEPVTVGASALDDSRAYFS
NYGKCTDIFAPGLSIQSTWIGSKYAVNTISGTSMASPHIAGLLAYYLSLQPAGDSEFAVA
PITPNKKKDLITIATQGTLSDIPADTPNLLAWNGAGCS*NYS*QIVKAGGYKAKAQKQTKLP
ATIEELEDAIEGDFEVVSGKIVKGAKSFGSKAEKFAKKIHDLVEEEIDEFMSELSE

Summary

The proteomic study of proteins from *T. atroviride* P1 showed that three proteases were differentially expressed when the fungus was grown on the cell wall material of the phytopathogen *R. solani* when compared to growth on glucose (Grinyer et al. 2005, Publication V, Section 3.2.1). The gene sequences encoding vacuolar protease A (*vpa1*) and vacuolar serine protease (*vsp1*) were isolated and information showing predicted introns, translated amino acid sequences, signal peptide presence and modification sites has been presented. The presence of predicted signal peptide sequences on both proteases indicates that these proteases may be secreted from the fungus to play an active role in the biological control response of *T. atroviride*. Recently, an aspartyl protease was identified as the most dominant protein expressed when *T. harzianum* CECT 2413 was grown on the cell walls of *R. solani*, *P. ultimum* and *B. cinerea* (Suarez et al. 2005). However, to say conclusively that VPA and VSP are secreted, their presence must be confirmed in secreted protein fractions, either by antibody detection or further mass spectrometry analysis. The potential biological role of both proteases warrants further exploration.

3.3 Summary and concluding remarks

This thesis describes the development and application of several methods to study the biological control fungus *Trichoderma* spp. by proteomics:

1) Methods for extracting proteins from *Trichoderma harzianum* were developed for their study by two-dimensional electrophoresis (2DE). A) An acid-based protein extraction method was developed to reduce the solubilisation of carbohydrates, such as β-1,3-glucans contaminating the whole cell proteome from the fungal cell wall. Acid extraction minimised protein precipitation and streaking on 2D gels at an acidic pH (Herbert et al. 2006, Publication I, Sections 3.1.1; Grinyer et al. 2004a, Publication II, Section 3.1.2). B) Methods were also developed for extracting proteins from a partially purified subcellular fraction of mitochondria (Grinyer et al. 2004b, Publication III, Section 3.1.3), as well as from the cellular membrane, Tris-soluble and Tris-insoluble protein fractions (Section 3.1.5). C) Sample prefractionation by the multi-compartment electrolyser and protein identification by mass spectrometry on a membrane preparation of *S. cerevisiae* enabled the visualisation and identification of low abundance proteins and alkaline membrane-bound or associated proteins (Pedersen et al. 2003, Publication IV, Section 3.1.4).
2) Protein reference maps for each of the proteomic fractions of *Trichoderma* spp. described above were produced using a combination of MALDI-TOF-MS and, where possible LC-MS/MS (Grinyer et al. 2004a, Publication II, Section 3.1.2; Grinyer et al. 2004b, Publication III, Section 3.1.3; Section 3.1.5).
3) Protein expression levels were monitored by differential gel electrophoresis in order to identify novel proteins involved in the biocontrol response of *T. atroviride* (Grinyer et al. 2005, Publication V, Section 3.2.1). Several known

4) The genes encoding vacuolar protease A and vacuolar serine protease were isolated by PCR and genomic walking PCR. The gene sequence and protein sequence derived from each gene (*vpa1* and *vsp1*) have been characterised (Section 3.2.2).

3.3.1 Mapping the *Trichoderma harzianum* proteome

This is the first time that an extensive array of sub-cellular proteomes has been displayed for the filamentous fungus, *Trichoderma harzianum* A3091. Methods for the display of proteomes of the whole protein content, enriched mitochondrial protein, Tris-soluble, Tris-insoluble, and membrane proteins have been developed and protein identification from the resulting 2D gels have been made by MALDI-TOF-MS and LC-MS/MS. Table 8 below summarises the information obtained from the proteome mapping of *T. harzianum*.

Table 8: All proteins identified from the whole and sub-proteome fractions of *T. harzianum*. An identification frequency of 20.2 % was achieved with a combination of mass spectrometry techniques including MALDI-TOF-MS and LC-MS/MS.

Total protein spots processed for MS	573
Total proteins identified	116
Unique proteins / gene products identified	42
% protein identified from *Trichoderma* genus	43 %
% protein identified by CSI from other fungal and yeast species	57 %
% proteins identified in more than one protein extract	29 %
% proteins identified from a single protein extract	71 %

Subcellular fractionation for proteomic studies has become an essential tool to visualise a greater proportion of the proteome than whole-cell protein extracts. A single protein extract will contain both highly abundant cytoplasmic proteins and less abundant organellar proteins. However, when identifications are made from a 2D gel of a whole cell protein extract, the highly abundant proteins will dominate and hide the presence of the low abundance proteins, resulting in a distorted view of the proteome. Subcellular fractionation is conducted to either remove those highly abundant cytoplasmic proteins or to display them on separate 2D gels. Fractionation allows a more realistic view of the subcellular proteome, whether compartmentalised into organelles or based on protein solubility. The importance of subcellular fractionation has been shown in this work, where a further 24 proteins were identified from all combined subcellular fractions in addition to the 18 protein identifications that were made from the whole cell protein fraction. Subcellular fractions can be further divided into narrower pH ranges, using equipment similar to the multi-compartment electrolyser (MCE) or narrow range pH gradients. Sample prefractionation using the MCE has allowed the display of difficult protein fractions such as the alkaline membrane proteome of the yeast *S. cerevisiae* (Pedersen et al. 2003, Publication IV, Section 3.1.4).

The identification of a protein in more than one subcellular fraction is a common occurrence and has recently been described in a *S. cerevisiae* study. Of the 1200 yeast proteins known to localise within an organelle, more than 25 % co-localised with other organelles or the cytoplasm (Karlberg et al. 2000). This co-localisation explains why some proteins identified from the cytoplasmic, Tris-soluble fraction were also identified in organelle or protein fractions from cellular membranes in the current

work. Many proteins have more than one function that involves co-localisation between the cytoplasm and other cellular compartments. Co-localisation was shown in the work discussed in Section 3.1.5, where cytoplasmic proteins were found to contaminate membrane preparations. Co-localisation of cytoplasmic proteins was implicated from a search of the recent literature. One such example was the 14-3-3-like protein which was found to associate with plasma and organelle membranes in plants (Pertl et al. 2005), bacteria (Assossou et al. 2003) and in a green alga (Voigt et al. 2001), indicating that cytoplasmic proteins can interact with cellular membranes and are not merely cytoplasmic contaminants as previously thought.

3.3.2 Importance of protein databases for proteomics

In the early stages of this work, very little information was available on proteins from *Trichoderma* spp. and only 269 proteins had been submitted into the Swiss-Prot / TrEMBL protein database as of September 2003. Of these 269 proteins, approximately 60 were from *T. harzianum*. The lack of protein information available in databases during this time severely limited the number of protein identifications that could be obtained using a proteomic approach. The number of proteins in the *Trichoderma* spp. database has slowly risen to 675 as of May 2006. However, only 103 and 37 proteins are from *T. harzianum* and *T. atroviride*, respectively. The majority of *Trichoderma* proteins in Swiss-Prot / TrEMBL come from *T. reesei*, with 134 protein entries. While the public announcement of the release of the *T. reesei* genome was made by Genencor on the 25th July 2003 (http://www.genencor.com/wt/gcor/pr_1059584144), predicted protein information from the genome release is still not publicly available on protein databases including Swiss-Prot / TrEMBL. The availability of the *T. reesei* genome sequence has not been sufficient for identifying proteins by peptide mass

