Manipulation of the host cell response to infection with *Toxoplasma gondii*

Morag M Nelson

BSc

University *of*

Glasgow

This thesis is submitted for the degree of Doctor of Philosophy

Division of Infection and Immunity

Faculty of Biomedical and Life Sciences

University of Glasgow

June 2006

Abstract

The principal aim of the work presented in this thesis was to investigate the host-pathogen interaction in *Toxoplasma gondii* infected cells and to elucidate which proteins and pathways were involved in the host response to infection. *Toxoplasma gondii* is of clinical importance for both human and veterinary disease. There are currently no drugs available to treat the chronic stage of disease. The transcriptional response of host cells to *Toxoplasma* infection has previously been investigated using microarrays. Proteins are the functional molecules in the cell and the protein expression response is likely to differ from the transcriptional response due to gene expression control and post translational modifications. The advantage of a global proteomic expression profiling study of *Toxoplasma gondii* infected cells is that, in principle, all possible host responses could be identified without pre-supposing a specific mechanism. Advanced proteomic methods were developed and used to investigate the host protein response to infection with *Toxoplasma gondii*. A conventional gel-to-gel approach was used in concert with difference gel electrophoresis (DIGE) to investigate the protein expression in infected cells. This part of the work concluded that the patterns of protein expression in infected cells showed clear differences from that of the non-infected cell. This resulted in the unambiguous identification of one hundred and fifty seven host proteins which changed in expression during infection with *Toxoplasma gondii*. Many proteins not previously implicated in response to infection were identified. The results from the proteomic studies were compared with the transcriptional study in order to determine the gene to protein correlation in *Toxoplasma gondii* infected cells. A weakly positive correlation between gene and protein expression was determined. To investigate whether the host response is uniquely tailored to an individual pathogen or whether cells have a common response to infection, the protein response to a second intracellular parasite, *Leishmania major*, was investigated. Far fewer changes were seen in the *L.major* infected cell proteome, proving that the host response to infection is unique to the pathogen. Phosphorylation is one the most common post-translational modifications therefore the host cell phosphoproteome was investigated alongside the steady state proteome of infected cells using three techniques and proteins specifically phosphorylated due to the presence of the parasite were identified. Functional analysis of the modulated proteins revealed extensive modulation of many pathways and functions within the infected host cell including glycolytic, mitotic and structural proteins. Over one third of all the modulated proteins identified were mitochondrial. Overall the findings detailed in this thesis indicate that the host cell has a highly specific response to *Toxoplasma gondii* infection and that many host cell functions are modulated due to the presence of the parasite and reveal the intimate molecular relationship between host and parasite.

Table of Contents

List of Tables

Chapter I Introduction

Chapter II Optimisation of proteomic techniques for human foetal foreskin fibroblasts (HFFs)

Chapter III Proteomic analysis of *Toxoplasma gondii* infected cells

Chapter IV Analysis of differential protein expression in *Toxoplasma gondii* infected cells by Difference Gel Electrophoresis

Chapter V Investigation into the gene to protein correlation in HFF cells during infection with *Toxoplasma gondii*

Chapter VI The proteome of *Leishmania* sp infected cells

Chapter VII Analysis of the phosphoproteome of *Toxoplasma gondii* infected cells

Chapter VIII Functional analysis of the differentially expressed proteins in *Toxoplasma gondii* infected cells

Chapter IX Summary and Forward Perspectives

List of Figures

Chapter I Introduction

Chapter II Optimisation of proteomic techniques for human foetal foreskin fibroblasts (HFFs)

Chapter III Proteomic analysis of *Toxoplasma gondii* infected cells

Chapter IV Analysis of differential protein expression in *Toxoplasma gondii* infected cells by Difference Gel Electrophoresis

Chapter V Investigation into the gene to protein correlation in HFF cells during infection with *Toxoplasma gondii*

Chapter VI The proteome of *Leishmania* sp infected cells

Chapter VII Analysis of the phosphoproteome of *Toxoplasma gondii* infected cells

Chapter VIII Functional analysis of the differentially expressed proteins in *Toxoplasma gondii* infected cells

Chapter IX Summary and Forward Perspectives

List of Abbreviations

#	Number
<	Less than
µA	Micro amperes
µg	Microgram
µl	Microlitre
1-DE	One dimensional electrophoresis
2-DE	Two dimensional electrophoresis
AA	Amino Acid
ADP	Adenosine di-phosphate
AMP	Adenosine mono-phosphate
APS	Ammonium persulfate
ATP	Adenosine tri-phosphate
Bac	Benzyl hexadecyl ammonium chloride
BCA	Bicinchonic acid
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
BVA	Batch variation analysis
C	Centigrade
CA	Carrier ampholyte
cDNA	Complementary DNA
CHCA	α cyano hydroxyl cinnamic acid
CID	Collision induced dissociation
cm	Centimetre
CMV	Cytomegalovirus
CO_2	Carbon dioxide
Da	Dalton
ddH2O	Double distilled water
DIA	Differential In-Gel Analysis
DIGE	Difference gel electrophoresis
DMEM	Dulbecco's modified eagles medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
e.g.	for example
ECACC	European Collection of Cell Cultures
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol-bis N,N,N',N'-tetra acetic acid
EMBL	European Molecular Biology Laboratory
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
EST	Expressed sequence tag
FAs	Fatty acids
FCS	Foetal calf serum

FS	Filter sterile
gI number	Gene identification number
GRO1	Growth-regulated protein alpha
GRO2	Macrophage inflammatory protein 2
H	Hours
HCl	Hydrochloric acid
HFF	Human foetal foreskin fibroblast
HIV	Human Immunodeficiency Virus
hpi	Hours post infection
HPRD	Human Protein Reference Database
HSP	Heat shock protein
HUGO	Human Genome Organisation
i.e.	Id est, that is
IAA	Iodoacetamide
ICAM-1	Intracellular adhesion molecule 1
ICAT	Isotope coded affinity tags
Id.	Identified
IEF	Isoelectric focusing
IF	Intermediate filaments
IFNγ	Interferon gamma
IgG	Immunoglobulin G
IL	Interleukin
IMDM	Iscove's modified dulbecco's medium
Inf	Infected
IPG	Immobilised pH gradient
IR	Immune response
ISG	Immune stimulated genes
K	1000
Kb	Kilobase
KCl	Potassium chloride
KH_2PO_4	Potassium hydrogen phosphate
L	Litre
LB	Lysis buffer
LDL	Low density lipoprotein
LIF	Leukemia inhibitory factor
LN_2	Liquid nitrogen
M	Methionine
MA	Microarray
mA	MilliAmp
MAGE	Microarray and gene expression
MALDI	Matrix-assisted laser desorption ionisation
Mb	Megabase
MCP-1	Monocyte chemotactic protein 1
MeOH	Methanol
mg/ml	Milligrams per millilitre
$MgCl_2$	Magnesium chloride
MHC	Major histocompatibilty complex

Min	Minutes
MIPS	Munich Information Centre for Protein Sequences
Ml	Millilitre
mM	MilliMolar
MOWSE	Molecular Weight Search
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MS/MS	Tandem MS
MTs	Microtubules
MW	Molecular weight
Na_2HPO_4	Sodium hydrophosphate
NaCl	Sodium chloride
NCBI	National Centre for Biotechnology Information
NFκβ	Nuclear factor kappa beta
Ng	Nanogram
NK	Natural killer
NL	Non-linear
Nm	Nanometre
No.	Number
NO	Nitric oxide
non-inf	Non-infected
O_2	Oxygen
P	Passage number
P.M	Peptides matched
P^{32}	Radiolabelled phosphate
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Pen/Strep	Penicillin/streptomycin
pg	Picogram
pI	Isoelectric point
PMF	Peptide mass fingerprint
pmoles	Picomoles
PMSF	Phenylmethylsulfonyl fluoride
PMT	Photomultiplier tube
PP	Phosphorylated
PRDX	Peroxiredoxin
PTM	Post translational modifications
PV	Parasitophorous vacuole
PVM	Parasitophorous vacuolar membrane
r.t	Room temperature
rpm	Revolutions per minute
RT-PCR	Reverse transcriptase polymerase chain reaction
s.c	Sequence coverage
SAP	Shrimp alkaline phosphatase
SARS	Severe acute respiratory syndrome
SB3-10	N-decyl-N,N-dimethyl-3-ammonio-1-propanesulfonate
SDS	Sodium dodecyl sulphate

SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
sec	Seconds
STAT1α	Signal transducer and activator of transcription
STY	Serine/threonine/tyrosine
T.cruzi	*Trypanosoma cruzi*
T.g	*Toxoplasma gondii*
T.gondii	*Toxoplasma gondii*
TBE	Borate-EDTA Buffer
TBS	Tris Buffered Saline
TEMED	*N,N,N',N'-tetramethyl ethylenediamidine*
TF	Transcription factor
TFA	Trifluroacetic acid
Th1	T helper 1 cells
TOF	Time of flight
Tris-HCl	Tris Hydrochloride
unPP	unphosphorylated
UV	Ultraviolet
V	Volts
v/v	By volume
V/h	Volts per hour
VhrT	Volt hour total
w/v	Weight per volume
WR	Working reagent
ZnCl2	Zinc chloride
°	Degrees

It is noted that throughout the thesis the abbreviated initial for a genus term(e.g. *Toxoplasma* = *T.*) and the specific name (e.g. *gondii*) should be separated i.e. *T. gondii*.

It is noted that throughout the thesis numbers and units (e.g. 2ml) should be separated i.e. 2 ml.

Presentations

Some of the work in this thesis has been the subject of the following presentations:

Conference Proceedings

2005	British Society of Parasitology 'Manipulation of host cell response during infection with *Toxoplasma gondii*'. *Nottingham, UK.*
2004	Human Proteome Organisation 'Modelling the mammalian cellular proteome during invasion by an intracellular pathogen'. *Beijing, China.* Young Scientist Award winner.
2004	British Society of Parasitology 'Modelling the mammalian cellular proteome during invasion by *Toxoplasma gondii*'. *Chester, UK.* Oral presentation prize winner.

Acknowledgements

It is with great pleasure that I acknowledge the efforts of the many people who have contributed to the development of this thesis. First and foremost I would like to express my gratitude to my two supervisors Dr Jonathan M Wastling and Dr Richard Burchmore. Without their expertise and guidance this work would not have been possible and I thank them for their supervision and support. I would also like to thank all of the members of academic staff in the Department of Infection and Immunity, in particular Prof. Paul Hagan and Prof. Michael Turner for their advice and support.

Thank you to the BBSRC for the financial support during the tenure of my PhD.

For their help and expertise with the proteomic side of my work I would like to thank Dr Andy Pitt and Dr Chris Ward. Special thanks to Dr Andy Jones for his help with the bioinformatics. Thank you to Janice Brock, Anne McIntosh and Sarah McLellan for their help in the laboratory.

I wish to express my gratitude to everyone I have shared both laboratory and office space with throughout my PhD. Thank you to Anne, Adele, Jennifer, Dan, Jane, David, Sophia, Gordon, Manus, Siddiki and everyone in North Lab for their help in the lab, over coffee breaks for the scientific discussion and in the pub for the not so scientific discussions!

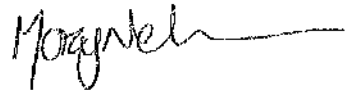
The three years I spent in Glasgow during my PhD contain some wonderful memories, and for that I have mainly to thank Joanne McGregor, who was always there with words of encouragement (and a bottle of wine!) after a hard day, Walt Adamson and Jon Housely. For keeping me sane outside of working hours my thanks must go to all the members of Team Walt over the years, Dr Paul and all of the regular quiz teams.

Many thanks to the Linlithgow girls and the Cambridge gang for their continuing support and friendship and to my new friends in Liverpool. Thank you Mairi for the three (and a bit!) years of PhD banter.

My final thanks must go to my wonderful parents (and best friends) Moira and Graeme and sister Fiona for their encouragement and unquestioning support. I am greatly appreciative to have been given every opportunity in life and I could not have wished for more.

Author's Declaration

The work presented in this thesis was performed solely by the author except where the assistance of others has been acknowledged.



Morag M Nelson, June 2006

Dedication

I dedicate this thesis to Lily & George and Lilian & John

Chapter I Introduction

1.1 Introduction

In nature three types of symbiotic relationship exist; (1) mutualism, where the two associates cannot survive without one another. This association has two-way benefit to the organisms and causes no detriment to either, (2) commensalism, this involves one way benefit, but as in the case of mutualism, no harm to either party, and (3) parasitism, an intimate relationship between two or more species, in which one (the parasite) temporarily or permanently exploits the resources of the host. In principle, a parasite can meet its living requirements exclusively from the connection with a host. The host is however not inactive against the parasite and via different mechanisms aims to hinder parasite development and growth. In a host-parasite relationship, the parasite is metabolically or physiologically dependent on the host. The parasite may have lost the genetic ability to make certain vital metabolic intermediates and has to rely on the host genome for provisions.

Parasitism is widespread throughout nature and it has been estimated that more than 50% of all known species are parasitic at some stage of their life cycle. Almost 100% of plants and animals are themselves parasitized as individuals. The high prevalence of host-parasite relationships provides a biologically relevant system to study host-pathogen interactions.

The interaction between host and parasite is finely balanced; a parasite which stimulates too strong an immune response, or cannot protect itself from the host, will be destroyed. A parasite which multiplies rapidly leading to host death will ultimately destroy itself. The host endeavours to eliminate the pathogen while the parasite employs new ways to maintain the connection with the host. The host can hamper the parasite both mechanically and chemically e.g. cleaning of skin, peristaltic contraction of the digestive tract, antibody development and other immune responses. Parasites respond to these defences by anchoring themselves with hooks and suckers onto skin, or mucous membranes, by developing protective defences which lessen the protective capacities of their host or by hiding from the immune system within host cells.

Pathogens have evolved multiple strategies to avoid host defence mechanisms including host cell invasion, host mimicry and antigenic variation. Host cell invasion by an intracellular pathogen requires a complex series of events. The infectious organism must be able to penetrate the protective barriers of the host and then survive the innate and acquired

immune defences. It will then need to find an environmental niche suitable for growth and/or replication within the host cell. Upon host-pathogen contact, the host innate immune response is activated. Pathogens that survive this undergo assault from the adaptive immune response. Through selective pressure, many successful pathogens have evolved mechanisms to alter and modulate these host immune responses and to evade phagocytosis, thus impeding the host immune system. Underlying this complex host defence and pathogen survival interaction are unique proteomic signatures in both the host and pathogen (Zhang *et al.*, 2005). Invasion involves physiological changes in both the host cells and invading micro-organisms. These changes are relayed by signal transduction systems and are mirrored by gene and protein expression changes in the host and the pathogen. Analysis of the host cell transcriptome and proteome during invasion by an intracellular pathogen will give an insight into the functioning of the cell, and highlight modulation in both gene and protein expression. This will undoubtedly lead to a better understanding of the host-pathogen interaction, identification of the host defence strategies and highlight the mechanisms by which the pathogen modulates the host cell behaviour to its own advantage. Investigating the host response to several pathogens will help deduce whether the host produces a consistent broad response to all intracellular pathogens and to what extent the host response is uniquely tailored to individual pathogens. Comparison of host responses to different strains of the same pathogen could highlight differences in pathogenesis or reveal mechanisms of virulence. Understanding why certain strains of a particular organism are pathogenic and under which conditions the pathogen causes disease may help to develop new tools for disease prevention or management (Kato-Maeda *et al.*, 2001).

There are now several large-scale methods available to investigate the host response to infection, and these techniques involve both genomics (the term used to describe the study of the genome, the entire genetic complement of an organism) and proteomics (the study of the proteome, the entire protein complement of an organism) (Abbott, 1999). It is now possible to study the gene expression of both the host and pathogen at the whole genome level using DNA microarrays and chip technologies (discussed later in this chapter) and the collaborative genome sequencing projects. Large scale gene expression studies give insight into an organism's phenotypic state under any given condition (e.g. age/ disease/ stress).

The transcriptional profile of an organism shows its gene expression at a given time-point. The complement of all the genes expressed at any one time is termed the transcriptome. Within the cell genes are translated into proteins and these encoded proteins carry out the host cell functions. To understand how cells function, one must study what proteins are

present, how they interact with each other and what they do. Many drugs act at the cellular protein level.

Several studies have used microarrays (MA) to investigate the gene expression profile of human cells during invasion by an intracellular parasite. To date, there have been no such investigations into the protein profile. Results from the microarray studies looking at various cell types and pathogens have shown a very wide range of modulation of gene expression in the host cell during invasion. It is not yet clear whether different cell types have a similar response to an invading pathogen, and also whether one cell type will have a general response to invasion, or whether the host response is tailored to each individual pathogen species.

The gene to protein correlation in all organisms remains unclear due to post-translational modifications (PTM), splicing events and translational processes (Ferguson *et al.*, 2005). Therefore it is now recognised that the protein expression must also be investigated alongside the transcriptome, to gain a true overall picture of the infected cell. Such global data, in combination with advanced statistical analyses offers the ability to gain new insights into the interaction between pathogen and host. As defined by Tyers and Mann in the March 2003 issue of Nature Insight: "The study of the proteome, called Proteomics, now evokes not only all the proteins in any given cell, but also the set of all protein isoforms and modifications, the interactions between them, the structural description of proteins and their higher-order complexes, and for that matter almost everything post-genomic" (Tyers and Mann, 2003).

Genomic and proteomic technologies provide a wealth of data and combining these datasets through bioinformatics will yield a comprehensive database of gene and protein function. This will serve as a powerful reference of the behaviour of the cell during invasion and will allow researchers to build and test hypotheses. Studying transcriptomic and proteomic changes associated with the pathogen-host interaction through proteomics may reveal unusual mechanisms for exploiting host cell pathways and diverting host cell organelle functions. These novel modifications could also be potential targets for new drugs.