fingerprinting (PMF) and mass spectrometry, as protein information has not yet been translated into a suitable format for PMF analysis. Due to the limited information in public protein databases, a strategy of cross species identification (CSI) in combination with a mixed mass spectrometry approach was developed to enhance protein identifications from the proteome of *T. harzianum*. This approach involved the use of two separate mass spectrometry techniques, MALDI-TOF-MS and LC-MS/MS, to maximise the amount of peptide and amino acid sequence information from each protein. This PMF data was submitted to protein databases where all fungal and yeast protein information publicly available at the time was searched. Protein identifications were substantially increased using this approach when compared to using only MALDI-TOF-MS and database searching from *Trichoderma* spp. only (discussed in Grinyer et al. 2004a, Publication II, Section 3.1.2; Grinyer et al. 2004b, Publication III, Section 3.1.3). However, as shown in Section 3.1.5, the use of a combined mass spectrometry approach and CSI resulted in a relatively low overall protein identification frequency of 20 %. It is clear that the only way to enhance the number of proteins identified from *T. harzianum* / *T. atroviride* P1 is to have a complete genome sequence from which the predicted protein information can be derived. At the recently held 8th European Fungal Genetics Conference in Vienna, Austria, the release of the *T. atroviride* P1 genome sequence was scheduled for the second half of 2006 (personal communication, Herrera-Estrella). In future work, the PMF data for each *T. harzianum* protein extract should be re-searched against the *T. atroviride* P1 protein database to provide more protein identifications and increase the protein identification frequency.

3.3.3 Differential gel electrophoresis to identify proteins in the biological control response of *T. atroviride* when grown on *R. solani* cell walls

Some of the proteomic methods developed above were used to find potential proteins playing a role in biocontrol from the biological control agent *T. atroviride* P1 (formerly *T. harzianum* P1) when grown on the cell wall material of the phytopathogen *R. solani*. *T. atroviride* was grown on medium containing 2 % (w/v) glucose for 48 h to provide enough mycelial mass before transferred onto medium containing either 2 % (w/v) glucose or 0.1 % (w/v) cell wall material of *R. solani* as the sole carbon source. After a further 48 h of culturing, the mycelia were washed and the whole cell proteome extracted by the acid extraction method developed in the work described in Publications I and II (Herbert et al. 2006, Publication I, Section 3.1.1; Grinyer et al. 2004a, Publication II, Section 3.1.2). Differential gel electrophoresis (DIGE) was carried out and proteins that were up-regulated when grown on the cell wall of *R. solani*, in comparison to protein expression levels when grown on glucose, were targeted for identification. Proteins were identified using the combined mass spectrometry approach of MALDI-TOF-MS and LC-MS/MS. N-terminal sequencing provided additional protein information. Twenty four up-regulated protein spots were identified, some of which had a known function in the biological control response of *T. atroviride* based on previous research. Known cell wall-degrading enzymes (CWDEs) included N-acetyl-β-D-glucosaminidase and 42 kDa endochitinase which were identified from the whole cell proteome and the secretome of *T. atroviride* (Grinyer et al 2005, Publication V, Section 3.2.1) when grown on *R. solani* cell wall material. Beta-1,3-glucanase, another important CWDE expressed during the biological control response of *T. atroviride* was identified from the secretome. The identification of

known biocontrol-related proteins provided evidence that a biocontrol-like response was induced when *T. atroviride* was grown in the presence of *R. solani* cell walls.

Catabolite repression will affect protein regulation and secretion and hence the secretome of *T. atroviride*. The secreted protein fraction was investigated purely to find glucanases, one of the major families among biological control proteins that was absent from the whole cell proteome. A more thorough investigation of secreted proteomes using different carbon sources including glycerol could be conducted in future research to eliminate the effects of catabolite repression.

Three novel proteases were identified from the DIGE study described in Grinyer et al. 2005, Publication V (Section 3.2.1). These proteases were vacuolar serine protease (VSP), vacuolar protease A (VPA) and trypsin-like protease (TLP). Two genes, *vsp1* encoding vacuolar serine protease and *vpa1* encoding vacuolar protease A were also isolated in this work (described in Section 3.2.2). The amino acid sequences of both proteases were derived from the isolated gene sequences. Vacuolar protease A (an aspartyl protease) has predicted molecular mass and pI values that correctly match its position on the 2D gel (Publication V, Section 3.2.1). It was found to contain a predicted signal peptide indicating that this protease could be secreted from the fungus during a biological control response. Recently, an aspartyl protease was found to be the predominant protein secreted when *T. harzianum* CECT 2413 was grown on the cell walls of *R. solani*, *P. ultimum* and *B. cinerea* (Suarez et al. 2005) supporting the theory that VPA may also be secreted. VPA also contains a predicted N- and O-linked glycosylation site which may indicate the presence of sugars attached to this protein. If they are attached to VPA and these sugars are shown to contain evidence of sulfation

or phosphorylation, this result would explain the presence of up to six protein isoforms differing by charge visible on the 2D gel. However, further work is required to confirm if VPA is secreted from *T. atroviride* during biocontrol and if the glycosylation forms prove that sugars contain sulfation and phosphorylation modifications exist to create the pI isoforms observed on 2D gels (see Section 3.2.2).

A signal cleavage site and two N-linked glycosylation sites were predicted in the protein vacuolar serine protease (VSP) sequence. However, as above, further work is needed to confirm if this protease is secreted and if attachment of charged groups to sugars can explain the presence of two pI isoforms. The predicted amino acid sequence of VSP (derived from the gene sequence) has a molecular mass approximately 15-20 kDa higher than observed from its position on the 2D gel (Grinyer et al. 2005, Publication V, Section 3.2.1). A lack of mass spectrometry information is observed for the first 250 amino acids of this protein, indicating that further downstream processing of this protease is occurring in the fungus to produce the final protein form. A family of eukaryotic serine proteases not belonging to the subtilisin family, called pro-protein convertases, are involved in the downstream processing of proteins. They include the Kex2 protein from *S. cerevisiae* which contains a signal sequence for secretion, a pro-domain, a subtilase catalytic domain and P-domain which is unique to pro-protein convertases (reviewed in Rockwell and Thorner 2004). The processing capabilities of these pro-protein convertases provide a likely explanation for the downstream processing of VSP observed during this work. Downstream processing is also required to activate an aspartic proteinase Sapp1p from *Candida parapsilosis*, which occurs autocatalytically at acidic pH conditions or in the presence of a membrane-bound processing protease (Dostal et al. 2005). It is possible that many fungal proteases are

modified by forms of downstream processing. The significance of such modifications will become clear as further fungal protease genes are isolated and the proteins they represent are further characterised.

Whilst the gene sequence of trypsin-like protease (TLP) was not isolated using the approach of designing degenerate primers and GWPCR, several trypsin-like proteases have been purified and their genes isolated from other *Trichoderma* spp. (discussed further in Section 3.2.2). Based on the N-terminal sequence of TLP from *T. atroviride* (Grinyer et al. 2005, Publication V, Section 3.2.1), the similarity between these trypsin-like proteases appear to be high with 90 % amino acid identity across the first 10 amino acids (Suarez et al. 2004; Uchikoba et al. 2001). PCR products were obtained using several different sets of degenerate primers when gene isolation was attempted. However, when these PCR products were cloned and sequenced, they did not align with trypsin-like proteases from closely related fungal species. Several types of DNA polymerases were used and annealing temperatures varied to increase PCR specificity. The lack of specific PCR products from the trypsin-like protease gene is due to the low levels of conserved sequence observed amongst orthologous fungal genes, making the design of effective degenerative primers very difficult.

Recent literature has confirmed that both the trypsin-like protease and vacuolar protease A (aspartyl protease) of *Trichoderma* spp. are secreted under conditions inducing biological control (Suarez et al. 2004; Suarez et al. 2005). The occurrence of this protease secretion provides evidence that the differential gel electrophoresis experiment designed and applied in this work actually stimulates a biological control-like response by *Trichoderma*.

Proteomic techniques to study differential protein expression are highly useful in pin-pointing responsive proteins involved in particular biological situations. Once protein extraction methodologies and protein identification strategies have been optimised for an organism, the discovery of multiple protein targets can be made quickly by monitoring protein expression patterns and levels. In the example of finding proteins that may be involved in the biological control response of *T. atroviride*, several known cell wall-degrading enzymes and three previously unknown proteases have been implicated in the cellular response to growth in the presence of *R. solani* cell wall material.

4. Literature Cited

A

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422:198-207

Assossou O, Besson F, Rouault JP, Persat F, Brisson C, Duret L, Ferrandiz J, Mayencon M, Peyron F, Picot S (2003) Subcellular localization of 14-3-3 proteins in *Toxoplasma gondii* tachyzoites and evidence for a lipid raft-associated form. FEMS Microbiol Lett 224:161-168

B

Baek JM, Howell CR, Kenerley CM (1999) The role of an extracellular chitinase from *Trichoderma virens* Gv29-8 in the biocontrol of *Rhizoctonia solani*. Curr Genet 35:41-50

Bara MT, Lima AL, Ulhoa CJ (2003) Purification and characterization of an exo-beta-1,3-glucanase produced by *Trichoderma asperellum*. FEMS Microbiol Lett 219:81-85

Benitez T, Rincon AM, Limon MC, Codon AC (2004) Biocontrol mechanisms of *Trichoderma* strains. Int Microbiol 7:249-260

Berggren KN, Schulenberg B, Lopez MF, Steinberg TH, Bogdanova A, Smejkal G, Wang A, Patton WF (2002) An improved formulation of SYPRO Ruby protein gel stain: comparison with the original formulation and with a ruthenium II tris (bathophenanthroline disulfonate) formulation. Proteomics 2:486-498

Bermudez A, Daban JR, Garcia JR, Mendez E (1994) Direct blotting, sequencing and immunodetection of proteins after five-minute staining of SDS and SDS-treated IEF gels with Nile red. Biotechniques 16:621-624

Bjellqvist B, Ek K, Righetti PG, Gianazza E, Gorg A, Westermeier R, Postel W (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. J Biochem Biophys Methods 6:317-339

Bolar JP, Norelli JL, Harman GE, Brown SK, Aldwinckle HS (2001) Synergistic activity of endochitinase and exochitinase from *Trichoderma atroviride* (*T. harzianum*) against the pathogenic fungus (*Venturia inaequalis*) in transgenic apple plants. Transgenic Res 10:533-543

Breci L, Hattrup E, Keeler M, Letarte J, Johnson R, Haynes PA (2005) Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing. Proteomics 5:2018-2028

Brunner K, Peterbauer CK, Mach RL, Lorito M, Zeilinger S, Kubicek CP (2003) The Nag1 N-acetylglucosaminidase of *Trichoderma atroviride* is essential for chitinase induction by chitin and of major relevance to biocontrol. Curr Genet 43:289-295

Brunner K, Zeilinger S, Ciliento R, Woo SL, Lorito M, Kubicek CP, Mach RL (2005) Improvement of the fungal biocontrol agent *Trichoderma atroviride* to enhance both antagonism and induction of plant systemic disease resistance. Appl Environ Microbiol 71:3959-3965

C

Carsolio C, Benhamou N, Haran S, Cortes C, Gutierrez A, Chet I, Herrera-Estrella A (1999) Role of the *Trichoderma harzianum* endochitinase gene, *ech42*, in mycoparasitism. Appl Environ Microbiol 65:929-935

Chernushevich IV, Loboda AV, Thomson BA (2001) An introduction to quadrupole-time-of-flight mass spectrometry. J Mass Spectrom 36:849-865

Cherry JR, Fidantsef AL (2003) Directed evolution of industrial enzymes: an update. Curr Opin Biotechnol 14:438-443

Chet I (1987) Innovative approaches to plant disease control. Wiley, New York, pp 137-160

Chet I, Inbar J (1994) Biological control of fungal pathogens. Appl Biochem Biotechnol 48:37-43

Chet I, Benhamou N, Haran S (1998) *Trichoderma* and *Gliocladium*. Enzymes, biological control and commercial application, Volume 2. Taylor and Francis, London, pp 153-171

Chevallet M, Santoni V, Poinas A, Rouquie D, Fuchs A, Kieffer S, Rossignol M, Lunardi J, Garin J, Rabilloud T (1998) New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. Electrophoresis 19:1901-1909

Cohen-Kupiec R, Broglie KE, Friesem D, Broglie RM, Chet I (1999) Molecular characterization of a novel beta-1,3-exoglucanase related to mycoparasitism of *Trichoderma harzianum*. Gene 226:147-154

D

Daubenberger CA, Tisdale EJ, Curcic M, Diaz D, Silvie O, Mazier D, Eling W, Bohrmann B, Matile H, Pluschke G (2003) The N-terminal domain of glyceraldehyde-3-phosphate dehydrogenase of the apicomplexan *Plasmodium falciparum* mediates GTPase Rab2-dependent recruitment to membranes. Biol Chem 384:1227-1237

de la Cruz J, Hidalgo-Gallego A, Lora JM, Benitez T, Pintor-Toro JA, Llobell A (1992) Isolation and characterization of three chitinases from *Trichoderma harzianum*. Eur J Biochem 206:859-867

de la Cruz J, Pintor-Toro JA, Benitez T, Llobell A, Romero LC (1995a) A novel endo-beta-1,3-glucanase, BGN13.1, involved in the mycoparasitism of *Trichoderma harzianum*. J Bacteriol 177:6937-6945

de la Cruz J, Pintor-Toro JA, Benitez T, Llobell A (1995b) Purification and characterization of an endo-beta-1,6-glucanase from *Trichoderma harzianum* that is related to its mycoparasitism. J Bacteriol 177:1864-1871

de la Cruz J, Llobell A (1999) Purification and properties of a basic endo-beta-1,6-glucanase (BGN16.1) from the antagonistic fungus *Trichoderma harzianum*. Eur J Biochem 265:145-151

Delahunty C, Yates JR, 3rd (2005) Protein identification using 2D-LC-MS/MS. Methods 35:248-255

de las Mercedes DM, Limon MC, Mejias R, Mach RL, Benitez T, Pintor-Toro JA, Kubicek CP (2001) Regulation of chitinase 33 (*chit33*) gene expression in *Trichoderma harzianum*. Curr Genet 38:335-342

Delgado-Jarana J, Rincon AM, Benitez T (2002) Aspartyl protease from *Trichoderma harzianum* CECT 2413: cloning and characterization. Microbiol 148:1305-1315

Denning D, Mykytka B, Allen NP, Huang L, Al B, Rexach M (2001) The nucleoporin Nup60p functions as a Gsp1p-GTP-sensitive tether for Nup2p at the nuclear pore complex. J Cell Biol 154:937-950

Donzelli BG, Harman GE (2001a) Interaction of ammonium, glucose, and chitin regulates the expression of cell wall-degrading enzymes in *Trichoderma atroviride* strain P1. Appl Environ Microbiol 67:5643-5647

Donzelli BG, Lorito M, Scala F, Harman GE (2001b) Cloning, sequence and structure of a gene encoding an antifungal glucan 1,3-beta-glucosidase from *Trichoderma atroviride* (*T. harzianum*). Gene 277:199-208

Dostal J, Dlouha H, Malon P, Pichova I, Hruskova-Heidingsfeldova O (2005) The precursor of secreted aspartic proteinase Sapp1p from *Candida parapsilosis* can be activated both autocatalytically and by a membrane-bound processing proteinase. Biol Chem 386:791-799

Draborg H, Kauppinen S, Dalboge H, Christgau S (1995) Molecular cloning and expression in *S. cerevisiae* of two exochitinases from *Trichoderma harzianum*. Biochem Mol Biol Int 36:781-791

Dubourdieu D, Desplanques C, Villetaz J, Ribereau-Gayon P (1985) Investigations of an industrial beta-D-glucanase from *Trichoderma harzianum*. Carbohydr Res 144:277-287

F

Flores A, Chet I, Herrera-Estrella A (1997) Improved biocontrol activity of *Trichoderma harzianum* by over-expression of the proteinase-encoding gene *prb1*. Curr Genet 31:30-37