1.2 *Toxoplasma gondii*

1.2.1 *Life cycle of Toxoplasma gondii*

Toxoplasma gondii (*T.gondii*) is an obligate intracellular protozoan parasite from the phylum Apicomplexa (other members of this phylum include *Plasmodia* sp, the causative agent of malaria, *Neospora* sp. and *Cryptosporidium* sp.). Apicomplexan parasites are so named due to the presence of a specialised structure, the apical complex, which is involved in host cell invasion. The parasite actively invades host cells and can infect and replicate within most nucleated cells. *T.gondii* has evolved a remarkable ability to survive in its host, typically for long periods of time, with minimal pathogenicity. In hosts with differing levels of immune competence, infection with *T.gondii* results in different outcomes of infection ranging from flu-like symptoms in an immune healthy host, to death, in an immunocompromised host (Yap and Sher, 2002). It is one of the most widespread parasites of wild and domestic animals and infection rates in humans are typically 10-30% in most populations but can be much higher (Carruthers, 2002). Approximately one third of the entire world population possess antibodies to *Toxoplasma* although there is considerable variation between and even within countries. The remarkably high infection rate makes *T.gondii* a serious threat to human health, and for the farming industry, it is a major cause of economic loss through neonatal mortality, particularly in sheep (Buxton, 1993).

The life cycle of the parasite is shown in Figure 1-1. The natural life cycle of the parasite is between felines, which serve as the definitive host, and sheep, pigs, rats and other small mammals. Humans can become accidentally infected by ingesting undercooked meat or infective oocysts passed in the cats faeces (Lyons *et al.*, 2002). Upon entering a new host, the sporozoites within the oocyst differentiate to the rapidly growing tachyzoite form of the parasite (Coppens and Joiner, 2001).

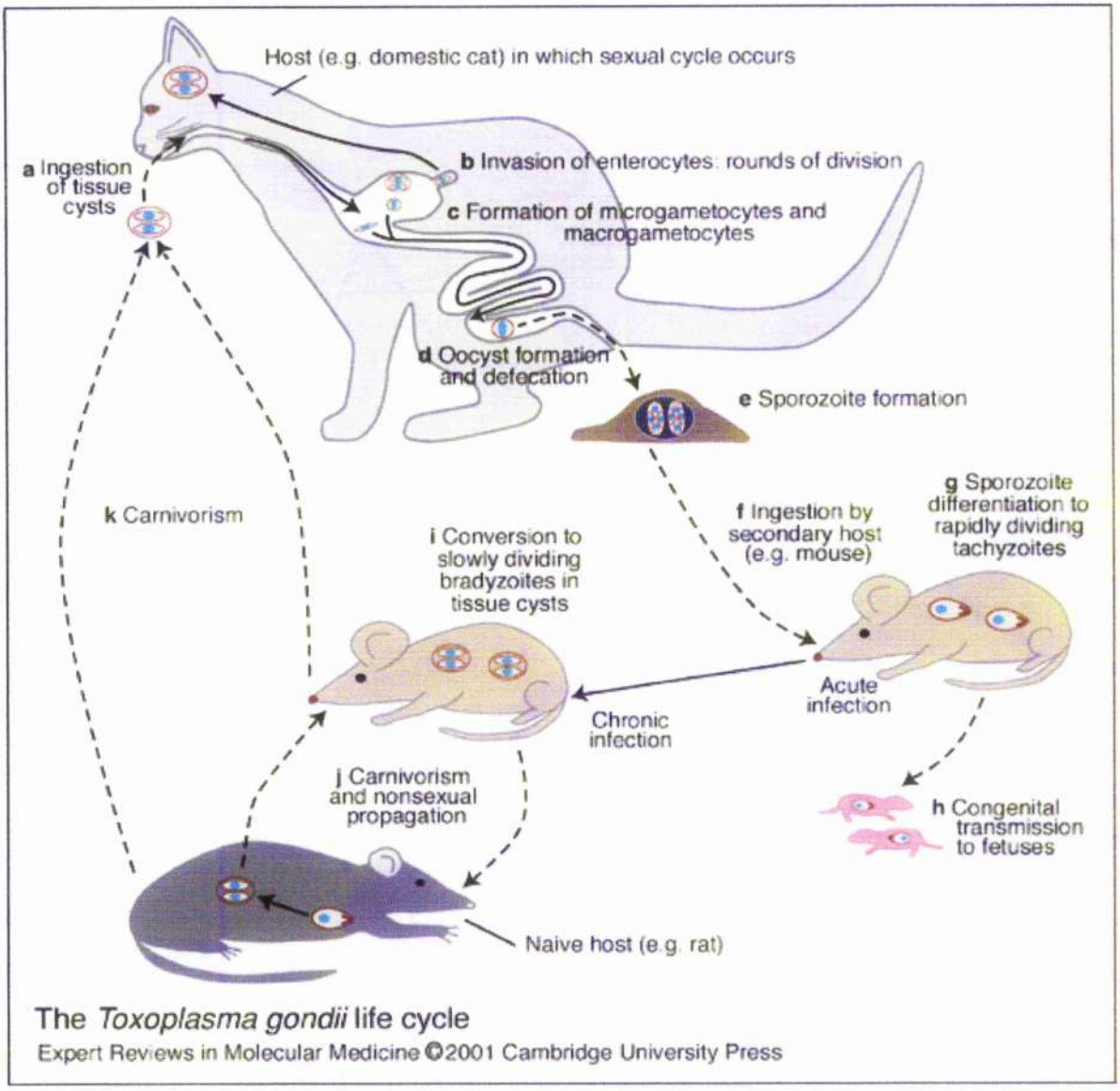


Figure 1-1 The *Toxoplasma gondii* life-cycle (Coppens and Joiner, 2001). The life-cycle of *T.gondii* has a sexual and an asexual stage. The sexual stage takes place in the definitive host (the cat) in the intestinal enterocytes where the parasite undergoes several rounds of division and differentiates into micro- and macrogametocytes. The gametocytes fuse to form an oocyst that is shed into the environment. The oocysts are ingested by an intermediate host and the sporozoites differentiate into the rapidly dividing tachyzoite stage which can infect almost all nucleated cells. Tachyzoites can differentiate into a slow-dividing form, termed the bradyzoite, which can persist as a cyst for the life span of the host. Ingestion of tissue cysts allows for indefinite nonsexual propagation of *T.gondii.*

Initial infection and acute disease are characterised by the presence of the fast-replicating tachyzoites, which take approximately 15-30 seconds to actively invade a host cell by an actin-dependent mechanism (Coppens and Joiner, 2001). Once inside the cell the parasite forms and resides within a parasitophorous vacuole (PV). The parasitophorous vacuolar membrane (PVM) is formed principally from host cell lipids augmented with the assistance of products secreted by the parasite's apical organelles. Pores in the vacuolar membrane allow the parasite access to the host cellular nutrients, including ATP and lipids (Coppens and Joiner, 2001). *Toxoplasma gondii* relies heavily on scavenging essential compounds from the nutrient rich soup of the host cell cytoplasm. The PVM has a dual function; to facilitate access to intracellular nutrients and to protect the parasite from the intracellular host defences. The restricted size of the pores of this membrane prevents access of host proteins to the vacuolar space but allows the diffusion of small molecules. As a bi-directional channel, the pore also prevents the accumulation of parasite waste metabolites within the PV, allowing the parasite to remain in the PV for indefinite periods. *Toxoplasma gondii* prevents the PV fusing with any compartment in the endocytic cascade, or to acidify, thereby preventing fusion, acidification and parasite killing (Coppens and Joiner, 2001).

Tachyzoites replicate intracellularly by endodyogeny (internal division into two) with a generation time of 6-9h *in vitro* until the cell bursts and the parasites infect neighbouring cells, usually after accumulating 64-128 parasites per cell (Coppens and Joiner, 2001). Around 10-14 days post-infection, tachyzoites differentiate into bradyzoites that replicate more slowly and form cysts in tissues throughout the body. It is believed that a mounted host immune response is responsible for the induction of differentiation (Kim and Boothroyd, 2005). Tissue cysts can exist for the life span of the host and are not associated with disease, unless the individual is immunocompromised, whereupon the bradyzoites differentiate back into tachyzoites due to the waning immune response resulting in disease reactivation and uncontrolled tachyzoite proliferation (Dzierszinski *et al.*, 2004; Weiss and Kim, 2000). As well as contributing to morbidity in immunocompromised hosts, the parasite can also cause disease in a second cohort of patients, pregnant women who are infected with the parasite for the first time. The parasite can cross the placental blood barrier and cause congenital defects or abortion.

The detection of *Toxoplasma* specific antibodies is the primary diagnostic method to determine infection with *Toxoplasma*. Sulfonamides and pyrimethamine (Daraprim) are two drugs widely used to treat toxoplasmosis in humans. They act synergistically by blocking the metabolic pathway involving *p*-aminobenzoic acid and the folic-folinic acid

cycle, respectively. The inability of currently, available treatments to eliminate tissue cysts in the chronic stage of infection highlights a continued inadequacy in existing chemotherapeutic agents (Lyons *et al*, 2002).

Several mechanisms exist within the host cell and the immune response that should ensure elimination and destruction of the parasite, including antigen presentation, lysosomal fusion and apoptosis, yet the parasite can maintain a chronic infection eliciting little pathogenicity. How this is achieved is not yet fully understood. It has been previously shown that infection with *T.gondii* causes the induction of numerous immune modulators that effect both the innate and adaptive immune responses. There is an increased need to understand the adaptations that enable *Toxoplasma* to invade and survive in host cells. The adaptive strategies developed by the parasite may range from passive avoidance mechanisms to extremely active modulation of host cell functions (Cummings and Rehman, 2000).

1.2.2 T.gondii *genomics*

Genetic analyses of *T.gondii* have revealed a very stable 87 megabase nuclear genome consisting of eleven chromosomes which show little variation across strains, a 6 kilobase (Kb) mitochondrial genome and a 35Kb plastid-like genome; the apicoplast. Population studies have demonstrated that the small amount of variation between the three clonal lineages defines phenotype such as virulence (Ajioka *et al*., 2001). A collaboration between TIGR (The Institute for Genomic Research, http://www.tigr.org/tdb/e2k1/tga1/) and the Sanger Centre (http://www.sanger.ac.uk/Projects/T_gondii/) has resulted in a complete draft version (10x assembly) of the *Toxoplasma gondii* genome. ToxoDB (http://www.toxodb.org/ToxoDB.shtml) provides access to this draft genome sequence of *T. gondii*.

1.2.3 Immunity to T.gondii

T.gondii induces a strong type-1 T-cell mediated response to promote a self-limiting infection and ensure survival of its host. For example, tachyzoite replication is restricted by interferon gamma (IFNγ) produced by antigen specific T cells. Biochemical studies have indicated that the cytokine interleukin-12 (IL-12), which induces natural killer (NK) cells to produce IFN-γ, is critical for control of both acute and chronic infection and that T lymphocytes (Th1 cells) producing this lymphokine dominate the response to *T.gondii*. Although early control of *T.gondii* infection is clearly mediated by IL-12 dependent IFN-γ

production, the nature of the intracellular mechanism(s) responsible for the restriction of pathogen growth remains poorly understood. The production of nitric oxide (NO) intermediates, a pathway involved in the control of other intracellular protozoa and bacteria, does not appear to be critical during this period as mice deficient in the NO producing enzyme inducible NO synthase survive into the chronic phase (Yap and Sher, 2002).

Chemokines function to recruit immune cells e.g. macrophages, dendritic cells and neutrophils to sites of infection, and their synthesis has been proposed as an important component in cell mediated immunity to *Toxoplasma gondii*. The cellular source of these chemokines is unknown but proposed to originate from both infected cells and other cells recruited to the site of infection.

In a study by Denney *et al*, human HeLa epithelium cell and fibroblast monolayers were infected with *T.gondii* which resulted in the marked increase of expression of IL-8 specific mRNA and secretion of pro-inflammatory and chemoattractant cytokines interleukin-8 (IL-8), GRO1 and MCP-1 (monocyte chemotactic protein 1) (Denney *et al.*, 1999). The authors suggest that pro-inflammatory cytokine secretion is an important host cell response to toxoplasmosis.

T.gondii induced alterations of host cell physiology include the inhibition of apoptosis (Blader *et al.*, 2001). The activity of the key protease caspase-3 was inhibited in infected fibroblasts and parasite induced inhibition of caspase-3-activation required survival-promoting activity of nuclear factor κβ (NFκβ) in murine fibroblasts (Luder and Seeber, 2001). Another parasite induced putative evasion mechanism in host cells is the down regulation of MHC class II expression in macrophages. This was mediated by interference of the parasite with the cellular trafficking of signal transducer and activator of transcription (STAT1α), thus representing a novel strategy to modulate the host-pathogen interaction (Luder and Seeber, 2001).

The mechanisms by which this obligate intracellular parasite manipulates the structures and pathways of the host cell to aid its own survival and to create a hospitable environment presents a difficult challenge to analyse in the background of a nucleated host cell using classical biochemical techniques.

1.3 Host-pathogen model system

Toxoplasma gondii was chosen to investigate the host response to infection with an intracellular pathogen as it is a globally important disease and has also recently emerged as a leading model parasite for other apicomplexans due to the fact it possesses many general apicomplexan features. It is relatively safe (if handled appropriately) and easy to culture *in vitro* and can be readily grown in nearly all cultured cells, is amenable to genetic experimentation and has excellent animal models available (Carruthers, 2002). *Toxoplasma* serves as a model system for genetic exploration of parasite biology and host-parasite interactions. *Toxoplasma* provides a highly amenable system for probing basic mechanisms common to all apicomplexan parasites related to the host cell interaction (Coppens and Joiner, 2001). The molecular genetics of *Toxoplasma* are exceedingly powerful and well developed. Understanding the host-parasite interaction may reveal unusual mechanisms for exploiting host cell pathways and diverting host organelle functions. These novel modifications could also be potential targets for new drugs (Coppens and Joiner, 2001). The host-parasite relationship can be explored using mutant host cells or mutant parasites. In parallel, using microarray technology, it is possible to reveal the nature of the host cell genes that are up or down regulated upon infection.

1.3.1 Human foetal foreskin fibroblasts

The host cells used in the model host-pathogen system detailed in this thesis were Human Foetal Foreskin Fibroblasts (HFF) (European Collection of Cell Cultures). This cell type was selected as it is a well characterised *T.gondii* cell culture model (Blader *et al.*, 2001; de Avalos *et al.*, 2002; Roos *et al.*, 1994), therefore the data from the proteomic study will be comparable to that of previous experiments investigating the changes in gene expression in *T.gondii* infected HFFs. HFFs are non-transformed and do not require pre-treatment with differentiation factors, such as growth factors or phorbol esters, which may increase the variability between cell preparations (Blader *et al*, 2001).

Although HFFs are not considered a major target for *T.gondii in vivo*, they do apparently reflect many features seen in other cell types as well as the infected animal. For example, many of the cytokine and chemokine genes induced in HFFs are the same as those induced in monocytes and *in vivo* during mouse infection (Blader *et al*, 2001).

Human foreskin fibroblasts (HFF cells) offer several advantages for culturing *T.gondii* (1) the large, flat morphology of HFF cells provides an extensive plasma membrane,

permitting multiple cycles of parasite replication within each cell before lysis. (2) HFF cells are strongly contact inhibited, permitting confluent monolayers to be prepared many weeks in advance and maintained until needed. (3) As primary human cells, HFF cells provide a host comparable in certain respects to a clinical infection. Disadvantages of HFFs include their limited replicative lifespan, slow growth rate (especially at high passage number) and differences between HFF cultures available in different labs (Roos *et al.*, 1994).

In vitro studies using cell based host models represent a versatile, inexpensive host-pathogen interaction model. However, there will be differences between this type of analysis and what occurs in an actual biological system (Zhang *et al.*, 2005). Despite this, *in vitro* studies can provide valuable information regarding the host-response.

In 2001, the first analyses of the working draft of the human genome sequence were published, jointly by Sanger and Celera (Stein, 2004). This presents a valuable resource for database searching for gene and protein identification, as one of the final experimental stages in proteomics involves prior knowledge of gene sequence.

1.4 Microarray studies

Until recently, gene or protein studies looked at one molecule at a time using classical biochemical techniques. In the past decade several new techniques have been developed that allow for the study of the entire genome or proteome. Microarrays allows for the global study of the host cell transcriptome at a single time point.

The underlying principle of all microarray experiments is that two different cDNA samples (for example, from infected and uninfected cells) are each labelled with a different fluorophore and hybridise, with high sensitivity and specificity, to complementary sequences immobilised on a solid substrate. The array is scanned at the two wavelengths corresponding to the excitation optimum of the dyes and the fluorescence volume of each spot on each of the two scans of the microarray is measured, relative to background, by analysis software and then compared to the data measurement from the same spot on the opposing scan. Differential transcriptional profiles highlight changes in gene expression and these genes can be identified and then further verified for example by Northern blotting or RT-PCR (reverse transcriptase polymerase chain reaction). Microarrays allows for parallel quantitative measurement of differential gene expression of thousands of genes

at one time. Two major advantages of microarrays are the very small quantity of RNA needed per experiment and the vast amount of data produced per experiment.

Inferring biologically meaningful information from microarray data requires sophisticated data exploration. Most global gene analyses use some form of clustering algorithm to find genes co-regulated across the dataset as shared expression often implies shared function. When a co-regulated class of genes is known, supervised clustering algorithms, which are trained to recognise known members of the class, can assign uncharacterised genes to that class (Cummings and Relman, 2000). Clustering of the data may identify groups of genes involved in activation or repression of key regulatory pathways.

One of the fields in which microarrays have had an increasingly important impact is in studying host-pathogen interactions. Microarrays have been used to look at the similarities and differences in gene expression responses of different host cells to diverse infection stimuli and to investigate fundamental features of innate immunity.

The basic model involves measurement of gene expression of host cells before and after a certain event, i.e. parasite invasion. By following the pattern of gene expression at different time-points, it is possible to elucidate which host genes are up or down regulated over the course of infection and to provide a molecular description of the events that follow infection. Identification of genes that are differentially regulated and the characterisation of their functions provides an insight into understanding pathogen survival.

1.4.1 Transcriptional studies investigating the host cell response to infection

Global transcriptional responses elicited in mammalian cells by pathogenic organisms are predicted to provide a unique signature of that particular interaction (de Avalos, 2002) and provide an invaluable resource for the functional characterisation of host cell pathways required to facilitate pathogen survival and for the further understanding of host defence mechanisms (de Avalos, 2002). The presence of a specific pathogen may also elicit adaptive changes in the host cell phenotype. These changes will be reflected in modulation of gene expression.

The development of an intracellular infection is the result of a series of molecular changes occurring in the cell. To evade the host immune response, intracellular pathogens have developed different strategies to exploit and manipulate host cell functions for infection,

growth and survival. Several groups have employed DNA microarrays to analyse the host response to infection and Table 1-1 lists some of these studies which used a range of parasites, bacteria and viruses to investigate the gene expression of host cells when invaded by an intracellular pathogen.