G

Galvani M, Rovatti L, Hamdan M, Herbert B, Righetti PG (2001a) Protein alkylation in the presence/absence of thiourea in proteome analysis: a matrix assisted laser desorption/ionization-time of flight-mass spectrometry investigation. Electrophoresis 22:2066-2074

Galvani M, Hamdan M, Herbert B, Righetti PG (2001b) Alkylation kinetics of proteins in preparation for two-dimensional maps: a matrix assisted laser desorption/ionization-mass spectrometry investigation. Electrophoresis 22:2058-2065

Gavin PD, Prescott M, Luff SE, Devenish RJ (2004) Cross-linking ATP synthase complexes in vivo eliminates mitochondrial cristae. J Cell Sci 117:2333-2343

Geremia RA, Goldman GH, Jacobs D, Ardiles W, Vila SB, Van Montagu M, Herrera-Estrella A (1993) Molecular characterization of the proteinase-encoding gene, *prb1*, related to mycoparasitism by *Trichoderma harzianum*. Mol Microbiol 8:603-613

Gerke V, Moss SE (2002) Annexins: from structure to function. Physiol Rev 82:331-371

Goldman GH, Hayes C, Harman GE (1994) Molecular and cellular biology of biocontrol by *Trichoderma* spp. Trends Biotechnol 12:478-482

Gooday GW (1995) The growing fungus. Chapman and Hall, London, pp 43-66

Gorg A, Weiss W, Dunn MJ (2004) Current two-dimensional electrophoresis technology for proteomics. Proteomics 4:3665-3685

Granger J, Siddiqui J, Copeland S, Remick D (2005) Albumin depletion of human plasma also removes low abundance proteins including the cytokines. Proteomics 5:4713-4718

Grinyer J, McKay M, Nevalainen H, Herbert BR (2004a) Fungal proteomics: initial mapping of biological control strain *Trichoderma harzianum*. Curr Genet 45:163-169

Grinyer J, McKay M, Herbert B, Nevalainen H (2004b) Fungal proteomics: mapping the mitochondrial proteins of a *Trichoderma harzianum* strain applied for biological control. Curr Genet 45:170-175

Grinyer J, Hunt S, McKay M, Herbert BR, Nevalainen H (2005) Proteomic response of the biological control fungus *Trichoderma atroviride* to growth on the cell walls of *Rhizoctonia solani*. Curr Genet 47:381-388

Guerlava P, Izac V, Tholozan JL (1998) Comparison of different methods of cell lysis and protein measurements in *Clostridium perfringens*: application to the cell volume determination. Curr Microbiol 36:131-135

Gutierrez A, del Rio JC, Martinez MJ, Martinez AT (1999) Fungal degradation of lipophilic extractives in *Eucalyptus globulus* wood. Appl Environ Microbiol 65:1367-1371

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17:994-999

H

Harman GE, Hayes CK, Lorito M, Broadway RM, Di Pietro A, Peterbauer C, Tronsmo A (1993) Chitinolytic enzymes of *Trichoderma harzianum*: Purification of chitobiosidase and endochitinase. Molecular Plant Pathology 83:313-318

Harman G, Latorre B, Agosin R, San Martin R, Riegel D, Nielsen P, Tronsmo A, Pearson R (1996) Biological and integrated control of *Botrytis* bunch rot of grape using *Trichoderma* spp. Biol Control 7:259-266

Harrison MJ, Wathugala IM, Tenkanen M, Packer NH, Nevalainen KM (2002) Glycosylation of acetylxylan esterase from *Trichoderma reesei*. Glycobiology 12:291-298

Hart C, Schulenberg B, Diwu Z, Leung WY, Patton WF (2003) Fluorescence detection and quantitation of recombinant proteins containing oligohistidine tag sequences directly in sodium dodecyl sulfate-polyacrylamide gels. Electrophoresis 24:599-610

Hart C, Schulenberg B, Patton WF (2004) Selective proteome-wide detection of hydrophobic integral membrane proteins using a novel fluorescence-based staining technology. Electrophoresis 25:2486-2493

Herbert BR, Molloy MP, Gooley AA, Walsh BJ, Bryson WG, Williams KL (1998) Improved protein solubility in two-dimensional electrophoresis using tributyl phosphine as reducing agent. Electrophoresis 19:845-851

Herbert B, Righetti PG (2000) A turning point in proteome analysis: sample prefractionation via multicompartment electrolyzers with isoelectric membranes. Electrophoresis 21:3639-3648

Herbert B, Galvani M, Hamdan M, Olivieri E, MacCarthy J, Pedersen S, Righetti PG (2001) Reduction and alkylation of proteins in preparation of two-dimensional map analysis: why, when, and how? Electrophoresis 22:2046-2057

Herbert B, Hopwood F, Oxley D, McCarthy J, Laver M, Grinyer J, Goodall A, Williams K, Castagna A, Righetti PG (2003) Beta-elimination: an unexpected artefact in proteome analysis. Proteomics 3:826-831

Herbert BR, Grinyer J, McCarthy JT, Isaacs M, Harry EJ, Nevalainen H, Traini MD, Hunt S, Schulz B, Laver M, Goodall AR, Packer J, Harry JL, Williams KL (2006) Improved 2-DE of microorganisms after acidic extraction. Electrophoresis 27:1630-1640

Huber LA, Pfaller K, Vietor I (2003) Organelle proteomics: implications for subcellular fractionation in proteomics. Circ Res 92:962-968

Hunt SM, Thomas MR, Sebastian LT, Pedersen SK, Harcourt RL, Sloane AJ, Wilkins MR (2005) Optimal replication and the importance of experimental design for gel-based quantitative proteomics. J Proteome Res 4:809-819

I

Inbar J, Chet I (1992) Biomimics of fungal cell-cell recognition by use of lectin-coated nylon fibers. J Bacteriol 174:1055-1059

Inbar J, Chet I (1995) The role of recognition in the induction of specific chitinases during mycoparasitism by *Trichoderma harzianum*. Microbiol 141:2823-2829

K

Karas M, Bachmann D, Bahr U, Hillenkamp F (1987) Matrix-assisted ultraviolet laser desorption of non-volatile compounds. Int J Mass Spectrom Ion Process 78:53-68

Karlberg O, Canback B, Kurland CG, Andersson SG (2000) The dual origin of the yeast mitochondrial proteome. Yeast 17:170-187

Kim KK, Fravel DR, Papavizas GC (1988) Identification of a metabolite produced by *Talaromyces flavus* as glucose oxidase and its role in the biocontrol of *Verticillium dahliae*. Phytopathol 78:488-492

Kitamoto Y, Kono R, Shimotori A, Mori N, Ichikawa Y (1987) Purification and some properties of an exo-beta-1,3-glucanase from *Trichoderma harzianum*. Agric Biol Chem 51:3385-3386

Kullnig C, Mach RL, Lorito M, Kubicek CP (2000) Enzyme diffusion from *Trichoderma atroviride* (= *T. harzianum* P1) to *Rhizoctonia solani* is a prerequisite for triggering of *Trichoderma ech42* gene expression before mycoparasitic contact. Appl Environ Microbiol 66:2232-2234

Kurland CG (1991) Codon bias and gene expression. FEBS Lett 285:165-169

L

Laschet JJ, Minier F, Kurcewicz I, Bureau MH, Trottier S, Jeanneteau F, Griffon N, Samyn B, Van Beeumen J, Louvel J, Sokoloff P, Pumain R (2004) Glyceraldehyde-3-phosphate dehydrogenase is a GABAA receptor kinase linking glycolysis to neuronal inhibition. J Neurosci 24:7614-7622

Lester PJ, Hubbard SJ (2002) Comparative bioinformatic analysis of complete proteomes and protein parameters for cross-species identification in proteomics. Proteomics 2:1392-1405

Lim D, Hains P, Walsh B, Bergquist P, Nevalainen H (2001) Proteins associated with the cell envelope of *Trichoderma reesei*: a proteomic approach. Proteomics 1:899-909

Limon MC, Limon MC, Chacon MR, Mejias R, Delgado-Jarana J, Rincon AM, Codon AC, Benitez T (2004) Increased antifungal and chitinase specific activities of *Trichoderma harzianum* CECT 2413 by addition of a cellulose binding domain. Appl Microbiol Biotechnol 64:675-685

Liu M, Sun ZX, Zhu J, Xu T, Harman GE, Lorito M (2004) Enhancing rice resistance to fungal pathogens by transformation with cell wall degrading enzyme genes from *Trichoderma atroviride*. J Zhejiang Univ Sci 5:133-136

Lora JM, de la Cruz J, Llobell A, Benitez T, Pintor-Toro JA (1995) Molecular characterization and heterologous expression of an endo-beta-1,6-glucanase gene from the mycoparasitic fungus *Trichoderma harzianum*. Mol Gen Genet 247:639-645

Lorito M, Harman GE, Hayes CK, Broadway RM, Woo SL, Di Pietro A (1993) Chitinolytic enzymes produced by *Trichoderma harzianum*. Antifungal activity of purified endochitinase and chitobiosidase. Phytopathol 83:302-307

Lorito M, Hayes CK, Di Pietro A, Woo SL, Harman GE (1994a) Purification, characterisation and synergistic activity of a glucan 1,3-β-glucosidase and an N-acetyl-β-glucosaminidase from *Trichoderma harzianum*. Mol Plant Pathol 84:398-405

Lorito M, Hayes CK, Zoina A, Scala F, Del Sorbo G, Woo SL, Harman GE (1994b) Potential of genes and gene products from *Trichoderma* spp. and *Gliocladium* spp. for the development of biological pesticides. Mol Biotechnol 2:209-217

Lorito M, Peterbauer C, Hayes CK, Harman GE (1994c) Synergistic interaction between fungal cell wall degrading enzymes and different antifungal compounds enhances inhibition of spore germination. Microbiol 140:623-629

Lorito M, Farkas V, Rebuffat S, Bodo B, Kubicek CP (1996) Cell wall synthesis is a major target of mycoparasitic antagonism by *Trichoderma harzianum*. J Bacteriol 178:6382-6385

Lorito M, Woo SL, Garcia I, Colucci G, Harman GE, Pintor-Toro JA, Filippone E, Muccifora S, Lawrence CB, Zoina A, Tuzun S, and Scala F (1998) Genes from mycoparasitic fungi as a source for improving plant resistance to fungal pathogens. Proc Natl Acad Sci U S A 95:7860-7865

Luche S, Santoni V, Rabilloud T (2003) Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. Proteomics 3:249-253

M

Mach RL, Peterbauer CK, Payer K, Jaksits S, Woo SL, Zeilinger S, Kullnig CM, Lorito M, Kubicek CP (1999) Expression of two major chitinase genes of *Trichoderma atroviride* (*T. harzianum* P1) is triggered by different regulatory signals. Appl Environ Microbiol 65:1858-1863

Mackintosh JA, Choi HY, Bae SH, Veal DA, Bell PJ, Ferrari BC, Van Dyk DD, Verrills NM, Paik YK, Karuso P (2003) A fluorescent natural product for ultra sensitive detection of proteins in one-dimensional and two-dimensional gel electrophoresis. Proteomics 3:2273-2288

Mano N, Goto J (2003) Biomedical and biological mass spectrometry. Anal Sci 19:3-14

Mao W, Lewis JL, Lumsden RD, Hebbar KP (1998) Biocontrol of selected soilborne diseases of tomato and pepper plants. Crop Protection 17:535-542

McCarthy J, Hopwood F, Oxley D, Laver M, Castagna A, Righetti PG, Williams K, Herbert B (2003) Carbamylation of proteins in 2-D electrophoresis-myth or reality? J Proteome Res 2:239-242

Migheli Q, Gonzalez-Candelas L, Dealessi L, Camponogara A, Ramon-Vidal D (1998) Transformants of *Trichoderma longibrachiatum* overexpressing the β-1,4-endoglucanase gene *egl1* show enhanced biocontrol of *Pythium ultimum* on cucumber. Phytopathol 88:673-677

Mignogna G, Giorgi A, Stefanelli P, Neri A, Colotti G, Maras B, Schinina ME (2005) Inventory of the proteins in *Neisseria meningitidis* serogroup B strain MC58. J Proteome Res 4:1361-1370

Molloy MP, Herbert BR, Walsh BJ, Tyler MI, Traini M, Sanchez JC, Hochstrasser DF, Williams KL, Gooley AA (1998) Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. Electrophoresis 19:837-844

Molloy MP, Herbert BR, Williams KL, Gooley AA (1999) Extraction of *Escherichia coli* proteins with organic solvents prior to two-dimensional electrophoresis. Electrophoresis 20:701-704

Molloy MP, Herbert BR, Slade MB, Rabilloud T, Nouwens AS, Williams KL, Gooley AA (2000) Proteomic analysis of the *Escherichia coli* outer membrane. Eur J Biochem 267:2871-2881

Monti M, Orru S, Pagnozzi D, Pucci P (2005) Interaction proteomics. Biosci Rep 25:45-56

Morris DD, Reeves RA, Gibbs MD, Saul DJ, Bergquist PL (1995) Correction of the beta-mannanase domain of the *celC* pseudogene from *Caldocellulosiruptor saccharolyticus* and activity of the gene product on kraft pulp. Appl Environ Microbiol 61:2262-2269

Morris DD, Gibbs MD, Chin CW, Koh MH, Wong KK, Allison RW, Nelson PJ, Bergquist PL (1998) Cloning of the *xynB* gene from *Dictyoglomus thermophilum* Rt46B.1 and action of the gene product on kraft pulp. Appl Environ Microbiol 64:1759-1765

Mukherjee PK, Latha J, Hadar R, Horwitz BA (2003) TmkA, a mitogen-activated protein kinase of *Trichoderma virens*, is involved in biocontrol properties and repression of conidiation in the dark. Eukaryot Cell 2:446-455

Mulholland J, Preuss D, Moon A, Wong A, Drubin D, Botstein D (1994) Ultrastructure of the yeast actin cytoskeleton and its association with the plasma membrane. J Cell Biol 125:381-391

Murray FR, Llewellyn DJ, Peacock WJ, Dennis ES (1997) Isolation of the glucose oxidase gene from *Talaromyces flavus* and characterisation of its role in the biocontrol of *Verticillium dahliae*. Curr Genet 32:367-375

Musante L, Candiano G, Ghiggeri GM (1998) Resolution of fibronectin and other uncharacterized proteins by two-dimensional polyacrylamide electrophoresis with thiourea. J Chromatogr B Biomed Sci Appl 705:351-356

Myer VE, Young RA (1998) RNA polymerase II holoenzymes and subcomplexes. J Biol Chem 273:27757-27760

N

Nandakumar MP, Marten MR (2002) Comparison of lysis methods and preparation protocols for one- and two-dimensional electrophoresis of *Aspergillus oryzae* intracellular proteins. Electrophoresis 23:2216-2222

Nikolaizik WH, Weichel M, Blaser K, Crameri R (2002) Intracutaneous tests with recombinant allergens in cystic fibrosis patients with allergic bronchopulmonary aspergillosis and *Aspergillus* allergy. Am J Respir Crit Care Med 165:916-921

Noronha EF, Ulhoa CJ (2000) Characterization of a 29-kDa beta-1,3-glucanase from *Trichoderma harzianum*. FEMS Microbiol Lett 183:119-123

P

Papavizas G (1985) *Trichoderma* and *Gliocladium*: Biology, ecology, and potential for biocontrol. Ann Rev Phytopathol 23:23-54

Peberdy JF (1994) Protein secretion in filamentous fungi--trying to understand a highly productive black box. Trends Biotechnol 12:50-57

Pedersen SK, Harry JL, Sebastian L, Baker J, Traini MD, McCarthy JT, Manoharan A, Wilkins MR, Gooley AA, Righetti PG, Packer NH, Williams KL, Herbert BR (2003) Unseen proteome: mining below the tip of the iceberg to find low abundance and membrane proteins. J Proteome Res 2:303-311