Intracellular pathogen	Author of paper
Toxoplasma gondii	Blader *et al.*, 2001
Trypanosoma cruzi	de Avalos, 2002
Bordetella pertussis	Belcher, 2000
Listeria monocytogenesis	Cohen *et al.*, 2000
Salmonella dublin	Eckmann *et al*, 2000
HIV	Cummings and Relman, 2000
Human cytomegalovirus	Zhu, 1998

Table 1-1 Studies investigating the gene expression of host cells during infection with an intracellular pathogen

Genes modulated during infection can be examined to refine hypotheses about their role in the host-pathogen interaction. This gene response details the degree of the host response to infection (Kato-Maeda *et al*, 2001). Comparing the gene expression profiles of host cell infection by wild type and mutant pathogen strain has also been used to explore the host-pathogen interaction (Kato-Maeda *et al*, 2001).

Zhu *et al* (1998) used microarrays to investigate the response of human fibroblasts to cytomegalovirus (CMV) infection (Zhu *et al.*, 1998). Four percent of the genes arrayed changed in expression by a factor of four or more after the onset of viral replication, as compared to uninfected HFF cells. In this study, modulation of mRNAs encoding components of prostaglandin E2 biosynthetic pathway suggested that CMV induced synthesis of this pro-inflammatory secondary messenger. Three potential explanations for this observation were proposed; (1) this pathway could be induced by a cellular response intended to limit spread of infection by promoting the killing of infected cells, (2) viral regulators could induce prostaglandin E2 production to lure monocytes, which could subsequently be infected, leading to viral dissemination within the host, (3) and these genes could be induced secondarily through induction of interleukin-1β since a similar pattern of regulation was observed in cells treated with that cytokine. Some of the changes, such as induction of cytokine stress-inducible proteins, and many interferon-inducible genes, were consistent with induction of cellular immune responses (Cummings and Relman, 2000). These studies show the potential power of using microarrays to characterise the host response but also suggest that interpretation of host gene expression profiles will be challenging as it is often difficult to determine whether the response is

stimulated by the pathogen to aid its own survival, or stimulated by the host as a protective response. Microarrays can identify interesting cellular events, but because expression patterns cannot distinguish between these mechanisms, the need for further investigation is necessary (Cummings and Relman, 2000).

A similar experimental design to that of Zhu *et al* (1998) examined the global effects of HIV-1 infection on cultured CD4-positive T cells. One study concluded that HIV-1 infection resulted in differential expression of 20 of the 1506 human genes monitored and that most of these changes occurred only after 3 days in culture. In contrast, the preliminary results of an independent study using a similar design indicated that substantial HIV-induced transcription changes began very early after inoculation. The latter study confirmed activation of nuclear factor –κβ (NF-κβ) , p68 kinase and RNase L (Cummings and Relman, 2000). NF-κβ is a transcription factor that is well known as a critical mediator of the common immune response in various cell types because it is activated in response to many micro-organisms and other stresses (Diehn and Relman, 2001). These contradicting results highlight the difficulties often seen when comparing microarray datasets as often different MAs or techniques have been used to investigate the same question.

DNA expression arrays have also been used to examine the response of host cells to infection by bacterial pathogens. The results from the various studies have shown both similarities and differences between the transcriptional profiles of host cells to different pathogens. Transcriptional profiling of macrophages and epithelial cells infected by *Salmonella* confirmed increased expression of many pro-inflammatory cytokines and chemokines, signalling molecules and transcription activators and identified genes previously unrecognised to be regulated by infection (Cummings and Relman, 2000).

One of these studies was by Eckmann *et al* who investigated the host cell response of human colorectal and colonic epithelial cells after infection with *Salmonella* (Eckmann *et al.*, 2000). They found from the gene expression profile that relatively few genes were induced upon infection. The authors interpret this as evidence that the epithelial mRNA response to infection is limited but specific. Those that were induced include cytokines, kinases, transcription factors and HLA class 1 genes. Over expression of Iκβ (an inhibitor of NF-κβ) blocked induction of gene expression for a number of regulated genes, underscoring the importance of NF-κβ in the pro-inflammatory response (Diehn and Relman, 2001).

Similar results were seen upon infection of human bronchial epithelial cells with *Bordetella pertussis* (Belcher *et al.*, 2000). The small number of genes that were induced (including monocyte chemotactic protein (MCP-1), tumour-necrosis-factor-∝-inducible DNA-binding protein A20 (TNFAIP3/A20) and the nuclear factor inhibitor ∝ (NFκβI∝)) included genes encoding pro-inflammatory cytokines with chemoattractant activities, which may explain the cellular infiltrate seen in patients with pertussis. In addition, the transcriptional analysis suggested that the physical properties of the respiratory tract mucus of the host might be altered by the activity of the pertussis toxin, creating a more advantageous microenvironment for the pathogen.

While each microarray study stands as a valuable dataset on its own, comparing the combination of data from different studies can be used to assess the similarity of genome-wide host gene-expression responses induced by different pathogens under different circumstances. It will be important to identify the central regulators of common immune responses and the ways in which pathogens can avoid or stimulate this process (Diehn and Relman, 2001).

1.4.2 Differential gene expression in Toxoplasma gondii *infected cells*

Microarray studies allow for investigation into the differential gene expression in *Toxoplasma* infected host cells. This parasite provides a highly amenable system for probing basic mechanisms common to all apicomplexan parasites related to their interaction with their specific host cells (Coppens and Joiner, 2001). Two published studies to date have used microarray technology to investigate changes in host cell gene expression upon infection with *T.gondii*; Blader *et al.*, 2001 and Gail *et al.*, 2001.

The investigation by Blader *et al* is titled 'Microarray analysis reveals previously unknown changes in *Toxoplasma gondii*-infected human cells' (Blader *et al.*, 2001). This study investigated the gene expression profile of Human Foetal Foreskin Fibroblasts (HFF) infected with *T. gondii* using human cDNA microarrays consisting of about 22,000 known genes and uncharacterised expressed sequence tags (ESTs). RNA was purified from HFFs that were mock or parasite (*T.gondii* tachyzoites) infected for a series of time points and a microarray experiment carried out to compare cDNA from mock versus infected HFFs. Following hybridisation, the data were filtered based on spot intensities three times greater than background and displayed a two-fold change or greater (relative to uninfected) in at least one time point during the experiment. The filtered data were organised into temporal gene expression clusters. Clustering of the time course data indicated that two distinct gene

expression patterns were present; genes modulated rapidly after infection (1--2 hpi) and those modulated later (>6 hpi).

During the first 2 hours of infection, 63 genes exhibited a 2-fold or greater increase in their abundance and 15 genes exhibited a 2-fold decrease. Many of the induced genes were previously known to be involved/associated with the immune response and included chemokines (GRO1, GRO2, LIF and MCP1), cytokines (IL1β, IL-6), cell matrix and adhesion proteins (ICAM-1 and matrix metalloproteinase) and apoptotic and transcriptional regulatory factors (superoxide dismutase 2). This data was confirmed by Northern Blot. Although the early induced genes were predominantly associated with the immune response, they represented a relatively small number of immune response genes spotted on the microarrays.

Twenty four hours after infection, 567 genes were modulated in expression. Categorising the genes by function indicated modulation in numerous cellular processes including carbohydrate and lipid metabolism, transcription regulation, protein synthesis, targeting and degradation, cell signalling, inflammation, cell adhesion and cytoskeleton, nucleotide and amino acid metabolism, cell cycle and apoptosis. It was however noted that this type of functional clustering may be misleading, because some genes may be involved in more than one process but were arbitrarily grouped into a single class.

To investigate the role of soluble factors in modulating the transcriptional response, media from 24 hpi mock and parasite infected HFFs were added to uninfected HFF monolayers. RNA was harvested at 4,14, and 24 hours from each conditioned media-treated monolayer as well as from the 24 hpi mock and parasite infected cells from which the conditioned media was obtained. Each microarray was hybridised with labelled cDNA prepared from each time point, analysed and clustered. Gene clustering indicated that 2 classes of gene were modulated during infection; (1) those regulated by both conditioned media and *T.gondii* and (2) those regulated only in the presence of *T.gondii* (therefore the transcriptional regulation of these genes is a direct consequence of intracellular infection and growth.) The mevalonate metabolic genes shown to be up regulated 24 hpi were confirmed to be regulated only in the presence of *T.gondii*, as were many of the glycolytic enzymes.

The host genes modulated during infection could be grouped into one of three major classes, (1) pro-host genes that protect the host from infection (e.g. IFN-γ and IL-12), (2) pro-parasite host genes which are necessary for parasite growth and may include those involved in inhibiting infected cells from undergoing apoptosis or those involved in metabolic

pathways whose products the parasite scavenges from the host cell to satisfy its auxotrophs and (3) bystander host genes which are incidentally regulated as a result of the first two classes. Thus, microarrays are instrumental in identifying pro-parasite genes and analysis of time course data sets has identified several candidate pro-parasite genes, including transcripts that encode anti-apoptotic proteins and glycolytic and mevalonate enzymes (Spear *et al.*, 2006).

To investigate whether the pro-inflammatory genes induced early during infection in HFFs infected with *T.gondii* could be a non-specific response of HFFs to intracellular parasitism (i.e. independent of the pathogen), the HFF response to a second protozoan intracellular parasite, *Trypanosoma cruzi*, was compared to the *T.gondii* response using replicate microarrays and the same data analysis methods as for the *T.gondii* experiments. Only 4% (23 of 614) of the genes modulated by *T.gondii* 24hpi were also modulated by *T.cruzi* 24hpi. Of these 23, only 3 (MCP1, pre-B cell-colony-enhancing-factor and bradykinin receptor B2) were up regulated 2h post invasion (hpi) with *T.gondii*. The overall results of this study indicated that the transcriptional response is unique to individual intracellular protozoan parasites. A more in-depth study investigating transcriptional changes in gene expression in host cells after invasion by *T.cruzi* (de Avalos, 2002) is discussed later in this chapter.

The study by Gail *et al*, titled 'Transcriptional profile of *Toxoplasma gondii*-infected human fibroblasts as revealed by gene-array hybridization', was similar in design to that of Blader *et al* (2001) although on a smaller scale (Gail *et al.*, 2001). Transcriptional profiles of infected and uninfected HFFs were determined by hybridisation to gene arrays representing about 600 genes. The results of this study revealed up regulation of 17 genes (>5 fold) after *T.gondii* infection, including genes encoding immunomodulatory factors; IL-6, macrophage inflammatory protein 2α (MIP2α) and neutrophil activating protein ENA 78, kinases, cell surface antigens (including transferrin receptor TfR), proteins involved in DNA synthesis or repair, cytokines, proteins involved in apoptosis and chemokine factors. The results were validated using RT-PCR.

Gail *et al* compared the results from the *T.gondii* infected HFF expression profiles with the transcriptional profiles of HFFs infected with either *Salmonella typhimurium* and *Chlamydia trachomatis*, to identify genes specifically induced in *T.gondii* infected cells. Most of the up regulated genes were induced by all three pathogens including several immunomodulatory factors; IL-6, MIP2α and neutrophil activating protein ENA 78. The genes for the transferrin receptor and myristoylated alanine-rich C-kinase substrate (MacMARCKS) were

up-regulated in *T.gondii* infected fibroblasts only and these may be involved in a specific host-cell interaction. For example, parasite induced TfR up regulation might result in increased iron uptake by the host cell and as a consequence the parasite might benefit from a better intracellular iron supply. It is possible that the differences in the pathogen-induced RNA expression by the host cell are related to the different lifestyles of the three pathogens i.e. *T.gondii* actively invades and *Salmonella* and *Chlamydia* are phagocytosed.

Comparing the MA studies by Blader and Gail, infection of HFFs by *T.gondii* resulted in increased gene expression of immunomodulatory factors including IL-6, MCP-1, TNFs and also cell adhesion genes. Many other genes showed differential gene expression but were not present on the arrays in both of the studies or the groups results differed from each other.

The results of these studies can be compared with a study by de Avalos *et al* (2002), titled 'Immediate/Early response to *T.cruzi* infection involves minimal modulation of host cell transcription', that investigated differential gene expression in HFFs after invasion by the intracellular parasite *Trypanosoma cruzi* to investigate whether the pro-inflammatory genes induced early during infection in HFFs infected with *T.gondii* were a specific or non-specific response of HFFs to intracellular parasitism.

Signalling pathways in both parasite and host cell are known to regulate *T.cruzi* entry into mammalian cells and some of these early signals are thought to regulate downstream host cell transcriptional events. To facilitate comparisons to existing microarray data concerning *Toxoplasma gondii* and human cytomegalovirus, primary HFFs were used in this study. The transcriptional response of *T.cruzi* infected HFFs at several time points post invasion was investigated. The study found no significant changes of HFF gene expression to *T.cruzi* at 2, 4 or 6 hpi. A significant increase in transcript abundance for 106 host cell genes was observed at 24hpi. Among the most highly induced is a set of interferon (IFN) stimulated genes (ISG), indicative of a type-1 IFN response to *T.cruzi*. No up regulation of these genes was observed at the earlier time points following infection, suggesting this response may follow IFN production by infected HFFs. These ISGs are likely to function as an important modulator of innate host defence mechanisms.

The increased gene expression seen in the *T.cruzi* infected HFFs at 24hpi showed several substantial differences from those evoked in *T.gondii*. Classification of the *T.cruzi* induced genes into functional categories revealed the conspicuous absence of several major groups of genes (cytokines, carbohydrate metabolism, cell cycle and apoptosis) all of which were

up regulated in HFFs by *T.gondii* on similar cDNA arrays. Also, the most highly represented group of genes induced by *T.cruzi*, the ISGs, were not induced by *T.gondii*, bacterial or fungal pathogens. Significant changes in host cell pro-inflammatory cytokine gene expression was not seen from this microarray experiment. The authors noted this to be an unusual response, as up regulation of these genes was commonly seen in other experiments investigating the host response to intracellular parasitism.

The marked delay in *T.cruzi* induced host cell transcriptional responses indicates that changes in host cell gene expression may correlate with a particular parasite-dependent event, such as differentiation or replication. Establishment of infection by *T.cruzi* involves several events that together take about 24 hours. Using a similar analogy, the rapid and robust gene expression response initiated in HFFs to *T.gondii* infection may correlate with the fact that parasite differentiation is not required for replication and there appears to be no lag phase following host cell entry associated with this parasite.

1.4.3 Host-pathogen gene expression conclusion

Microarray studies can provide a "genome wide view" of the host cell gene expression in a particular situation by use of a whole genome array. A number of studies have been carried out looking at host cell gene expression in response to a variety of intracellular pathogens. Intracellular pathogens can induce a variety of host cell modulations to aid their survival, including subverting the host immune response, altering host cell pathways and organelle redistribution. One similarity across all of the studies discussed here was some modulation of genes involved in the host immune response. Aside from this, it is very difficult to identify any common responses across the microarray studies as different arrays, types of host cells or methods of data analysis were used. Therefore this means that no overall conclusions or hypotheses can be generated from the microarray studies about the host cell response to infection with an intracellular pathogen.

When comparing datasets such as those above, it is important to recognise the possibility that gene expression responses that appear to be induced by a pathogen stimulus may actually be due to the conditions under which the experiments were performed. Because the results of pathogen gene expression and of host-pathogen interactions are influenced by the model system used, such results must be interpreted with caution (Kato-Maeda *et al*, 2001). Analyses such as these are quite sensitive to the characteristics of the experimental design and these characteristics can be divided into three groups; (1) DNA microarray

related experimental issues, (2) infection-related experimental issues and (3) biological host response related issues (Diehn and Relman, 2001).

Each study provided a valuable dataset of host gene expression in response to a specific pathogen. As proteins are the functional molecules within the cell and the gene to protein correlation remains unknown it is now necessary to look at the protein expression in host cells, and this proteomic data, in combination with the microarray gene expression study data, will lead to an overall picture of the functional molecules in the cell and specifically the genes and proteins whose expression is modulated by the presence of an intracellular pathogen.

1.5 Proteomics

Proteomics can be defined as the large scale study of proteins and their properties such as post-translational modifications, expression levels and interaction with other molecules to obtain a global view of cellular processes at the protein level (Abbott, 1999).

Although the global analysis of mRNA is extremely powerful, it has limitations. The 'one gene one protein' hypothesis is not valid because of RNA editing, splicing and post-translational modifications of proteins (Graveley, 2001). Gene expression can be controlled by various mechanisms in the pathway from DNA to RNA to protein. (1) Transcriptional control determines when and how often a gene is transcribed, (2) RNA processing control, how the RNA transcript is spliced, (3) RNA transport control, which completed mRNAs will be exported from the nucleus to the cytoplasm, (4) translational control, determining which mRNAs in the cytoplasm will be translated by ribosomes, (5) mRNA degradation control, selectively destabilising certain mRNA molecules in the cytoplasm and (6) protein activity control, selectively activating, inactivating or compartmentalising proteins (Alberts *et al.*, 1999). The study of gene expression using microarrays allows for whole cell profiling of the transcriptional response to a certain stimulus. However, due to the numerous levels of possible gene expression control, the transcriptional profile does not always provide an accurate representation of the cellular protein expression.

It has been estimated that each protein encoded by the human genome can exist in 5 or more isoforms due to PTM and each of these protein isoforms may have different biological activities in cellular metabolism (Cash, 2002). Post-translational modification of proteins is important for biological processes, i.e. in cellular signalling where the attachment of a phosphate group to a protein can trigger either activation or inactivation of a signalling cascade. Another reason for the poor level of correlation between mRNA and protein levels (generally lower than 0.5) is that the rates of degradation of individual mRNAs and proteins differ. Even in the simplest self-replicating organism, *Mycoplasma genitalium*, there are 24% more proteins than genes, and in humans there could be at least three times more (Abbott, 1999). Thus there is an increase in the complexity of human proteome without an increase in the actual genome size.

It should therefore not be expected that there will be a strict linear relationship between mRNA levels and the proteome of a cell. Protein analysis mirrors the real protein profile displaying key regulatory events in cellular mechanism. As most drug targets are proteins, methods to analyse the protein complement of cells efficiently should contribute directly to

drug development. Therefore direct studies at the level of the functional proteins themselves are required to support mRNA analyses.