Pertl H, Gehwolf R, Obermeyer G (2005) The distribution of membrane-bound 14-3-3 proteins in organelle-enriched fractions of germinating lily pollen. Plant Biol (Stuttg) 7:140-147

Peterbauer CK, Lorito M, Hayes CK, Harman GE, Kubicek CP (1996) Molecular cloning and expression of the *nag1* gene (N-acetyl-beta-D-glucosaminidase-encoding gene) from *Trichoderma harzianum* P1. Curr Genet 30:325-331

Pozo MJ, Baek JM, Garcia JM, Kenerley CM (2004) Functional analysis of *tvsp1*, a serine protease-encoding gene in the biocontrol agent *Trichoderma virens*. Fungal Genet Biol 41:336-348

Prins TW, Tudzynski P, Von Tiedmann A, Tudzynski B, Ten Have A, Hansen ME, Tenberge K, van Kan JAL (2000) Fungal Pathology. Kluwer, Dordrecht, pp 33-64

Q

Qian WJ, Qian WJ, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, Kangas LJ, Petritis K, Camp DG 2nd, Smith RD (2005) Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J Proteome Res 4:53-62

R

Rabilloud T, Adessi C, Giraudel A, Lunardi J (1997) Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 18:307-316

Rabilloud T (1998) Use of thiourea to increase the solubility of membrane proteins in two-dimensional electrophoresis. Electrophoresis 19:758-760

Ramot O, Cohen-Kupiec R, Chet I (2000) Regulation of β-1,3-glucanase by carbon starvation in the mycoparasite *Trichoderma harzianum*. Mycol Res 104:415-420

Ramot O, Viterbo A, Friesem D, Oppenheim A, Chet I (2004) Regulation of two homodimer hexosaminidases in the mycoparasitic fungus *Trichoderma asperellum* by glucosamine. Curr Genet 45:205-213

Rawlings ND, Barrett AJ (1993) Evolutionary families of peptidases. Biochem J 290:205-218

Rescher U, Ruhe D, Ludwig C, Zobiack N, Gerke V (2004) Annexin 2 is a phosphatidylinositol (4,5)-bisphosphate binding protein recruited to actin assembly sites at cellular membranes. J Cell Sci 117:3473-3480

Righetti PG, Drysdale JW (1973) Small-scale fractionation of proteins and nucleic acids by isoelectric focusing in polyacrylamide gels. Ann N Y Acad Sci 209:163-186

Righetti PG, Castagna A, Herbert B, Reymond F, Rossier JS (2003) Prefractionation techniques in proteome analysis. Proteomics 3:1397-1407

Righetti PG, Castagna A, Herbert B, Candiano G (2005) How to bring the "unseen" proteome to the limelight via electrophoretic pre-fractionation techniques. Biosci Rep 25:3-17

Righetti PG (2006) Real and imaginary artefacts in proteome analysis via two-dimensional maps. J Chromatogr B Analyt Technol Biomed Life Sci (In Press)

Rockwell NC, Thorner JW (2004) The kindest cuts of all: crystal structures of Kex2 and furin reveal secrets of precursor processing. Trends Biochem Sci 29:80-87

S

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463-5467

Santoni V, Rabilloud T, Doumas P, Rouquie D, Mansion M, Kieffer S, Garin J, Rossignol M (1999) Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels. Electrophoresis 20:705-711

Schirmbock M, Lorito M, Wang YL, Hayes CK, Arisan-Atac I, Scala F, Harman GE, Kubicek CP (1994) Parallel formation and synergism of hydrolytic enzymes and peptaibol antibiotics, molecular mechanisms involved in the antagonistic action of *Trichoderma harzianum* against phytopathogenic fungi. Appl Environ Microbiol 60:4364-4370

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

Shanmugam V, Sriram S, Babu S, Nandakumar R, Raguchander T, Balasubramanian P, Samiyappan R (2001) Purification and characterization of an extracellular alpha-glucosidase protein from *Trichoderma viride* which degrades a phytotoxin associated with sheath blight disease in rice. J Appl Microbiol 90:320-329

Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 68:850-858

Shi Y, Xiang R, Horvath C, Wilkins JA (2004) The role of liquid chromatography in proteomics. J Chromatogr A 1053:27-36

Smejkal GB (2004) The Coomassie chronicles: past, present and future perspectives in polyacrylamide gel staining. Expert Rev Proteomics 1:381-387

Smejkal GB, Robinson MH, Lawrence NP, Tao F, Saravis CA, Schumacher RT (2006) Increased protein yields from *Escherichia coli* using pressure-cycling technology. J Biomol Tech 17:173-175

Stasyk T, Huber LA (2004) Zooming in: fractionation strategies in proteomics. Proteomics 4:3704-3716

Steinberg TH, Haugland RP, Singer VL (1996) Applications of SYPRO orange and SYPRO red protein gel stains. Anal Biochem 239:238-245

Steinberg TH, Agnew BJ, Gee KR, Leung WY, Goodman T, Schulenberg B, Hendrickson J, Beechem JM, Haugland RP, Patton WF (2003) Global quantitative phosphoprotein analysis using multiplexed proteomics technology. Proteomics 3:1128-1144

Stosz SK, Fravel DR, Roberts DP (1996) In vitro analysis of the role of glucose oxidase from *Talaromyces flavus* in biocontrol of the plant pathogen *Verticillium dahliae*. Appl Environ Microbiol 62:3183-3186

Suarez B, Rey M, Castillo P, Monte E, Llobell A (2004) Isolation and characterization of PRA1, a trypsin-like protease from the biocontrol agent *Trichoderma harzianum* CECT 2413 displaying nematicidal activity. Appl Microbiol Biotechnol 65:46-55

Suarez MB, Sanz L, Chamorro MI, Rey M, Gonzalez FJ, Llobell A, Monte E (2005) Proteomic analysis of secreted proteins from *Trichoderma harzianum*. Identification of a fungal cell wall-induced aspartic protease. Fungal Genet Biol 42:924-934

T

Tastet C, Charmont S, Chevallet M, Luche S, Rabilloud T (2003) Structure-efficiency relationships of zwitterionic detergents as protein solubilizers in two-dimensional electrophoresis. Proteomics 3:111-121

U

Unlu M, Morgan ME, Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis 18:2071-2077

Uchikoba T, Mase T, Arima K, Yonezawa H, Kaneda M (2001) Isolation and characterization of a trypsin-like protease from *Trichoderma viride*. Biol Chem 382:1509-1513

V

Villamon E, Villalba V, Nogueras MM, Tomas JM, Gozalbo D, Gil ML (2003) Glyceraldehyde-3-phosphate dehydrogenase, a glycolytic enzyme present in the periplasm of *Aeromonas hydrophila*. Antonie Van Leeuwenhoek 84:31-38

Viterbo A, Haran S, Friesem D, Ramot O, Chet I (2001) Antifungal activity of a novel endochitinase gene (*chit36*) from *Trichoderma harzianum* Rifai TM. FEMS Microbiol Lett 200:169-174

Viterbo A, Ramot O, Chemin L, Chet I (2002a) Significance of lytic enzymes from *Trichoderma* spp. in the biocontrol of fungal plant pathogens. Antonie Van Leeuwenhoek 81:549-556

Viterbo A, Montero M, Ramot O, Friesem D, Monte E, Llobell A, Chet I (2002b) Expression regulation of the endochitinase *chit36* from *Trichoderma asperellum* (*T. harzianum* T-203). Curr Genet 42:114-122

Viterbo A, Harel M, Chet I (2004) Isolation of two aspartyl proteases from *Trichoderma asperellum* expressed during colonization of cucumber roots. FEMS Microbiol Lett 238:151-158

Voigt J, Liebich I, Kiess M, Frank R (2001) Subcellular distribution of 14-3-3 proteins in the unicellular green alga *Chlamydomonas reinhardtii*. Eur J Biochem 268:6449-6457