Two central technologies in proteomics are two dimensional gel electrophoresis (2-DE) and Mass Spectrometry (MS). 2-DE is used to separate complex protein mixtures into individual proteins, MS can then be used to identify these proteins. Figure 1-2 is a schematic of a typical gel-based proteomics workflow.

Proteomics plays an important role in the pharmaceutical industry. The new methods for the global analysis of gene expression, which can analyse large numbers of parameters simultaneously, are ideal for accurately establishing drug action, as well as identify possible toxic side-effects at the molecular level. New techniques offer opportunities for screening gene expression at the mRNA level. However, as discussed above, mRNA analysis alone fails to provide a comprehensive view of gene expression. Consequently, proteomics is becoming an important tool in the areas of drug development and toxicology.

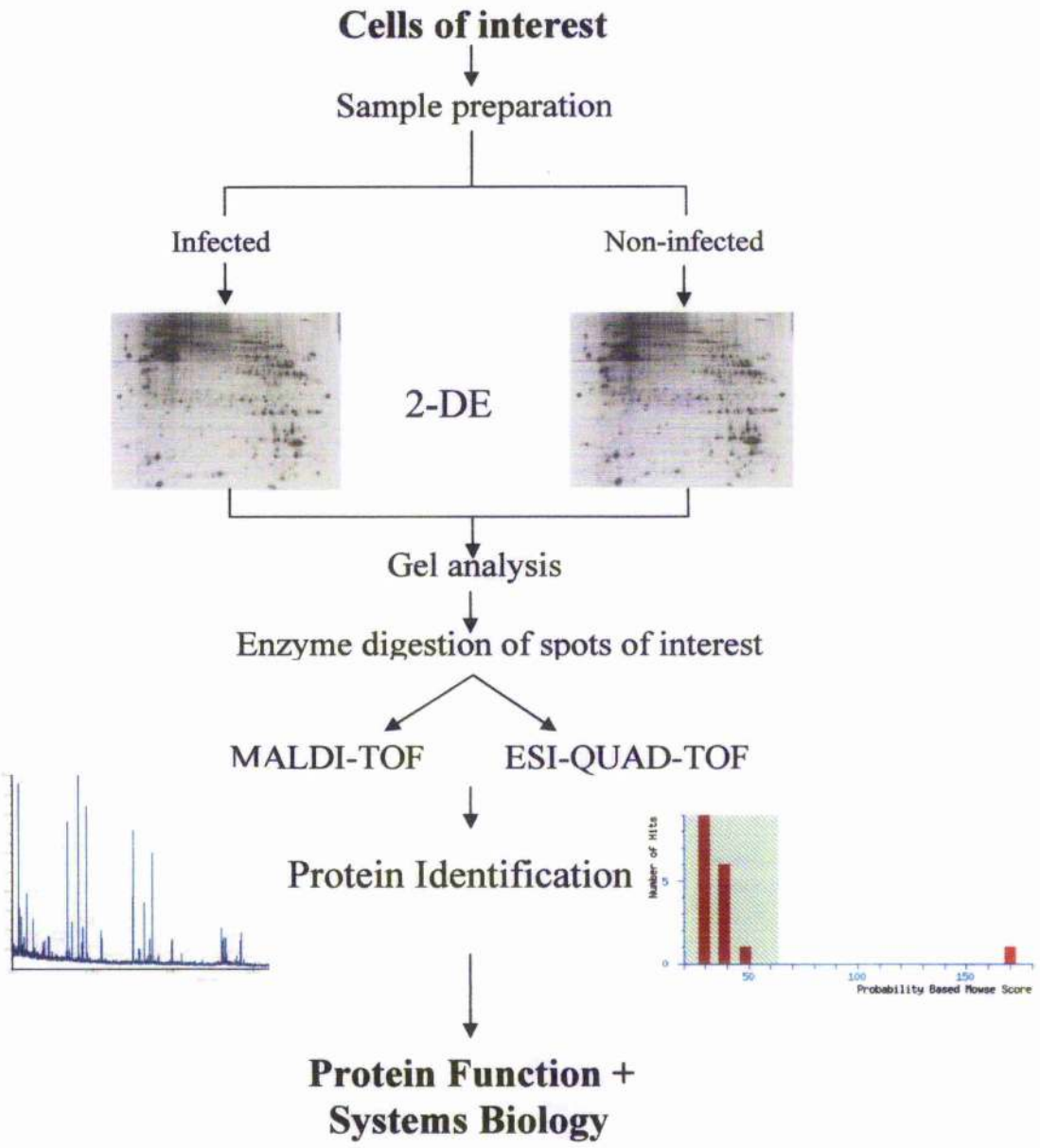


Figure 1-2 Schematic of a typical gel based proteomics workflow. Proteins are extracted from the sample of interest and separated by 2-DE. The proteins on the gel are visualised by staining and the gel images compared. Spots which are differentially expressed between samples are excised from the gels and the intact proteins digested into their component peptides using the enzyme trypsin. The peptides are then analysed using MS and bioinformatics resulting in protein identification. All data from the study is collated and used to investigate the overall proteome expression of the sample of interest.

1.5.1 2-DE

Two-dimensional electrophoresis (2-DE) was first introduced in 1975 by O'Farrell (O'Farrell, 1975) and generates image based 2-DE reference maps representing a typical protein expression pattern under specific conditions. 2-DE involves separating proteins firstly by charge (isoelectric focusing (IEF)) and secondly by molecular mass, using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). 2-DE has the capacity to separate and profile thousands of proteins (Wu *et al.*, 2002). The sample (e.g. cells, tissue extract) is solubilised (either chemically or mechanically) and the proteins denatured into their polypeptide subunits. The solubilisation method must be compatible with IEF and should not introduce modifications to proteins that could affect MS (Chambers *et al.*, 2000). The extract can then be incubated with nucleases to remove DNA or RNA molecules which can interfere with electrophoresis. This mixture is separated by isoelectric focusing (IEF) based on the proteins isoelectric point (pI). Proteins are amphoteric molecules, having positive, negative or zero charge depending on the pH of their surroundings. The pI is the pH at which there is no net electric charge on a protein.

When 2-DE was first introduced proteins were separated in the first dimension using polyacrylamide gels containing a mixture of hundreds of carrier ampholytes (CA) with differing pIs which span a specific pH range providing a pH gradient. When a voltage is applied to the gel, the CA separate according to their pI and produce a continuous pH gradient. The disadvantage of using CA was that the pH gradients produced were unstable and proteins could diffuse across the gradient. Most importantly IEF using CA was difficult to reproduce accurately.

In recent years the introduction of immobilised pH gradients (IPG) to replace the CA system has increased the resolution of separation and reproducibility of 2-DE (Gorg *et al.*, 2000). A pH gradient is generated by co-polymerisation of 6-8 well-defined chemicals (the immobilines) with the acrylamide matrix. An immobilised pH gradient is produced by covalently incorporating a gradient of basic and acidic buffering groups (of desired pH range between 3 and 11) into a polyacrylamide gel at the time it is cast. IPG strips allow for increased loading capacity and reproducibility compared to the CA system. IPG strips are plastic backed and the pH gradient is fixed which provides mechanical strength and does not allow for pH gradient drift. An electric field is applied to the gradient and the charged polypeptide subunits migrate in the polyacrylamide gel strip that contains an immobilised pH gradient (IPG) to the position equivalent to the pI. The technique has good resolving power as the pI of a protein is based on its amino acid sequence. The resolution

can be adjusted further by changing the range of the pH gradient (e.g. to look specifically at basic proteins (pH 7-11) or acidic proteins (pH 3-4)). IPGs result in highly reproducible protein patterns and are capable of much higher protein load (up to 5mg) than CA-IEF. High protein loads are essential if less abundant proteins components are to be identified using mass spectrometry (Molloy, 2000). Disadvantages of IPGs include the poor representation of hydrophobic and membrane proteins in 2-DE gels. There is also an exclusion of larger proteins when loaded onto the gel, this is thought to be made worse by in-gel rehydration.

The second step of 2-DE is to separate proteins based on molecular weight using SDS-PAGE. The IPG strip is applied to the edge of a polyacrylamide gel and the focused polypeptides migrate in an electric current into the second dimension gel and are separated on the basis of molecular size. Individual proteins are then visualised by staining the gel e.g. with Coomassie blue or silver staining (this will be discussed in further detail in chapter II). Several thousand spots may be resolved on a single 2-DE gel and the initial choice of sample solubilisation conditions and pH range of the gel strip used for the first dimension will determine which particular subset of proteins from the proteome are profiled.

The main problem in the visualisation of total cell or tissue extract proteins is the highly dynamic range and chemical diversity of proteins over a wide range of molecular weights, pI and solubility and although 2-DE is a very successful technique to profile thousands of proteins from one sample, there are several limitations to the technique. These include; (1) difficulties in gel reproducibility, (2) limited dynamic range i.e. 2-DE often fails to detect proteins with molecular masses in excess of 100 thousand Daltons and those with pI values lower than 4 or higher than 9 (Yanagida, 2002), (3) certain proteins stain poorly or not all, (4) hydrophobic proteins are under-represented, (5) post-translational modifications can result in the same proteins appearing at multiple positions on a gel and (6) low-abundance proteins which tend to do the most interesting jobs in cells are drowned out by high-abundance 'house-keeping' proteins which can be present at 10,000 times greater concentration (Abbott, 1999).

Gel images are analysed with specialised software which allows comparisons of multiple gels, therefore it is possible to compare gel images to identify changes in protein expression between samples e.g. healthy versus diseased. The proteins of interest can then be excised from the gel, enzymatically digested into peptides using trypsin, and identified using MS.

1.5.2 Mass Spectrometry

Protein identification by MS is a high-throughput method which provides extremely sensitive measurements of the mass of molecules and this data can be used to search protein and nucleotide databases directly to identify a protein. Mass spectrometers consist of three essential parts; (1) an ionisation source that converts molecules into gas phase ions, (2) a mass analyser which resolves ions based on their mass/charge (m/z) ratio and (3) an ion detector which detects the ions resolved by the mass analyser. A mass analyser uses a physical property such as time-of-flight (TOF) to separate ions of a particular m/z value that then strike the detector. The magnitude of the current that is produced at the detector as a function of time is used to determine the m/z value of the ion. The peptide or peptide fragment mass can then be calculated from this value depending on the charge state. The two most common types of MS are matrix-assisted laser desorption ionisation-time of flight (MALDI) which is useful for identifying proteins from organisms with sequenced genomes and electrospray ionisation tandem MS (MS/MS), which can provide some actual sequence data.

1.5.2.1 MALDI

MALDI is an MS method which is used to determine the mass of peptides (peptide mass fingerprinting (PMF)). The molecules to be ionised are desiccated in an energy absorbing crystalline matrix and a laser causes excitation of the matrix and the ejection of ions into the gas phase. This method of ionisation is often used in conjunction with TOF detection to identify proteins by peptide mass fingerprinting. The mass of the peptide is determined and searched against databases.

MALDI-TOF is a fast and reliable method of protein identification which relies on comparing peptide mass fingerprints from analysis of enzymatically digested protein samples to computer generated theoretical fingerprint profiles from protein/EST databases. Protein matches are ranked according to algorithms that take into account protein properties such as species of origin and expected mass range, in addition to the percentage sequence coverage and mass accuracy, to calculate the probability (and confidence level) of a correct identification.

1.5.2.2 Tandem MS (MS/MS)

Tandem MS can perform two-stage mass analysis of ions (Wu *et al.*, 2002). ESI-MS/MS can be coupled to several types of mass analyser including, TOF, ion trap and quadrupole. The samples are introduced to the mass spectrometer in liquid phase. Each peptide analysed by MS is subjected to further fragmentation to give partial information about the peptide sequence. Peptide fragmentation is induced in a collision cell and the molecular masses of the resultant fragment ions are compared with the theoretical masses of fragments from each protein sequence in the databases. This method is an accurate way to identify proteins and is applicable to short sequence data such as those of expression sequence tags (ESTs) (Yanagida, 2002). MS/MS can provide actual sequence data and can identify large proteins.

A typical MS approach is to analyse all samples initially by MALDI-TOF MS because of the speed of the technique. Proteins that cannot be identified by this method, perhaps because they are present in low amounts because several proteins are present in the one spot or because the protein is not present in any of the public domain databases, can then be subjected to further analysis by tandem MS. Therefore the two types of MS, MALDI and MS/MS, are complementary techniques.

The identification of proteins using mass spectrometry data is simplified by the use of specialised software algorithms that compare actual peptide mass fingerprinting data, or peptide fragmentation data with theoretical patterns generated from databases. Protein identifications may be obtained by searching protein/DNA databases.

Comparative non-gel based approaches which use MS/MS include MUDPIT (multidimensional protein identification technology) and isotope coded affinity tags (ICAT) (discussed later in this chapter). MUDPIT is a separative proteomic technique analogous to 2-DE which couples 2D liquid chromatography to tandem MS to resolve and identify peptides from complex mixtures (Wu *et al.*, 2002). MUDPIT is not quantitative. ICAT is a quantitative approach and is analogous to DIGE (difference gel electrophoresis).

1.5.2.3 Post-translational modifications

Post-translational modifications (PTM) play a critical role in cellular function therefore identification is an important goal of proteomics. Proteins can undergo a wide range of post-translational modifications such as phosphorylation, glycosylation, sulphonation,

palmitoylation and ADP ribosylation. Such modifications are vital for the correct functioning of many proteins. Although post-translational modifications are determined to some extent by the proteins amino acids, it is not currently possible to predict reliably which, if any, modifications occur for any particular protein (Cash, 2002). MALDI-TOF can identify when a protein is phosphorylated (Palzkill, 2002) as the expected fragment sizes with and without a phosphate can be calculated. More commonly tandem MS is used to investigate PTMs with the help of specific database searching against modified theoretical databases.

1.5.2.4 Quantitative proteomics

Quantitative proteomics is necessary to determine actual volume differences between proteins of interest in comparative proteomics. Several quantitative proteomic techniques exist. (1) Radioactive labelling. The proteins are labelled either *in vivo* during cell growth or *in vitro* within the protein lysate with radioactive amino acids e.g. cells are grown in the presence of [^{35}S] methionine and the protein extract separated by 2-DE. The spots are excised from the gel and the amount of radioactivity per spot determined using a scintillation counter which allows quantitation of the amount of protein in a spot. The exact amount of protein in each spot is calculated from this number by comparison to protein standards of known concentration. (2) SILAC (stable isotope coded amino acids in cell culture) allows for in vivo incorporation of a label into proteins for MS based quantitative proteomics. (3) Isotope-coded affinity tags (ICAT). In this approach, the two samples of interest are separated by 2-DE and protein spots of interest excised, digested and peptides labelled at the N-terminus with H4NiCNHS (light isotope containing hydrogen) and D4NiCNHS (heavy isotope containing deuterium) respectively. MS analysis of the combined digests then provides information on the expression levels of those proteins by comparison of H4/D4 ratios of individual peptides. (4) Difference Gel Electrophoresis (DIGE). Two samples of interest are labelled with fluorophore CyDyes, mixed together and run on the same gel. The ratio of fluorescent signal intensity of matching spots between samples is calculated along with accurate spot volumes (Unlu *et al.*, 1997). The spots can then be excised from the gel and the proteins identified using MS.

Jiang *et al* (2005) compared DIGE and ICAT and found that ICAT analysis showed a clear bias for proteins with high molecular mass, whereas DIGE could separate proteins in certain low molecular mass ranges and also identified cysteine-free proteins that were transparent to the ICAT analysis. Moreover ICAT analysis quantifies the sum of the protein species of one gene product, whereas DIGE quantifies at the level of resolved

protein species, including post-translationally modified and processed polypeptides (Jiang *et al.*, 2005). Therefore DIGE is a better technique to look at PTMs.

Proteomics can be used to investigate the cellular protein expression on a proteome wide scale. An essential step in discovering protein networks and their alterations in disease is to generate image based 2-DE reference maps representing a typical protein pattern of cells and tissues under normal conditions. These cellular models can then be compared to either different cell populations or the same cell population under different conditions (e.g. heat/stress/disease/infection). Many of these cellular models have been created using proteomics and include maps of human hippocampus, blood platelets and liver cells (Lehr *et al.*, 2002). These maps can be used to understand regulation of protein networks by extracellular stimuli as well as drugs and their alterations in disease. A study by Lehr *et al* (2002) looked at the use of primary skin fibroblasts as a human model system for proteome analysis. The resulting reference images of human primary fibroblasts provided a basis for investigating regulation by extracellular stimuli and drugs as well as their alterations in patients with different diseases. Primary human fibroblasts derived from skin biopsies represent the only cell system, which can be cultured as a monolayer over a long period of time without the need to immortalise the cells. The results proved that they could generate reference images of protein expression in human primary fibroblasts, which can then be used to identify possible alterations in protein profiles of cells obtained from patients with different diseases.

1.6 Bioinformatics

As well as the sophisticated software used to identify proteins from MS data, bioinformatics can be used to pull together the information gained from the genomic and proteomic studies. Specific bioinformatic objectives that will allow for this are (1) development of annotation standards and indexing techniques to enable the integration of microarray and proteomic data sets, (2) development of indexing and data linking techniques required for combined statistical data analysis of proteomic and microarray data, and (3) to examine statistical tools for an integrated study of the changes in the transcriptome and proteome of a cell.

Microarray experiments are, in the most part, annotated according to guidelines set out by MAGE (Microarray and Gene Expression). This enables experimental results to be reanalysed and reproduced by other groups, and provides a common platform for data exchange. Similar work is being carried out at present to develop an ontology (shared data format) for proteomics (Jones A., 2003). A common proteomics ontology would facilitate

data queries, data integration and data mining. Several aspects of MAGE are applicable to proteomics although will need to be adapted. Proteomics is more difficult due to the wealth of information provided by the 2-DE and MS data.

The wealth of information produced poses challenges for data management, and necessitates publicly accessible databases that use agreed standards to describe protein data, allowing data comparison and integration (Marte, 2003). Such databases are essential to enable detailed cross-comparisons between different cellular expression patterns under various conditions. This will be important in host-pathogen studies, in which integrated analyses of normal and infected cells, pathogen-expressed genes and host immune system genes will need to be compared. Integration of other post-genomic information, such as proteomic data, will also be needed. Furthermore, the eventual integration of gene-specific information from other databases in regard to structure, function, and biological processes, and of specialist data relating to the pathogens, will hopefully allow for sufficient understanding of host-pathogen interactions and the ability to generate further testable hypotheses (Kellam, 2000).

1.6.1 Protein Bioinformatics

Several software packages and databases now exist which allow for identification of proteins and further characterisation ranging from the calculation of basic physicochemical properties to the prediction of potential post-translational modifications and 3-D structures. Annotated protein and 2-DE databases are the bioinformatic core of proteome research and SWISS-PROT is a typical example of such an annotated database. Malmstrom *et al* (2004) have developed a 2-DE database which profiles the fibroblast proteome during health and disease which provides annotation, sequence and structure identification levels of proteins expressed within the fibroblast proteome (www.2DDB.org) (Malmstrom et al., 2004) In 2001, the first analyses of the working draft of the human genome sequence were published, jointly by Sanger and Celera (Stein, 2004). Many databases containing human genome information are now publicly available on the internet (www.HUGO.com, www.ncbi.ac.uk, www.harvester.embl.de). These databases provide information about the human genes and proteins and can be searched against for gene and protein identification. Proteomics based approaches offer great potential in unravelling complex biological problems such as the nature of particular molecular complexes or pathways in disease pathogenesis. The Harvester search engine at EMBL (www.harvester.embl.de) cross-links public bioinformatic databases and prediction servers to provide fast access to protein specific bioinformatic information. Harvester collates the information from these databases to provide a universal database profiling information about protein sequence, protein

domains and protein localisation. HPRD (www.hprd.org) provides information on protein domain architecture, post-translational modifications, interaction networks and disease association for each protein in the human proteome. The Gene Ontology (GO) project (http://www.geneontology.org) is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The three organizing principles of GO are molecular function, biological process and cellular component.

1.7 Combining genomics, proteomics and bioinformatics

Many methods are needed to understand the complex behaviour of biological systems, meaning that the future will involve a mix of genomics, proteomics, bioinformatics and many other new technologies and disciplines (Russell, 2002).

As microarrays allow for whole genome profiling they are promising tools for sifting through the genome to identify candidate genes as targets for drug development. However, this type of data has limitations because mRNA levels may not reflect protein levels, and expression of a protein may not always have a pathological consequence, therefore traditional pathology and toxicity studies will remain necessary (Kato-Maeda *et al*, 2001).

A study by Le Naour *et al* (2001) combined oligonucleotide microarray and proteomic approaches to uncover novel genes associated with dendritic cell differentiation and maturation, which, in turn, allowed analysis of post-translational modification of specific proteins as part of these processes. The study identified novel genes and proteins involved in dendritic cell differentiation and noted that the proteomic part of the study was paramount as it provided information unavailable at the RNA level (Le Naour *et al.*, 2001).Genomic studies in combination with proteomic analyses of host-pathogen interactions are essential to fully understand virulence and pathogenesis mechanisms (Zhang *et al.*, 2005).

1.8 Proteomics to study host-pathogen interactions

Several studies have profiled the proteome of intracellular pathogens including apicomplexan parasites e.g. *Toxoplasma gondii* (Cohen *et al.*, 2002), *Neospora caninum* tachyzoites and *Eimeria tenella* sporozoites (Belli *et al.*, 2005). Proteomics has been used to investigate the host protein response to several bacteria and viruses and these are reviewed in Zhang *et al* (2005). Jiang *et al* (2005) used DIGE and ICAT to investigate host cell protein expression during infection with severe acute respiratory syndrome (SARS) (Jiang *et al.*, 2005). To date there are no published studies which use proteomics to

investigate the host response to infection with an apicomplexan parasite. A more in-depth understanding of host-pathogen interactions has the potential to improve the understanding of pathogenicity and virulence, thereby defining novel therapeutic and vaccine targets. In addition, proteomic characterisation of the host response could provide pathogen specific host biomarkers for rapid pathogen detection and characterisation, as well as for early and specific detection of infectious diseases (Zhang *et al.*, 2005).

Rabilloud *et al* (1999) and Florens *et al* (2004) investigated the erythrocyte membrane during infection with *Plasmodium falciparum* using 2-DE (Rabilloud *et al.*, 1999; Florens *et al.*, 2004). However these studies only looked at the parasite proteins present in the erythrocyte membrane, and not at any differentially expressed host cell membrane proteins in infected or non infected erythrocytes. Investigating the host cell membrane proteins in the infected erythrocytes may lead to an increased understanding on the host-pathogen intracellular reaction, how the parasite proteins are translocated to the host cell surface, and how the infected host cell interacts with cells from the immune system.

1.9 Conclusion

Proteomics complements genomics-based approaches, providing additional information but presenting different technical challenges. For example, there is no protein equivalent of PCR for amplification of low-abundance proteins, so a range of detection from one to several million molecules per cell is needed. Proteins have properties determined by their 3D structure, and post-translational modifications also add to the analysis challenges. However, many complementary technologies are being developed, along with development of high-quality bioinformatic facilities.

Proteomics benefits from integration with the genomics initiative, the gain from this interface as genome sequencing and functional genomics projects are completed, together with the introduction of national proteomic funding initiatives, should allow proteomics-based approaches to realise their potential in biomedical applications and translation into clinical practice (Banks *et al.*, 2000).

1.10 Aim

The aim of the work presented in this thesis was to model the proteomic changes that occur in mammalian cells during infection with an intracellular pathogen. This is an important area of study as many of the mechanisms by which pathogens survive within the host and modulate host cell functions remain unknown. Comparison of *Toxoplasma gondii* infected and non-infected cells using proteomics will allow for identification of the modulated host cell proteins and pathways during infection and help towards a better understanding of the host response to this pathogen.

The specific objectives of the project are:

- Develop and optimise a reproducible and sensitive method of sample preparation for proteomic analysis of mammalian HFF cells

- Identify changes in the protein profile of *Toxoplasma gondii* infected cells as compared to non-infected cells using 2-DE and mass spectrometry

- Compare the differentially expressed proteins with data from previously published transcriptional studies to determine the gene to protein correlation during parasite infection

- Compare the protein expression response of cells infected with *T.gondii* to the proteome of cells infected with a second intracellular pathogen, *Leishmania major*, to determine whether the host has common response to infection

- Investigate the phosphoproteome of *Toxoplasma gondii*-infected cells to determine the level of post translation modifications and which proteins and pathways are active in the host cell

- Use functional analysis to determine the roles the modulated proteins may be playing within the infected cell and whether these proteins form part of a pro-host or pro-parasite response

- Identify novel proteins and pathways not previously implicated in the host response to infection with *Toxoplasma gondii*

Chapter II Optimisation of proteomic techniques for human foetal foreskin fibroblasts (HFFs)

2.1 Introduction

The aim of the work in this chapter was to optimise proteomic techniques to profile the proteome of human foetal foreskin fibroblasts (HFF). Optimisation of the sample preparation, determination of protein concentration and protein separation by two-dimensional electrophoresis (2-DE) are described.

2.1.1 Proteomics

To profile the HFF proteome, gel-based proteomic technologies were utilised. Figure 2-1 is a schematic of the proteomics workflow. Proteins are separated in the first dimension by isoelectric focusing on a plastic backed polyacrylamide strip containing an immobilised pH gradient. A current is applied to the strip and the proteins migrate in an electric field until they reach their pI. The strip is then applied to a sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel and the proteins separated in the second dimension by their molecular mass. The proteins on the gel are visualised by staining, for example with colloidal Coomassie, and the spots of interest excised. The proteins are digested with trypsin into their component peptides and analysed by mass spectrometry (MS). The resulting peak list from the peptide mass fingerprint is subjected to computer database searching and the protein identified.

2.1.1.1 2-DE

Two-dimensional electrophoresis (2-DE) was first introduced in 1975 by O'Farrell (O'Farrell, 1975) and involves separating proteins firstly by charge (isoelectric focusing (IEF)) and secondly by molecular mass, using SDS-PAGE. This technique has the capacity to separate thousands of proteins (Wu *et al.*, 2002).

2.1.1.2 Sample preparation

One of the most critical steps in 2-DE is sample preparation. The main problem in the visualisation of total cell or tissue extract proteins is the high dynamic range of proteins and their chemical diversity over a wide range of molecular weight, pI and solubility

(Rabilloud *et al.*, 2002). To achieve reproducible protein separation the method of sample preparation must be optimised for the specific sample of interest. Ideally, for successful 2-DE all the proteins in the sample must be soluble, unfolded, have a neutral charge and the molecular interactions be completely disrupted to obtain single polypeptides. This state must then be maintained throughout the separation process. This is however rare to achieve for complex protein mixtures. If only a subset of proteins is of interest (i.e. one subcellular compartment) then prefractionation is carried out during lysis.

Table 2-1 details several published 'classical' lysis methods for tissue cell culture sample preparation for 2-DE.

Cell lysis method	Cell type applicable	Protocol	Reference
Osmotic lysis	Tissue culture cells	Resuspend cells in a hypo-osmotic solution (i.e. water)	Amersham Biosciences 2D Electrophoresis Handbook, Edition AB, 2002
Freeze-thaw lysis	Tissue culture cells	Rapidly freeze cells in LN_2, then thaw (<30°C). Repeat if necessary	Amersham Biosciences 2D Electrophoresis Handbook, Edition AB, 2002
Detergent lysis	Tissue culture cells	Lyse cells in a lysis solution containing detergent (i.e. SDS)	Amersham Biosciences 2D Electrophoresis Handbook, Edition AB, 2002
Sonication and denaturation	Human fibroblast cells	Lyse cells in 2 volumes w/v of lysis buffer 0.025M Tris, 4% CHAPS/ w/v, 1% DTT w/v, 8M urea, 2M thiourea, by brief sonication	(Lehr *et al.*, 2002)

Table 2-1 A selection of published methods of tissue culture cell lysis

For the sample to be fully compatible with 2-DE then there must be no contaminating substances present i.e. salts, nucleic acids, lipids, polysaccharides, small ionic molecules, ionic detergents and phenolic compounds. Protein precipitation (e.g. acetone precipitation) can be used to remove contaminants that may affect IEF.

Choice of lysis of buffer is paramount for successful protein separation by 2-DE. Standard lysis buffer always includes urea and one or more detergents (Amersham Biosciences handbook, Edition AB, 2002). Urea is a neutral chaotrope and denatures proteins causing unfolding and solubilisation of the proteins to their native conformation, exposing the ionisable groups. Thiourea used alongside urea has often been found to improve protein solubilisation (Rabilloud, 1998). A reducing agent (most commonly dithiothreitol (DTT)) is included to break any disulfide bonds present and to maintain proteins in their fully

denatured state. 40mM Tris is included in the lysis buffer to provide a basic environment for full solubilisation and to minimise proteolysis. Zwitterionic amphiphilic compounds (i.e. CHAPS or SB3-10) which prevent aggregation through hydrophobic interactions are also normally included. The combination of all of these components is vital to solubilise most proteins from a wide range of organisms (Rabilloud, 1998). When using IPG immobiline pH gradients the inclusion of IPG buffer, specific to the desired pH range, in the lysis buffer enhances protein solubility by reducing protein aggregation due to charge-charge interactions. Proteolysis greatly complicates 2-DE analysis and therefore must be inhibited. As proteases may be released during cell lysis, a cocktail of protease inhibitors is often included in the lysis buffer, along with nucleases.

2.1.1.3 Isoelectric focusing

As previously mentioned in Chapter I, IEF separates proteins based on their isoelectric point (pI). Proteins are amphoteric molecules, having negative, positive or zero net charge depending on the pH of their environment. The overall charge of a protein is the sum of all the positive and negative charges of its constituent amino acid side chains and carboxyl and amino terminal groups. The pH at which the proteins net charge is zero is termed the isoelectric point. In a pH gradient under current, a protein will move to the position in the gradient where its net charge is zero (pI). A positively charged protein will migrate towards the cathode, losing its charge as it nears its pI and a negatively charged protein will migrate towards the anode. If the protein should diffuse away from its pI it will immediately gain charge and migrate back to position.

IEF is performed in individual immobilised pH gradient (IPG) gel strips, 3mm wide and cast on GelBond PAG film (Immobiline DryStrips) using the Amersham Biosciences system. Two techniques of sample loading are available. The first is cup loading, where the IPG strip is rehydrated in a solution containing urea, tris, DTT, CHAPS and immobilines. The sample is then added to the rehydrated strip using sample cups just prior to focusing. In the second technique, in-gel rehydration, the sample is mixed with the same rehydration buffer as with cup loading and the strip placed on top of the solution which allows the gel to rehydrate to its original thickness of 0.5mm. Proteins move to their pI and are thus separated by charge. Advantages of in-gel rehydration are that larger amounts of proteins can be loaded onto the strip and it is an easier technique to carry out. However, cup-loading may give better resolution of basic proteins.

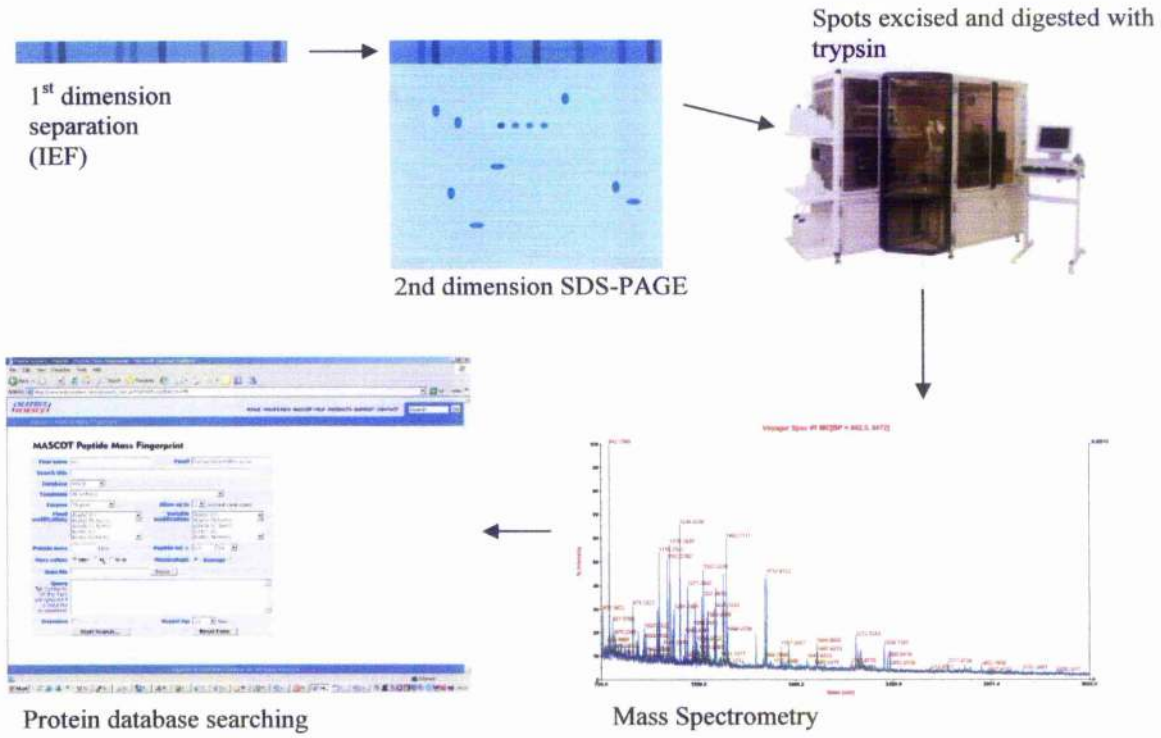


Figure 2-1 Schematic showing the proteomics work flow. Proteins are separated in the first dimension by isoelectric focusing on a plastic backed polyacrylamide strip containing an immobilised pH gradient. A current is applied to the strip and the proteins migrate in an electric fleid until they reach their pI. The strip is then applied to an SDS-PAGE gel and the proteins separated in the second dimension by their molecular mass. The proteins on the gel are visualised by staining the gel, for example with colloidal Coomassie, and the spots of interest excised. The proteins are digested with trypsin into their component peptides and analysed by MS. The resulting peak list from the peptide mass fingerprint is subjected to computer database searching and the protein identified.

Table 2-2 shows suitable sample loads for 2-DE using in-gel rehydration when using silver and Coomassie staining.

Immobiline DryStrip	pH	Suitable sample load (µg protein) Silver stain	Suitable sample load (µg protein) Coomassie stain
7cm	4-7	4-8	20-120
	6-11	8-16	40-240
	3-10, 3-10NL	2-4	10-60
24cm	4-7, 3-7	45-90	200-1300
	6-9, narrow interval*	80-170	400-2000
	3-10, 3-10NL	20-40	100-600

Table 2-2 Suitable 2-DE sample loads for in-gel rehydration using silver and Coomassie. * Immobiline DryStrip narrow intervals pH: 3.5-4.5, 4.0-5.0, 4.5-5.5, 5.0-6.0 and 5.5-6.7 (Amersham Biosciences Handbook, AB Edition, 2002)

To complete IEF a current is applied across the strip which allows movement of the proteins to their isoelectric point on the IPG strip and is performed on an Ettan IPGPhor flatbed system (Amersham Biosciences).

2.1.1.4 Second dimension separation

Separation in the second dimension is based on protein molecular mass. The proteins have previously been separated by their pI during IEF, in an immobilised pH gradient set onto a polyacrylamide gel strip. After focusing is complete the strips are equilibrated which prepares the proteins for the second dimension separation. The strip is washed in an equilibration solution containing buffer (50mM Tris-HCl, pH8.8), urea, glycerol (together with urea reduces electroendosmosis and improves transfer of protein from the 1st to 2nd dimension), DTT (reductant which cleaves disulfide bonds) and SDS (denaturant required for SDS-PAGE). A second wash where DTT is replaced with iodoacetamide (IAA) is also performed. IAA alkylates thiol group that have been exposed by the cleaving of disulfide bonds due to the presence of DTT and ensures that reoxidation is prevented. Protein reoxidation during electrophoresis can result in streaking. This alkylation also prevents reactions of cysteine residues which would affect identification of the protein by mass spectrometry.

The equilibrated strip is then placed onto an SDS-polyacrylamide gel and the proteins separated by electrophoresis. Gels can be cast in the lab or pre-cast gels purchased. Amersham Biosciences pre-cast gels are 12.5% polyacrylamide gels cross-linked with

bisacrylamide. They are compatible with the Ettan Dalt II electrophoresis unit and Ettan Dalt II Buffer Kit. When the pre-cast gel is used in concert with the buffer kit, a discontinuous buffer system is produced. On completion of the second dimension the gel is stained and the image scanned to visualise the proteins.