W

Wa C, Cerny R, Hage DS (2006) Obtaining high sequence coverage in matrix-assisted laser desorption time-of-flight mass spectrometry for studies of protein modification: Analysis of human serum albumin as a model. Anal Biochem 349:229-241

Washburn MP, Wolters D, Yates JR, 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19:242-247

Westermeier R, Marouga R (2005) Protein detection methods in proteomics research. Biosci Rep 25:19-32

Wildgruber R, Harder A, Obermaier C, Boguth G, Weiss W, Fey SJ, Larsen PM, Gorg A (2000) Towards higher resolution: two-dimensional electrophoresis of *Saccharomyces cerevisiae* proteins using overlapping narrow immobilized pH gradients. Electrophoresis 21:2610-2616

Wilkins MR, Williams KL (1997) Cross-species protein identification using amino acid composition, peptide mass fingerprinting, isoelectric point and molecular mass: a theoretical evaluation. J Theor Biol 186:7-15

Williams J, Clarkson JM, Mills PR, Cooper RM (2003) Saprotrophic and mycoparasitic components of aggressiveness of *Trichoderma harzianum* groups toward the commercial mushroom *Agaricus bisporus*. Appl Environ Microbiol 69:4192-4199

Wu C, Te'o VSJ, Farrell RA, Bergquist PL, Nevalainen KMH (2006) Improvement of the secretion of extracellular proteins and isolation and characterization of the amylase I (*amy1*) gene from *Ophiostoma floccosum*. Gene (In Press)

Wu CC, Yates JR, 3rd (2003) The application of mass spectrometry to membrane proteomics. Nat Biotechnol 21:262-267

Wysocki VH, Resing KA, Zhang Q, Cheng G (2005) Mass spectrometry of peptides and proteins. Methods 35:211-222

Z

Zeilinger S, Galhaup C, Payer K, Woo SL, Mach RL, Fekete C, Lorito M, Kubicek CP (1999) Chitinase gene expression during mycoparasitic interaction of *Trichoderma harzianum* with its host. Fungal Genet Biol 26:131-140

Zolotarjova N, Martosella J, Nicol G, Bailey J, Boyes BE, Barrett WC (2005) Differences among techniques for high-abundant protein depletion. Proteomics 5:3304-3313

5. Appendices

5.1 β-elimination: an unexpected artefact in proteome analysis (Publication VI)

Publication VI

Due to copyright laws, the following article has been removed from this thesis. Please refer to the following link for abstract details.

Herbert B, Hopwood F, Oxley D, McCarthy J, Laver M, Grinyer J, Goodall A, Williams K, Castagna A, Righetti PG (2003) Beta-elimination: an unexpected artefact in proteome analysis. Proteomics 3:826-831

http://dx.doi.org/10.1002/pmic.200300414

5.2 Genbank entry for *vpa1*

```
LOCUS       DQ661006                2209 bp    DNA     linear   PLN 30-JUN-2006
DEFINITION  Trichoderma atroviride vacuolar protease A (VPA1) gene, complete
            cds.
ACCESSION   DQ661006
VERSION     DQ661006
KEYWORDS    .
SOURCE      Trichoderma atroviride
  ORGANISM  Trichoderma atroviride
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; mitosporic Hypocreales;
            Trichoderma.
REFERENCE   1  (bases 1 to 2209)
  AUTHORS   Grinyer,J., Te'o,J. and Nevalainen,H.
  TITLE     Proteome analysis of the biological control fungus Trichoderma
            atroviride
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2209)
  AUTHORS   Grinyer,J., Te'o,J. and Nevalainen,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2006) Chemistry and Biomolecular Sciences,
            Macquarie University, Herring Rd, Sydney, NSW 2109, Australia
FEATURES             Location/Qualifiers
     source          1..2209
                     /organism="Trichoderma atroviride"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:63577"
     gene            <399..>1720
                     /gene="VPA1"
     mRNA            join(<399..497,564..717,786..>1720)
                     /gene="VPA1"
                     /product="vacuolar protease A"
     CDS             join(399..497,564..717,786..1720)
                     /gene="VPA1"
                     /codon_start=1
                     /product="vacuolar protease A"
                     /protein_id="ABG57251"
                     /translation="MKSALIAAAALVGSAQAGVHKMKLQKVSLEQQLEGSSIEAQVQQ
                     LGQKYMGVRPTSRVDVMFNDNVPKVKGGHPVPVTNFMNAQYFSEITIGSPPQTFKVVL
                     DTGSSNLWVPSQSCNSIACFLHSTYDSSSSSSYKKNGSDFEIHYGSGSLTGFISNDVV
                     TIGDLQIKGQDFAEATSEPGLAFAFGRFDGILGLGYDTISVNGIVPPFYQMVNQKLLD
                     EPVFAFYLGSGDEGSVATFGGVDESHYSGKIEYIPLRRKAYWEVDLDSIAFGDEVAEL
                     ENTGAILDTGTSLNVLPSGIAELLNAEIGAKKGYGGQYTIDCAKRDSLPDITFSLAGS
                     KYSLPASDYILEVSGSCISTFQGMDFPEPVGPLVILGDAFLRRYYSVYDLGKGAVGLA
                     KAK"
ORIGIN
        1 ctgctcgggc ggtactcaaa gggccccctt gcagggccgc cgtgcgtcat cagcccgtca
       61 gcacaggcac caataggcgc gacagcacgc aaacggcgcg agcacgaggc ctaacccccc
      121 gttcgccctg tttgattagc ccaggtccag tgacctcttg gtctgtagca gcggtagcat
      181 caccgccagc caccgatgct acaccagagg ccgaccagta ctaagctgcc tgccgccttc
      241 ttatctaagt tcccgctact gctgctggcg ccactacaaa ggcagcttcg tcatcccttc
      301 tcctcgggct gctcttccaa tccatcaact caccatcctc catccatcac ctctcacccc
      361 aaaaacaccc atcatcaagt tcccaagtgt cgcccagtat gaagagcgca ttgattgccg
      421 ctgcggcgct tgtcggctcc gcccaagccg gcgtccacaa gatgaagctg caaaaggttt
      481 ccctggagca gcagctggta tgaacccgcc ccaatctgcc tctgcaccgt cgcacgagca
      541 ctaacgcttc tcttttcaa caggagggct catccatcga ggcccaagtc cagcagctcg
      601 gccagaagta catgggcgtg cgccctacta gccgtgtcga tgtcatgttc aacgacaacg
      661 tgcccaaggt caagggcggc cacccggtcc ccgtcaccaa cttcatgaat gcccaatgtg
      721 agtcaacgtg caatggcact actccagcaa gcgagtgtag cgctttactg acttgaatgt
      781 gacagacttc tccgagatca ccatcggctc tcctccccag actttcaagg tcgtccttga
      841 cacgggaagc tccaacctct gggttccctc gcagtcctgc aacagcattg cctgcttcct
      901 gcactccacg tacgattcgt cttcctcgtc atcgtacaag aagaatggat ccgatttcga
      961 aatccactac ggatcgggta gcttgactgg attcatctcc aacgatgttg tcactatcgg
     1021 tgacctccag atcaagggcc aggactttgc cgaggctacc agcgagcccg gccttgcctt
     1081 tgcctttggc cgctttgacg gcattcttgg ccttggctac gataccattt ccgtcaatgg
     1141 tattgttcct ccctttacc agatggtgaa ccagaagctt ctggacgagc ctgttttcgc
     1201 cttctacctt ggaagcggcg acgaaggttc tgtggctacc tttggtggcg ttgatgagtc
     1261 ccactactcg ggcaagattg agtacattcc tctccgccgc aaggcttatt gggaggttga
     1321 ccttgactcc attgcctttg gcgatgaggt cgccgagctt gagaacactg gtgccatcct
     1381 cgacactggt acctcgctca acgttctccc ctctggcatc gccgagctgc tcaatgctga
     1441 gattggcgcc aagaagggct acggtggcca gtacaccatc gactgtgcca agcgcgactc
     1501 cctccccgac attaccttca gccttgccgg ctccaagtac agcctccctg cttctgacta
     1561 catccttgag gtgtctggca gctgcatctc taccttccag ggcatggact tccccgagcc
     1621 cgtgggcccc ctggtcatcc tcggcgatgc tttcctgcgc cgatactact cagtctatga
```

```
1681 cctcggcaag ggagctgtcg gtcttgccaa ggccaaataa aggaataata tagtacacag
1741 tggaatgaat gtgttgcaag gatgggcaca gcttatatgt tgctagaagg cagttttggt
1801 taattgagcg cctttgccta tatagtactg cgaggtgtac ggatcaagat tgcagcgccg
1861 tttatgtaat ataatggaac acgggtactt gttttgctgc tctaatttta gttgatattc
1921 acaaattgat cctgttttca atcttttgca tttatttggt gaatggcttg acgcgtaaac
1981 ttgaggtggc tacagcgttt gtggctgctt tgtgtgtatt tggtttccag taggtattcc
2041 tttttggtgg tgtctggtac aacaatctac ccatcgccta cctggtaggt actcaaaaga
2101 atatggttaa tacaggtttg tatcacggta agatctggta atcagtgaga tttcgtgcat
2161 taacattggc accctgcctt ggtacagagc tgaaagtcag cgtgggatc
```