2.1.1.5 Protein quantification

It is essential to have an accurate value of protein concentration for comparative 2-DE. Protein concentration can be determined using many commercially available kits which differ in their compatibility to reagents including detergents and denaturants present in 2-DE sample buffers. In this chapter, three methods of determining protein concentration were evaluated; BCA Assay (Pierce), Coomassie Plus Bradford method (Bio-Rad) and 2-D Quant Kit (Amersham Biosciences).

2.1.1.6 Protein detection on gels

To visualise the protein spots on a 2-DE gel there are several different protein stains available. The ideal stain will have high sensitivity, be compatible with MS and have a wide linear range for quantification.

Silver staining is the most sensitive non-radioactive method and can detect proteins present in quantities less than 1ng. However this method of staining is incompatible with MS. It is possible to omit glutaraldehye from the sensitiser and formaldehyde from the silver nitrate to make it compatible with MS but this decreases the sensitivity of the stain. Coomassie staining is 50-100 fold less sensitive than silver staining but is relatively simple to use and more quantitative than silver staining (Lopez, 2000). Colloidal Coomassie is more sensitive down to 100ng/protein spot and appears to stain the broadest spectrum of proteins. Fluorescent labelling and staining using SYPRO dyes have sensitivity between colloidal Coomassie and silver staining. SYPRO dyes are compatible with MS and show a wide dynamic range for quantification. SYPRO Ruby protein gel stain can detect 1-10ng of protein and can be linear over 3 orders of magnitude (Bio-Rad, 2-D Electrophoresis for Proteomics Manual)

2.1.1.7 Protein identification

Several types of software are available to analyse the 2-DE gel image and to detect and calculate spot volume. The gel image is loaded into the software and the spots detected.

Dust and scratches can be removed and background subtraction and spot normalisation carried out resulting in the calculation of spot volume. Spots which match between gels of interest can then be compared by spot intensity. Spots of interest can be excised from the gel, either manually or using a robotic spot picker (e.g. Ettan Spot Handling Workstation (Amersham Biosciences)). The protein gel spots are then digested into their component peptides using a specific protease, most commonly trypsin. Trypsin specifically cleaves the peptide bond at the c-terminus of arginine and lysine residues, unless they are followed by a proline. The peptides can then be subjected to MS protein identification.

Various mass spectrometry techniques are available for identification of proteins from 2-DE. The method most commonly used in a large scale identification project where the genome of interest has already been sequenced is MALDI-TOF (Matrix-Assisted-Laser-Desorption-Ionisation-Time-Of-Flight). Using MALDI-TOF, a peptide mass fingerprint (PMF) of the protein can be generated, which consists of the mass of some of the peptides produced by the tryptic digestion of the protein. This method of protein identification, using PMF generated by MS, was first demonstrated in 1993 (Yates, III *et al.*, 1993).

Search engines have been developed to enable the rapid searching of protein and nucleotide sequence databases with PMF data. These common steps are involved for all search engines; (1) each protein in the database is digested *in-silico* based on the specificity of the enzyme and the masses of the resulting peptides are calculated enabling the construction of a theoretical peptide mass fingerprint, (2) the experimentally determined peptide mass fingerprint is compared with the theoretical and a score is calculated which reflects the accuracy of the match between the experimental and theoretical peptide masses and (3) the protein sequences in the database are sorted according to their score and the protein sequence with the best score is selected. The database used to identify the proteins from their component peptides in this study was MASCOT® http://www.matrixscience.com/. The MASCOT® database is also available locally on the University of Glasgow server: http://fun gen1.ibls.gla.ac.uk/mascot/cgi/search_form.pl An example of the MASCOT® search engine can be seen in Figure 2-2.

An alternative MS approach is to use MS/MS. This technique fragments the tryptically produced peptide and calculates the mass to charge ratio for each peptide ion. MS/MS is very useful where there is not a lot of starting material or if there is a mixed spot (more than one protein species present per spot), however it is much more time consuming than MALDI and therefore is lower throughput. MS/MS data can also be searched using the MASCOT® search engine.

The aim of this chapter was to determine the optimal method for the extraction and solubilisation of HFF proteins and therefore to profile the HFF proteome using 2-DE with consistent reproducibility. The protein concentration from each method was first determined and the proteins then separated on 1-DE and 2-DE gels. The lysis methods were then compared by examining the number of protein spots visible, the resolution of the spot pattern and the distribution of protein spots at high and low pI/molecular weight. Reproducibility of the method was also investigated. The result of these optimisations was the production of an HFF protein map and the subsequent identification of 49 proteins by mass spectrometry.

2.2 Materials and Methods

2.2.1 Commonly used reagents

All equipment and reagents for 2-DE (Agarose scaling solution, IPGphor, DALT6 and 12 electrophoresis tanks, ImageMaster 2DElite V3.01 software, Typhoon scanner 9400, Labscan, Immobiline DryStrips, IPG Buffers, TEMED, CHAPS, DTT, APS, urea, thiourea, glycine, SYPRO Orange, Tris and Protein Silver Staining kit) were purchased from Amersham Biosciences. Acetonitrile, ammonium bicarbonate, Coomassie Brilliant Blue G250, EDTA, EGTA, TFA, L-glutamine and Pepstatin A were purchased from SIGMA. DMEM (catalogue# D6546 with 4500mg/L glucose, 110mg/L sodium pyruvate + sodium bicarbonate, substitutes pyroxidine HCl for pyridoxal HCL), Penicillin/Streptomycin and foetal calf serum were purchased from GibcoBRL. SYPRO Ruby, 1-DE gel equipment and 30% polyacrylamide solutions containing bis were purchased from Bio-Rad. Trypsin and 1-DE molecular weight protein markers were purchased from Promega. ZipTip™ pipette tips were purchased from Millipore. DN/RNases were purchased from SIGMA (2000 units/ml DNase, 1,750 units/ml Raise, 50mM $MgCl_2$, 0.5M Tris pH 8.0). The COMPLETE protease inhibitor cocktail was purchased from Roche, and complemented with 10µl of EDTA, EGTA and PMSF solution.

2.2.2 In vitro *culture of HFFs*

HFFs were purchased from the European Collection of Cell Cultures (ECACC, cat#86031405) at passage number 14. HFFs have a finite lifespan as they are a non-immortalised cell line, after passage 20 their replication rate slows down and at higher passage numbers the cells begin to die.

HFFs were grown in 75cm^2 or 175cm^2 culture flasks in 15ml or 45ml of DMEM (supplemented with 10% heat inactivated (56°C, 30 min) foetal calf serum (FCS), 1% Penicillin/Streptomycin (5000units/ml Penicillin, 5000µg/ml Streptomycin) and 0.584g/l, L-glutamine) respectively. Flasks were incubated at 37°C, 5% CO_2.

Confluent HFFs (passage < 20) were passaged by washing the cells in FS HEPES buffered saline solution (NaCl 8g, KCl 0.4g, d-Glucose 1g, Hepes 2.38g, 0.5% Phenol Red 2ml, pH adjusted to 7.5 using 5M NaOH in 1L ddH_2O and filter sterilised), this was poured off and

versine:trypsin solution (4:1) added and left for approximately 6 min at 37°C, 5% CO_2. The flask was tapped several times to dislodge all cells and the cells and liquid collected. The number of cells harvested was counted using a haemocytometer to give number of cells per ml and then pelleted at 2900rpm, 4°C for 10 min. The versine:trypsin solution was discarded and the cell pellet resuspended in pre-warmed DMEM at a concentration of $1.5x10^6$/ml. $2x10^4$cells were added per cm^2 to a flask (75cm^2 or 175cm^2 flasks grown in 15 ml or 45ml of DMEM correspondingly).

To photograph the HFFs, the cells were grown in 1cm^2 chamber slides (Lab-tek, Nalge Nunc International) in 500µl DMEM + 10% FCS, Pen/Strep and L-glutamine). Once confluent the media and chambers were removed from the slide and the slide left to air dry. The slide was then placed in absolute methanol for 5 min and allowed to air dry. The slide was then placed directly into Geimsa stain (10% Geimsa stain (BDH), 20% Geimsa buffer (3g Na_2HPO_4 (anhydrous), 0.6g KH_2PO_4 in1L ddH_2O) and 70% ddH_2O) for 15min. The slide was washed 3 times in tap water and left to air dry. The fixed cells were then viewed and the images captured using a Zeiss Axiovert 200M light microscope at 200x and 400x magnification.

For storage of cells in liquid nitrogen (LN_2), at passage numbers 14-17, HFFs were harvested as described previously, washed three times in phosphate buffered saline ((PBS) 8g NaCl, 0.2g KCl, 1.44g Na_2HPO_4, 0.24g KH_2PO_4 in 1L ddH_20, pH adjusted to 7.4 using HCl) and stored at -80°C overnight (at a concentration of 2-$4x10^6$/ml of freezing medium; DMEM, 20% FCS, 10% DMSO) and then transferred for storage in LN_2.

For all 1-DE SDS PAGE and 2-DE analysis HFFs were harvested using versine:trypsin as described, the cells counted, pelleted by centrifugation at 2900rpm, 4°C for 10 min and the cells washed in PBS and then repelleted by centrifugation. This last wash step was carried out three times in total. All PBS was removed from the cell pellet and the sample stored at -80°C until ready for use.

HFFs were routinely tested for Mycoplasma contamination using the Stratagene PCR Mycoplasma test and no mycoplasma were found. The manufacturer's protocol was followed.

2.2.3 *Protein Extraction Methods*

There are several published classical cell lysis methods (Table 2-2). In this chapter four lysis methods were tested; a published protocol taken from Lehr *et al.* (2001) and three modified protocols which combine elements of previously published protocols on the protein extraction from tissue culture cells for use with 2-DE.

Method 1 (taken from the paper by Lehr *et al*, 2001). In the culture flask the HFF cells were washed twice in FS-HEPES, harvested by scraping in solution and collected in a centrifuge tube. The flask was then washed once again and the wash collected to ensure all the cells were removed. The cells were pelleted by centrifugation at 8,000 g for 5 min at 4°C and the pellet stored at –70°C. The cell pellet was lysed in 2 volumes w/v of lysis buffer 3 (lysis buffer 3: 8M urea, 2M thiourea, 1% DTT, 4% CHAPS, 40mM Tris-base) and sonicated briefly using a water bath sonicator. The lysate was centrifuged at 13,000 g for 40 min at 4°C. The protein concentration of the lysate was determined and the supernatant stored at –70°C.

Method 2 HFFs were harvested as described in section 2.2.2. The cell pellet was washed three times in PBS. To $1x10^6$ cells, 60µl of lysis buffer 2 were added (7M urea, 2M thiourea, 4% CHAPS, 40mM Tris-base) along with 5µl DNase/RNase and 2.5µl protease inhibitors. The sample was left at room temperature for one hour during which it was vortexed several times. After one hour it was centrifuged at 13,000 g for 3 min to collect any insoluble material. The protein concentration of the supernatant was determined and then frozen at –70°C.

Method 3 HFFs were harvested as described in section 2.2.2. To $1x10^6$ cells, an aliquot of 60µl of lysis buffer 1 (8M urea, 4% CHAPS, 40mM Tris-base) was added, along with 5µl DNase/RNase and 2.5µl protease inhibitors. The sample was mixed by pipette, frozen in LN_2, heated to 29°C to thaw, and then vortexed for 20 sec. This was repeated 5 times in total. The sample was centrifuged at 13,000 g for 3 min to pellet any insoluble material. The protein concentration was determined and the supernatant stored at –70°C.

Method 4 As Method 3, but using lysis buffer 2 (7M urea, 2M thiourea, 4% CHAPS, 40mM Tris-base).

A sonicator water bath was used to sonicate the samples prepared by methods 1, 2 and 4 described above, but gave no significant advantages in protein extraction.

2.2.4 Protein concentration methods

Three methods to determine protein concentration of the cell lysate were assessed. (1) BCA Protein Assay (Pierce), (2) Bio-Rad Protein Assay and (3) Amersham Biosciences 2-D Quant kit.

2.2.4.1 BCA Protein Assay (Pierce)

The BCA Protein Assay kit was purchased from Pierce (catalogue # 23225). Bovine serum albumin (BSA) standards were prepared at concentrations of 0µg/ml – 2000µg/ml, using the same diluent as present in the sample of interest (e.g. sample lysis buffer) (Table 2-3).

BCA (bicinchoninic acid) working reagent (WR) was prepared by mixing 50 parts BCA reagent A with 1 part BCA reagent B. Each standard or unknown sample was dispensed by pipette into a microwell plate well in triplicate. 200µl WR were added to each well and the plate incubated at 37°C for 30 min on a plate shaker. The absorbance at 570nm was measured using an ELISA plate reader. The protein concentration of the sample of interest was determined with reference to the standard curve of the absorbance versus the concentration of the BSA standard.

2.2.4.2 Bio-Rad Protein Assay

Standards of BSA (Pierce, catalogue # 23209) ranging from 0-5µg/ml were prepared in the same diluent as present in the sample of interest (e.g. sample lysis buffer) (Table 2-4). Coomassie working reagent was prepared by mixing 1 part Coomassie Bio-Rad Assay Reagent with four parts ddH_2O. 10µl of each standard were added into the appropriate wells of a 96 well microplate (in duplicate) as indicated in the table. Sample was added into an empty well and serial dilutions of the sample made. 200µl of diluted working reagent were added to each well. Within an hour of adding the working reagent to the plate, the absorbencies of the standards and samples were read at 590nm. A straight line graph was produced by plotting known protein concentration of the standards against the absorbencies of the standards. From this graph the protein concentration of the samples of interest was determined.

BSA	diluent	final BSA concentration
300µl(stock)	0µl	2000µg/ml
375µl(stock)	125µl	1500µg/ml (A)
325µl(stock)	325µl	1000µg/ml (B)
175µl(of A)	175µl	750µg/ml (C)
325µl(of B)	325µl	500µg/ml (D)
325µl(of D)	325µl	250µg/ml (E)
325µl(of E)	325µl	125µg/ml (F)
100µl(of F)	400µl	25µg/ml (G)
0µl	400µl	0µg/ml (H) blank

Table 2-3 BCA Assay (Pierce) standards

well	µl BSA	µl diluent	dilution of BSA	[BSA] µg/µl	BSA µg
A1/B1	0	10	0	0	0
C1/D1	2.5	7.5	1 IN 10	0.05	0.5
E1/F1	5	5	1 IN 10	0.1	1
G1/H1	7.5	2.5	1 IN 10	0.15	1.5
A2/B2	10	0	1 IN 10	0.2	2
C2/D2	2.5	7.5	1 IN 1	0.25	2.5
E2/F2	3	7	1 IN 1	0.3	3
G2/H2	3.5	6.5	1 IN 1	0.35	3.5
A3/B3	4	6	1 IN 1	0.4	4
C3/D3	4.5	5.5	1 IN 1	0.45	4.5
E3/F3	5	5	1 IN 1	0.5	5

Table 2-4 Bio-Rad Protein Assay Coomassie Plus standards

tube	1	2	3	4	5	6
BSA (2mg/ml)	0µl	5µl	10µl	15µl	20µl	25µl
quantity	0µg	10µg	20µg	30µg	40µg	50µg

Table 2-5 2-D Quant Kit (Amersham Biosciences) standards

2.2.4.3 2-D Quant Kit (Amersham Biosciences)

The 2-D Quant Kit was purchased from Amersham Biosciences (Catalogue # 80-6483-56). The working colour reagent was prepared by mixing 100 parts of colour reagent A to 1 part of colour reagent B. BSA standards were prepared as according to Table 2-5. To each of the sample and the standards (in duplicate), 500µl of precipitant were added, vortexed briefly and incubated at room temperature for 2-3 min. 500µl of co-precipitant were added and mixed by inversion, then centrifuged at a minimum of 10,000 g for 5 min to sediment the protein into a pellet. Following centrifugation the supernatant was immediately discarded and then pulse centrifuged to collect the remaining liquid. This was removed to leave a protein pellet. 100µl of copper solution and 400µl of ddH_2O were added to each pellet and vortexed to dissolve the protein. 1ml of working colour reagent was added and mixed immediately by inversion. The samples and standards were incubated at room temperature for 15-20 min. The absorbance was read at 480nm, using a Hewlett Packard UV ChemStation Spectrophotometer. A straight line graph of standard concentration versus absorbance was plotted and the protein concentration of the samples of interest calculated from this.

2.2.5 1D SDS PAGE mini gels

Cell extracts were separated on 1-DE SDS PAGE mini gels to assess the quality of extraction methods 1-4. The proteins were extracted from the HFFs using the four methods described in section 2.2.3, and the protein concentration determined using the BioRad Protein Assay.

Proteins were separated on a 15% resolving gel (0.375% Tris-HCl, pH8.8, 1% (w/v) SDS, 1% (w/v) ammonium persulfate, 10µl TEMED, 15% polyacrylamide-bis) with a 5% stacking gel (0.19M Tris HCl, pH 6.8, 1% (w/v) SDS, 1% (w/v) ammonium persulfate, 5µl TEMED, 5% polyacrylamide-bis). Volumes of sample containing 2µg and 10µg were mixed with SDS loading buffer (62.5mM Tris-HCl, pH6.8, 10% glycerol, 2% (w/v) SDS, 0.4ml 2-mercaptoethanol, 0.05% (w/v) bromophenol blue) and then boiled for 10 min at 100°C and centrifuged at 12,000 g to collect any insoluble material. The ladder (broad range protein molecular weight markers from Promega, catalogue # V8491) did not require boiling. 10µl of each sample (containing 2µg or 10µg of sample) and 5µl of ladder were loaded onto the SDS-PAGE stacking gel.

Separation was performed by electrophoresis at 120V in 1x SDS electrophoresis buffer (25mM Tris-HCl, pH 8.3, 192mM glycine, 0.1% SDS) until the dye front reached the bottom of the gel. Gels were stained in colloidal Coomassie (described in section 2.2.7). Images were acquired at 300 dots per inch using Amersham Biosciences Labscan v3.0 software on an Amersham Biosciences Umax flatbed scanner with integrated transparency adapter.

2.2.6 2-DE of HFFs

2.2.6.1 2-DE 7cm gel

Protein extracts were prepared using methods 1-4 as previously described and the protein concentration determined. For a 7cm 2-DE mini gel, between 10 and 60µg of protein were loaded when the gel was to be stained with colloidal Coomassie. Rehydration buffer (8M urea, 2% CHAPS, 0.002% bromophenol blue) was added to the cell lysate so the final sample volume was 130µl. 2mg/ml DTT and 1% IPG buffer (of the same pH as the Immobiline DryStrip to be used) were added to the rehydration buffer just prior to use. The sample was left at room temperature for 1h then centrifuged at 13,000 g for 3 min to pellet any insoluble material. 125µl of rehydration solution containing sample were applied to a 7cm ceramic strip holder ensuring no bubbles were introduced, and an IPG Strip of appropriate pH was placed on top of the sample. IPG DryStrip cover fluid was applied to the entire surface of the strip to help minimise evaporation and urea crystallisation in the sample. Rehydration was carried out at 30V for 10-15h, followed by IEF at 500V for 1h, 1000V for 1h and 5000v for 4h at 20°C with current limited to 40µA/strip.

On IEF completion, the rehydrated IPG strip was first equilibrated for 15 min by rocking in equilibration buffer (2% SDS, 50mM Tris-HCl pH 8.8, 6M urea, 30% (v/v) glycerol, 0.002% bromophenol blue) containing 10mg/ml DTT (to reduce the proteins) and then a second 15 min wash with equilibration buffer including 25mg/ml iodoacetamide (IAA) (to alkylate the cysteine residues). The equilibrated strip was then rinsed in 1x SDS electrophoresis buffer (25mM Tris-HCl, pH 8.3, 192mM glycine, 0.1% SDS) and placed onto the top of the 15% SDS-PAGE gel. The strip was sealed into place using agarose sealing solution and run at 30 mA in a BioRad mini gel tank until the dye-front had reached the bottom of gel.

2.2.6.2 2-DE 24cm gel

For a 24cm 2-DE gel, approximately 300µg of protein per gel was optimal for use with colloidal Coomassie stain. Sample preparation method 3 was followed, with a starting material of $3x10^6$ HFFs. To the cells, 80µl lysis buffer (8M urea, 4% CHAPS, 40mM Tris-base), 10µl DNase/RNase mix and 5µl protease inhibitor mix were added. The sample was immediately frozen in LN_2, thawed at 29°C and then vortexed for 20secs. This freeze/thaw/vortex step was repeated five times in total. The protein concentration of the extract was determined using the BioRad Protein Assay. Rehydration buffer (8M urea, 2% CHAPS, 0.002% bromophenol blue, 2mg/ml DTT, 1% IPG buffer) was added to the sample to give a final volume of 460µl.

The sample was left at room temperature for 1h then centrifuged at 13,000 g for 3 min to pellet any insoluble material. 450µl of rehydration solution containing sample were applied to a clean 24cm ceramic strip holder, ensuring no bubbles were introduced and an IPG strip placed on top. IPG DryStrip cover fluid was applied to cover the entire surface of the strip which minimises evaporation and urea crystallisation in the sample. Rehydration and IEF for a 24cm strip with in-gel rehydration was performed at 20°C and using 40µA/strip. Rehydration was carried out at 30V for 10-15hours, followed by IEF at 500V for 1h, 1000V for 1h and 8000V for 8h at 20°C with current limited to 40µA/strip.

When the IEF was complete, the rehydrated IPG strip was first equilibrated for 15 min by rocking in equilibration buffer (2% SDS, 50mM Tris-HCl pH 8.8, 6M urea, 30% (v/v) glycerol, 0.002% bromophenol blue) containing 10mg/ml DTT and then a second 15 min wash with equilibration buffer including 25mg/ml iodoacetamide. The equilibrated strip was washed in electrophoresis buffer and placed onto the top of the 24cm pre-cast gel and the strip sealed into place using agarose sealing solution.

Gels were run in the Ettan Dalt 6 or 12 tanks using the SDS anode-cathode buffer kit (Amersham Biosciences). The lower part of the tank was filled with ddH_20 + anode buffer. The gels were placed into the tank and the remainder of the lower chamber filled with water. The upper chamber was filled with ddH_20 + cathode buffer to provide a discontinuous buffer system. The gels were run at a maximum of 14W per gel until the dye front reached the bottom of the gel.

2.2.6.3 2-DE gel staining and image acquisition

Gels were stained with Coomassie Brilliant Blue, colloidal Coomassic, silver or SYPRO stain.

Gels stained with Coomassie blue were incubated overnight in a Coomassie staining solution (0.5g Coomassie Brilliant blue G250, 100ml Acetic acid, 125ml isopropanol and ddH_2O up to 1L). Gels were destained by several washes in a solution of 10% acetic acid, 20% methanol until background staining had disappeared.

Gels stained with colloidal Coomassie were fixed for 2h in 40% ethanol, 10% acetic acid. The gel was washed twice, for 10 min each, in ddH_2O. The gel was then left in stain (4:1 colloidal stock:methanol) for 1-7 days (Coomassie stock: 5% Coomassic G250 in ddH_2O, colloidal stock: 50 g ammonium sulphate, 500ml ddH_2O, 6ml phosphoric acid, 10ml Coomassie stock.) The gel was rinsed with water to remove excess stain.

Silver staining was performed using the Protein Silver Staining Kit according to the manufacturer's instructions (Amersham Biosciences). All solutions were made up in 250ml ddH_2O, ingredients marked with an * were added immediately before use. The gel was fixed by soaking in fixing solution (10% glacial acetic acid, 40% ethanol) for 30 min. The fix solution was removed and sensitising solution added (30% ethanol, 0.005% glutaraldehyde (25%w/v)*, 0.04% sodium thiosulphate (5%w/v), 1 packet of sodium acetate (17g)) and left shaking for at least 30 min. The solution was removed and the gel washed in ddH_2O three times for 5 min each. Silver solution was then added (10% silver nitrate solution (2.5%w/v), 0.0004 formaldehye (37%w/v)*) and left shaking for 20 min. The silver solution was removed and the gels rinsed in ddH_20 twice for 1 min each time. To develop the stain, developing solution was added to the gel (1 packet of sodium carbonate (6.25g), 0.0002% formaldehyde (37%w/v)*) and left shaking for 2-5 min until spots developed and before background got too high. To halt the developing, the solution was removed and stop solution added (1 packet EDTA-$Na_2 2H_20$ (3.65g)) and left shaking for 10 min. The gel was washed 3x for 5 min each in ddH_20.

Staining with SYPRO Orange and SYPRO Ruby was per the manufacturer's instructions. For SYPRO Orange staining the gel was fixed in 10% MeOH, 7.5 % Acetic acid for 1h. The gel was then probed with 0.05% SDS for 1h. The gel was stained with SYPRO Orange for 2h (1 part SYPRO Orange in 5000 parts 7.5% acetic acid). The gel was then washed in 7.5% acetic acid for 1 min to remove any excess stain. For SYPRO Ruby staining the gel

was fixed in 40% methanol, 10% acetic acid for 1h. Gels were then stained with SYPRO Ruby protein gel stain for 4h to overnight with continuous gentle agitation. Post staining the gels were washed in 10% methanol, 7% acetic acid for 1h, this step was repeated.

Prior to scanning, reference markers were applied to the plastic gel back. These markers are included in the spot detection and the Ettan Spot handling workstation uses them to orientate around the gel. Coomassie Brilliant blue, colloidal Coomassie and silver stained gel images were acquired using an AmershamPharmacia Biotech Labscan. Images were acquired at 300 dots per inch using Labscan sv3.0 software on a Umax flatbed scanner with integrated transparency adapter. Images from the fluorescently labelled SYPRO gels were acquired using a Typhoon scanner 9400 (variable mode imager). Settings for SYPRO Ruby scanning were emission filter 610BP30, PMT 450-800, laser green 532, sensitivity normal, pixel size 100 microns and for SYPRO Orange scans the emission filter was 580BP30, laser green 532nm, sensitivity normal, PMT 450-550.

Spots were detected on the gel using the 2DElite automatic spot detection programme. A small section of the gel is manually selected by the user, and the software then detects all spots present in the area. The user is then able to manually adjust the sensitivity of the spot detection. Those parameters are then applied to the whole gel. The reference markers are manually annotated by the user. The reference marker on the left (acidic end) of the gel is annotated as IR1, and the marker on the right (basic) is annotated as IR2. Dust particles identified as spots by the software are removed manually by the user. Background subtraction and normalisation are carried out using the software, and spot volumes calculated. The spot co-ordinates are recorded and can then be used in the spot picking workstation for automated spot picking.

2.2.6.4 Protein identification using mass spectrometry

Protein spots were excised from the gels using two methods; manually or using the Amersham Biosciences Ettan Spot Handling Work Station. Spots were manually excised from the gel using a scalpel. The gel piece was then transferred to an Eppendorf tube and washed for 1h in 100mM ammonium bicarbonate, 1h in 500µl 50% acetonitrile/100mM ammonium bicarbonate, 10 min in 100% acetonitrile and then dried in a vacuum centrifuge. The gel pieces were rehydrated with 0.2µg trypsin in 25mM ammonium bicarbonate for 15 min then 20µl of 25mM ammonium bicarbonate were added to cover the gel pieces. In-gel digestion was carried out overnight at 37°C on a horizontal shaker. The sample was then acidified by addition of 3% TFA and the resulting peptides were

concentrated and desalted using ZipTip™ pipette tips containing C_{18} reverse phase media. The ZipTip™ was equilibrated by washing twice with 50% acetonitrile and then twice with 0.1% TFA. Peptides were bound to the reverse phase by aspirating the sample approximately 15 times into the tip. The bound peptides were then washed twice with 0.1% TFA. Peptides were eluted in 2µl of elution solution (50% acetonitrile/0.1% TFA). The concentrated desalted sample was mixed with an equal volume of matrix solution (10mg/ml α-cyanohydroxycinnimic acid (CHCA) in 50% acetonitrile/0.1% TFA). 1µl of this mixture was then spotted onto a MALDI plate and left to air-dry for 30min.

The Ettan Spot Handling Work Station can be used for automated picking of the protein spots from the gel, tryptic digestion of the proteins, desalting and concentration of the peptide sample, and spotting of the sample mixed with matrix onto a MALDI target plate. The work station also produces two microplates, the first containing the picked gel plugs which the proteins were digested from, and the second, containing the dried down peptides. The peptides can then be used for further protein analysis if necessary (e.g. MS/MS).

Two types of MS were available for protein identification; MALDI MS (PerSeptive Biosystems Voyager-DE™ Pro, Biospectrometry Workstation) and tandem MS (MS/MS) (QSTAR PulsarI MS, Applied Biosystems) a quadrupole time-of-flight mass spectrometer. MS/MS will be discussed in Chapter III. Peptides were crystallised in matrix consisting of a saturated solution of α-cyano-4-hydroxycinnamic acid (CHCA) prepared in 50% (v/v) acetonitrile containing 1% (v/v) aqueous trifluoroacetic acid (TFA) and applied directly to the MALDI plate. Three sets of 150 shots/spectrum were collected for each spot. The spectra from the MALDI were first baseline corrected, smoothed, calibrated against internal trypsin autolysis peptides and deisotoped. The search parameters used for all protein searches can be seen in Table 2-6.

The peptide mass fingerprint obtained by MS was subjected to database searches by pasting the peptide masses into a protein database (MASCOT®, Matrix Science), which uses a probability based scoring system to define a significant match, and therefore protein identification. The data entry form for MASCOT® which allows selection of these parameters can be seen in Figure 2-2.

Parameter	Selected Option
Database	NCBInr
Enzyme	Trypsin
Protein Mass Range	800-3000KDa
Tolerance	50-70ppm
Missed cleavage	1
Fixed modification	Carboxymethylation
Variable modification	Oxidation of methionine

Table 2-6 Parameters selected for database searches using MS data. These parameters were used with the MASCOT® search engine

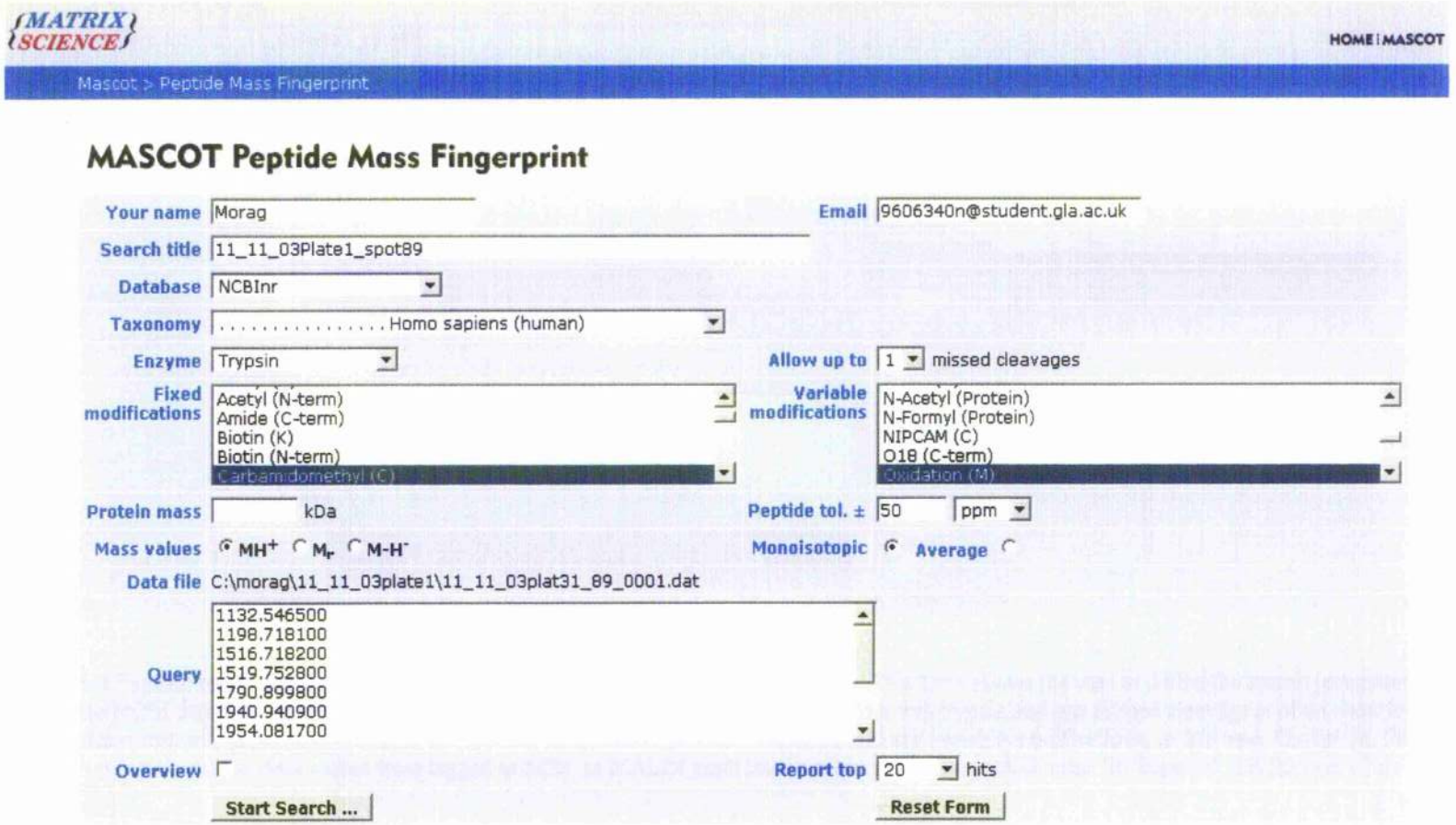


Figure 2-2 Peptide mass fingerprint search form from the MASCOT® search engine. This form allows the user to define the search parameters. In this case the MS data was searched against the NCBI database, specifically Homo sapiens, the enzyme used to digest the protein was trypsin and one missed cleavage is often observed with in-gel digestion. The fixed modifications that will occur on all proteins are entered, in this case, Carbamidomethyl, and any variable modifications, in this case, Oxidation. Peptide mass tolerance was set to 50 parts per million (ppm) and the mass values were tagged as MH+ as MALDI peptides are protonated. The peptide mass list is pasted into the query box and the data searched.

2.3 Results

2.3.1 HFFs

HFFs typically reach confluence (~ $3x10^4$ cells/cm^2) in ~5-10 days depending on passage number (Roos *et al.*, 1994). Therefore, in a confluent T75 flask (around 5-10 days old), there will be ~ $2.25x10^6$ cells. This number was found to vary between $2.25x10^6$ and $3x10^6$ HFF cells in a confluent T75 flask. Cells were counted using a haemocytometer.

HFFs were grown in $1cm^2$ chamber slides (Lab-tek, Nalge Nunc International) in 500µl DMEM + 10% FCS, Pen/Strep and L-glutamine) at 37°C, 5% CO_2 until confluent. The cells were then stained with Geimsa and viewed using a Zeiss Axiovert 200M light microscope at 200x and 400x magnification and the resulting HFF image can be seen in Figure 2-3.

2.3.2 Mycoplasma testing

Cells were routinely tested for *Mycoplasma* using the Stratagene PCR detection kit. *Mycoplasma* contamination has been shown to alter cellular growth rates, morphology, and viability, affect metabolism of amino and nucleic acids and induce chromosomal and membrane aberrations. This can lead to artifactual results in experimental systems (Raab, 1999). The results were always negative indicating that the HFFs were not infected with *Mycoplasma*. The 1-DE gel image in Figure 2-4 confirms that the HFFs were not infected with *Mycoplasma*.

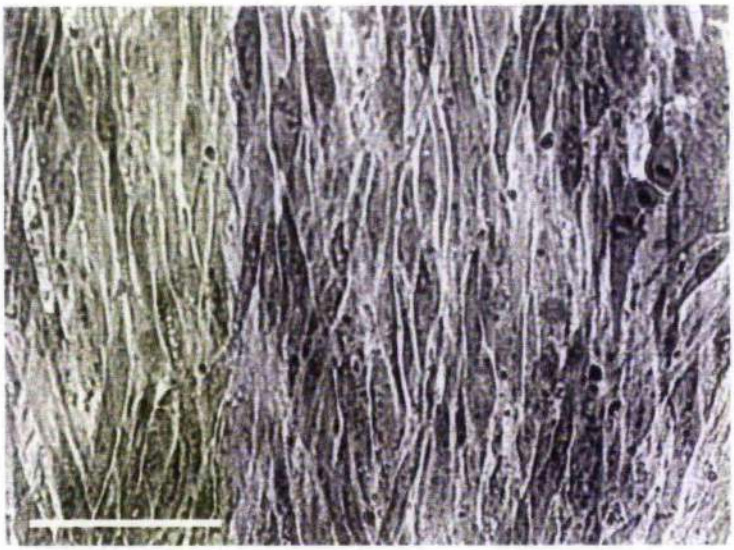


Figure 2-3 An image of confluent HFFs. Confluent HFFs grown in chamber slides were stained with Geimsa and photographed using a Ziess Axiovert 200M light microscope at 200x magnification. The cells grow parallel to each to other and are thin and elongated in shape. Bar equals 50μm.