5.3 Genbank entry for *vsp1*

```
LOCUS       DQ661007                2071 bp    DNA     linear   PLN 30-JUN-2006
DEFINITION  Trichoderma atroviride vacuolar serine protease (VSP1) gene,
            complete cds.
ACCESSION   DQ661007
VERSION     DQ661007
KEYWORDS    .
SOURCE      Trichoderma atroviride
  ORGANISM  Trichoderma atroviride
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; mitosporic Hypocreales;
            Trichoderma.
REFERENCE   1  (bases 1 to 2071)
  AUTHORS   Grinyer,J., Te'o,J. and Nevalainen,H.
  TITLE     Proteome analysis of the biological control fungus Trichoderma
            atroviride
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2071)
  AUTHORS   Grinyer,J., Te'o,J. and Nevalainen,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2006) Chemistry and Biomolecular Sciences,
            Macquarie University, Herring Rd, Sydney, NSW 2109, Australia
FEATURES             Location/Qualifiers
     source          1..2071
                     /organism="Trichoderma atroviride"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:63577"
     gene            <268..>1946
                     /gene="VSP1"
     mRNA            join(<268..633,699..>1946)
                     /gene="VSP1"
                     /product="vacuolar serine protease"
     CDS             join(268..633,699..1946)
                     /gene="VSP1"
                     /codon_start=1
                     /product="vacuolar serine protease"
                     /protein_id="ABG57252"
                     /translation="MRSVIALSVAAVAQASSFQVGTIHESSAPVLSNIEANAIPDNYI
                     IKFKDHVDEAGADKHQNWIQSIHDEGEQQRLELRKRSSIFGADEAFDGLKHTFKIGDF
                     KGYAGHFTEDIIEQVRNHPDVEYIERDTIVHTMLPLDSQDSVVVEDSCNPETEKQAPW
                     GLARISHRDTLGFGTFNKYLYTADGGEGVDAYVIDTGTNIEHVDFEGRAKWGKTIPAG
                     DEDEDGNGHGTHCSGTVAGKKYGVAKKAHVYAVKVLRSNGSGTMSDVVKGVEYAALSH
                     QEQVKQAKAGKRKGFKGSVANMSLGGGKTSALDAAVNAAVKAGVHFAVAAGNDNADAC
                     NYSPAAATEPVTVGASALDDSRAYFSNYGKCTDIFAPGLSIQSTWIGSKYAVNTISGT
                     SMASPHIAGLLAYYLSLQPAGDSEFAVAPITPNKLKKDLITIATQGTLSDIPADTPNL
                     LAWNGAGCSNYSQIVKAGGYKAKAQKQTKLPATIEELEDAIEGDFEVVSGKIVKGAKS
                     FGSKAEKFAKKIHDLVEEEIDEFMSELSE"
ORIGIN
        1 gctgaggcag cactggagcc tcgtcccagc cgcccttatc tactgcgcca tcgctcgtct
       61 cttcgtcctc tttctttctc ttcttccatc atccttctcc ccatccatcc atcgccgtct
      121 cttctcccat ctatctgccc atctcgcagg aggccgttgg agttttcttg atttccgtca
      181 cctttgtttt cgtctcttct ggccgcgtgt cgcaggttga ccaactccca tattccctca
      241 ttcatcacat ctatcaggac catcattatg cggtccgtca tcgctctctc tgtggcggcc
      301 gtcgcccagg ccagctcgtt ccaggttggc accatccacg agagctctgc tcccgtcctc
      361 agcaatattg aggccaacgc cattccggac aactacatca tcaaattcaa ggatcacgtt
      421 gatgaggctg gtgccgacaa gcaccagaac tggatccaga gcatccacga tgaaggcgag
      481 cagcagcgcc ttgagctccg caagcgaagc agcatcttcg gcgcggacga ggcctttgac
      541 ggcctgaagc acaccttcaa gattggtgat ttcaagggct acgctggtca ctttaccgag
      601 gacatcatcg agcaggtccg gaaccacccg gatgtaagtt cacgctacca tttccatcgc
      661 cactggcagg tcaattgagg ctaatccatc gtttccaggt cgagtatatt gagcgcgaca
      721 ccattgtcca cactatgctt cccctcgact cacaagacag cgttgttgtc gaggactcgt
      781 gcaaccccga gaccgagaag caggctccct ggggcttagc ccgtatctct caccgagaca
      841 ccctgggctt cggcacattt aacaagtacc tctacaccgc tgacggtggc gagggtgttg
      901 atgcctatgt cattgacact ggtaccaaca ttgagcacgt cgacttcgag ggtcgtgcca
      961 agtggggcaa gaccatccct gccggcgatg aggatgagga tggcaacggc cacggcactc
     1021 actgctctgg taccgttgct ggtaagaagt acggagttgc caagaaggcc catgtctacg
     1081 ccgtcaaggt cctccgatcc aacggctctg gcaccatgag cgatgtcgtc aagggcgtcg
     1141 agtacgccgc tctctcccac caggagcagg ttaagcaggc caaggccggc aagcgaaagg
     1201 gcttcaaggg ctcagtcgcc aacatgtctc ttggtggtgg caagacctct gctctcgacg
     1261 ccgctgtcaa cgctgccgtc aaggccggtg ttcacttcgc cgtcgctgct ggcaacgaca
     1321 acgccgatgc ctgcaactac tcccccgccg ccgccaccga gcccgtcacc gtcggtgcct
     1381 ctgctctgga cgacagccgt gcttacttct ccaactacgg caagtgcact gacatcttcg
     1441 cccctggctt gagcatccag tccacctgga tcggctccaa gtatgccgtc aacaccatct
     1501 ctggtacctc catggcgtct ccccacatcg ccggtctcct ggcctactac ctgtctctcc
```

```
1561 agcctgctgg tgactctgag tttgctgttg cccccatcac ccccaacaag ctcaagaagg
1621 acctcatcac cattgctacc cagggtactc tgtctgatat ccccgccgat acccccaacc
1681 tgcttgcctg gaacggtgct ggctgcagca actactccca gatcgtcaag gccggtggct
1741 acaaggccaa ggcccagaag cagaccaagc tccccgctac cattgaggag ctcgaggacg
1801 ccatcgaggg tgactttgag gttgtctctg gcaagattgt caagggtgcc aagtcctttg
1861 gctccaaggc tgagaagttc gccaagaaga tccacgacct cgtcgaggag gagattgacg
1921 agttcatgtc cgagctctcc gagtaagaat ctttattaag catatatttc ctttttattt
1981 gtaatggcca ttatgcgccg atgagtacgt cgagtacgac gggttatgga acgggaccag
2041 gagagactat ttgttttggg taattggcat g
```