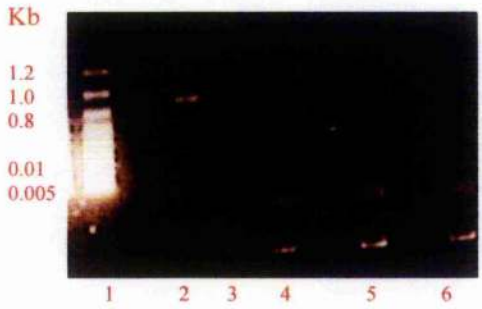


Figure 2-4 1-DE gel showing absence of Mycoplasma DNA in HFFs. Lane 1= ladder (Invitrogen 1Kb DNA ladder, catalogue # 15615-016), lane 2 = positive control (874b), lane 3 = negative control, lane 4 = HFF sample 1, lane 5 = HFF sample 2, lane 6 = HFF sample 3. A band of the correct size for a positive result using the Mycoplasma PCR test can be seen in lane 2, positive control. This band is not present in any of the HFF sample lanes, therefore we can conclude that the HFFs were not contaminated with Mycoplasma.

2.3.3 Assessment of the four sample preparation methods

The four methods of sample preparation for the extraction and solubilisation of protein from HFFs were compared by amount and quality of solubilised protein yielded, and compatibility with 1-DE and 2-DE electrophoresis. Quantitative 2-DE is totally dependant on accurate protein quantitation to ensure equal amounts of protein are compared between samples of interest. After isolating the proteins according to methods 1-4, the protein concentration of each HFF sample was determined. Due to differences in compatibility of each of the protein concentration protocols with the lysis buffer components, three protein concentration determination methods were assessed. The same amount of starting material (number of HFFs) was used for each lysis method.

2.3.3.1 BCA Protein Assay Kit (Pierce)

The BCA Protein Assay Kit is a detergent compatible assay which is based on bicinchonic acid (BCA) for the colorimetric detection and quantitation of protein. Protein in an alkaline environment reduces Cu^{+2} to Cu^{+1} (the biuret reaction). This change can be detected with high sensitivity and selectivity by colorimetric detection of the cuprous cation (Cu^{+1}) using a BCA reagent. This assay can detect protein between the concentrations of 20-2000µg/ml.

2.3.3.2 Bio-Rad Protein Assay

Bio-Rad Protein Assay has a detection range of 1µg/ml to 1500µg/ml of protein and is compatible with the common 2-DE sample components urea (up to 6M), thiourea and SDS (up to 0.1%). The Bio-Rad Protein Assay is based on the Bradford method (Bradford, 1976) which is based on an absorbance shift from 465 nm to 595 nm that occurs when the Coomassie Brilliant Blue G-250 dye reagent binds to proteins in an acidic solution. The mechanism of the reaction is the anionic form of the dye complexes with protein, resulting in a colour change from brown to blue. The dye interacts chiefly with arginine residues and weakly with histidine, lysine, tyrosine, tryptophan and phenylalanine residues. The reaction reaches a stable endpoint so absorbance does not shift over time. Absorbency was measured at 595nm using a spectrophotometer or microplate reader.

2.3.3.3 2-D Quant Kit (Amersham Biosciences)

The 2-D Quant kit was specifically designed to be compatible with sample preparation lysis buffers for 2-DE and is compatible with high concentrations of urea and thiourea. It

uses a combination of precipitant and co-precipitant to quantitatively precipitate sample protein whilst leaving interfering contaminants in solution. Protein is resuspended in an alkaline solution of cupric ions that bind to the polypeptide backbones of proteins. A colorimetric reagent, which reacts with unbound cupric ion, is then added. The colour density is inversely related to the concentration of protein in the sample. Protein concentration can then be accurately estimated by comparison to a standard curve.

The BCA Protein Assay was incompatible with methods 1, 2 and 4 and only gave a protein concentration for method 3. The concentration determined was 2.87μg/μl. The BioRad protein assay gave good reproducible results for methods 1, 2 and 3. The protein concentration of these samples was 1.7, 2.5 and 2.7μg/μl for sample preparation methods 1, 2 and 3 respectively. The 2D Quant kit gave very variable results between duplicate samples. Advantages of the Coomassie Plus Protein Assay over the BCA method are that the standard curve is much more linear, no reagent preparation is required, no dilution or filtering is necessary and the reaction is very fast. The Bio-Rad protein assay was therefore used for all further protein concentration determination in this thesis.

2.3.4 Comparison of sample preparation methods

Proteins were extracted from HFFs using the four sample preparation methods and the protein concentration of the supernatant determined. To compare the methods of protein extraction, 2µg and 10µg of each extract was run on a 1-DE 15% polyacrylamide gel (Figure 2-5). The gels were stained with colloidal Coomassie. As can be seen in the figure; lanes 3, 4 and 5 contain more bands with stronger staining than the bands visible in lane 1. This corresponds to methods 2, 3 and 4 profiling more of the HFF proteome when separated by 1-DE than method 1.

For further comparison of sample preparation, the proteins from each extract were separated using 2-DE. For each extract, 30µg of protein was loaded onto a 7cm Immobiline Strip pH3-10NL and 2-DE carried out. Gels were run in duplicate. To visualise the protein spots the gels were stained with Coomassie Brilliant blue and can be seen in Figure 2-6. Sample preparation method 1 shows poor resolution of lower weight molecular proteins. Method 2 shows horizontal streaking and again poor resolution of lower molecular weight proteins. Method 3 shows good separation and resolution of proteins from high to low molecular weight as well as good coverage of acidic through to basic proteins. Method 4 shows horizontal streaking, and what appears to be 'charge trains' which can be caused by protein phosphorylation or by carboxylation from urea breakdown. Sample preparation method 3 gave the best overall separation and resolution of proteins for 2-DE. Using the Amersham Biosciences 2DElite software, spot detection was carried out on the mini-gels. All artefacts (e.g. dust or scratches) were removed from the detection. The gel image from sample preparation method 3 showed 117 spots, present across the gel, from high to low pH, and small to large MW. This analysis was repeated using another four samples of HFF extracts prepared using methods 1-4. These extracts (30µg) were run on pH3-10NL gels, 7cm, and separated in the second dimension using 15% polyacrylamide gels, and stained with colloidal Coomassie. The gel images can be seen in Figure 2-7. The images were analysed using 2DElite software and the number of spots present detected. Again, method 3 has the widest range of spots present across the gel, from acidic to basic, and low to high MW. There is also no protein streaking present with Method 3. The number of spots detected on each set of mini-gels can be seen in Table 2-7.

For all subsequent analysis, sample preparation method 3 was chosen as it gave the highest number of spots and widest coverage of proteins, with good resolution.

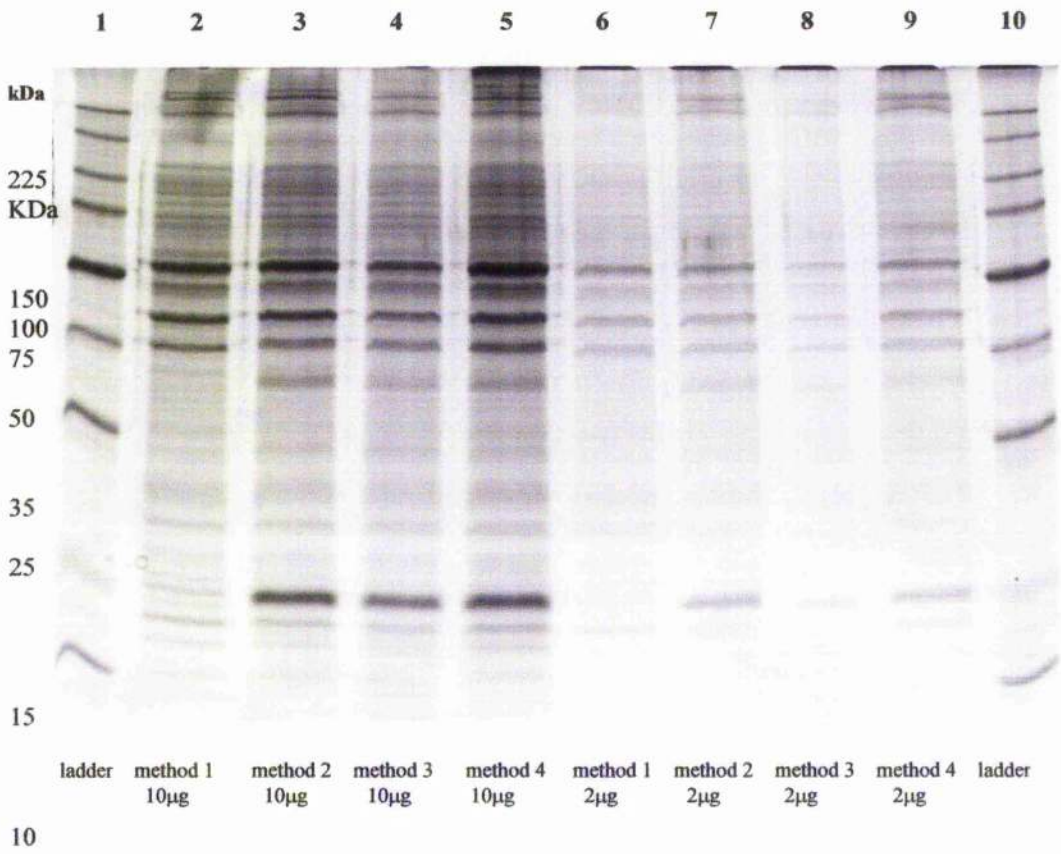


Figure 2-5 1-DE SDS-PAGE separation of proteins extracted using methods 1-4. Proteins from $3x10^6$ HFFs were extracted using methods 1 to 4 and the protein concentration determined. The same quantity of each sample (10µg or 2µg) was run on a 15% resolving gel to assess the quality of the extracts and therefore the method of sample preparation. The gel was stained with colloidal Coomassie. Lanes 1 and 11: broad range protein molecular weight marker (Promega, catalogue# V8491. Clear differences can be seen between methods 1 and methods 2-4. Method 1 profiles slightly fewer bands than the other three methods.

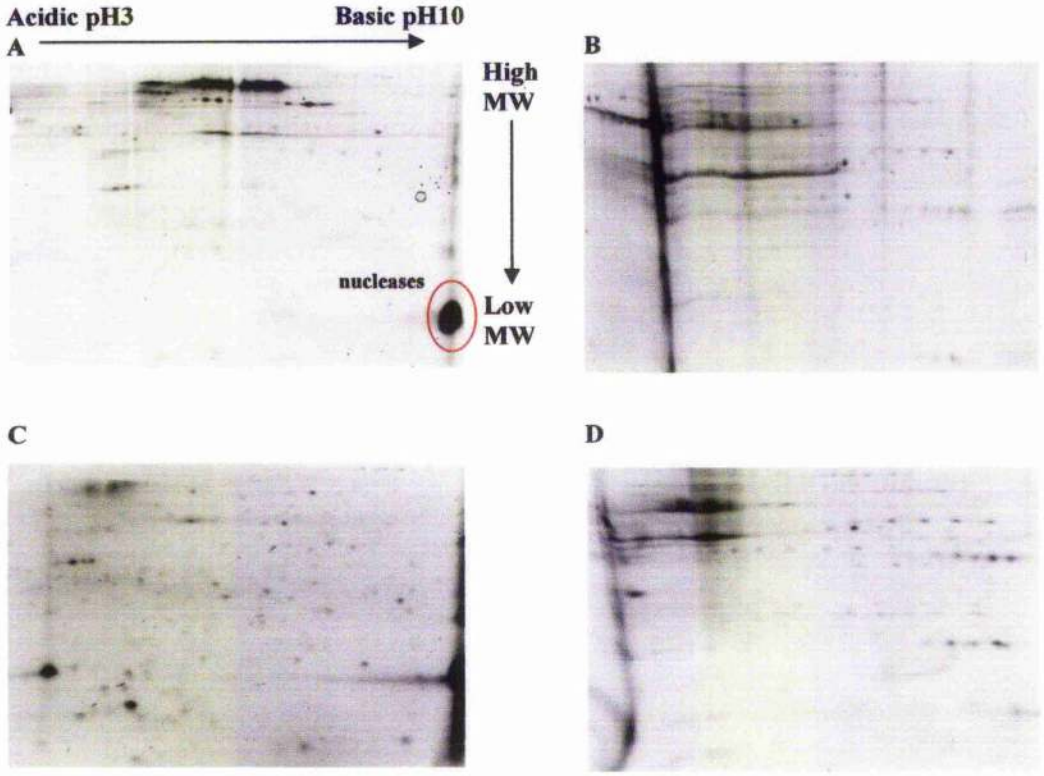


Figure 2-6 Comparison of the four methods of sample preparation by 2-DE. Starting material for all gels was $5.5x10^5$ cells which is approximately 55µg of protein. Proteins were separated on pH3-10NL IPG strips and run on 7cm mini-gels. **A.** Method 1 based on the method taken from Lehr et al 2001 and used lysis buffer 3. **B.** Method 2 used lysis buffer 2. **C.** Method 3 used lysis buffer 1. **D.** Method 4 used lysis buffer 2. Image C has the best resolution with spots present across the gel from acidic to basic pH and high to low molecular weight.

	Number of spots detected	Number of spots detected
Sample preparation method	**Replicate 1**	**Replicate 2**
1	122	113
2	103	87
3	117	105
4	137	76

Table 2-7 Number of spots per mini-gel detected using the 2DElite software. The images from replicate 1 can be seen in Figure 2-6 and replicate 2 in Figure 2-7.

A

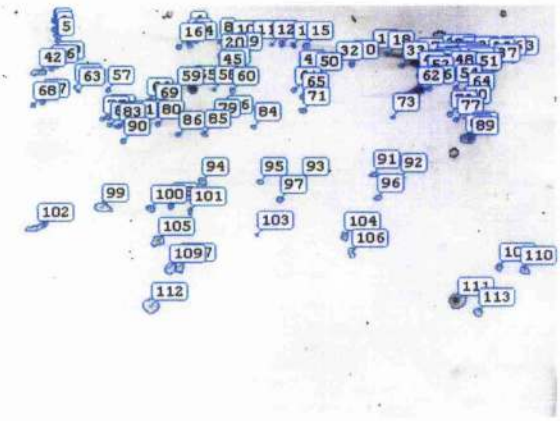
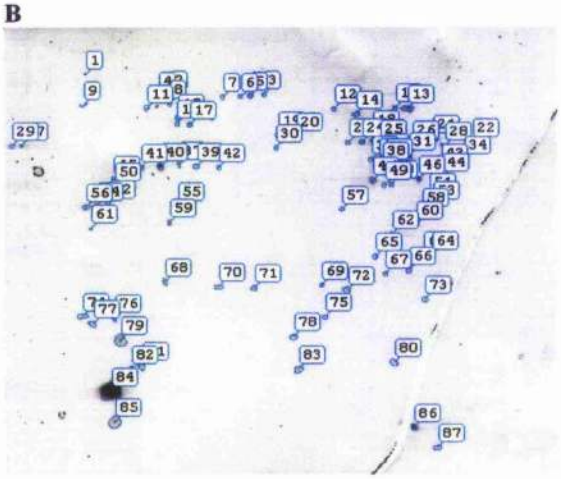


Figure 2-7 A comparison of sample preparation methods 1-4 by 2-DE mini-gels. 30µg of protein were loaded onto pH3-10NL 7cm strips for IEF and then run on a 15% resolving gels. Gels were stained with colloidal Coomassie. The gel images were analysed using ImageMaster 2DElite software.

A Method 1. 113spots detected.
Spots generally clustered towards mid to high molecular weight range.

B

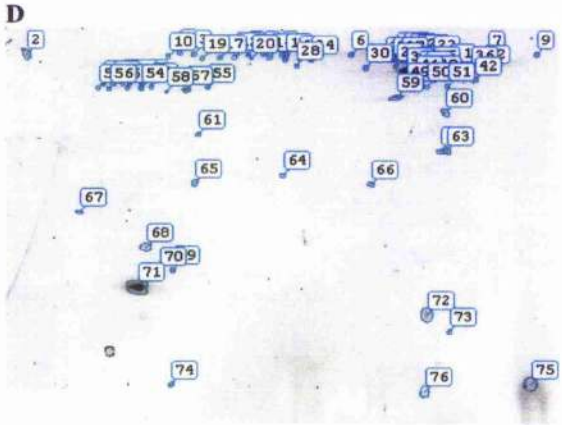


B Method 2. 87 spots detected.
Spots detected across the gel.

C

C Method 3. 105 spots detected.
Spots present across the gel from high to low molecular weight and acidic to basic pH.

D



D Method 4. 76 spots detected.
Most spots present in the high molecular weight range. This gel does not show good spot resolution

2.3.5 Assessment of the reproducibility of sample preparation method 3 and optimisation of the staining technique

To assess the reproducibility of sample preparation method 3, three independent HFF protein extractions were carried out. Samples were separated on 2-DE mini gels. The gels were stained with Coomassie Brilliant blue. The first gel image can be seen in Figure 2-6C. The other two images can be seen in Figure 2-8. All of the spots present on these images were detected using the 2DElite software. Between 100 and 200 spots were detected on each of the three gels.

To determine the optimal staining method, four staining techniques were assessed; colloidal Coomassie, silver staining, SYPRO Orange and SYPRO Ruby. Although silver staining allowed for the detection of many protein spots on the gel, this method of staining is incompatible with protein identification by mass spectrometry therefore the images have not been included.

Images of HFF proteins separated on 24cm pH3-10NL gels and stained with colloidal Coomassie and SYPRO Ruby can be seen in Figure 2-9 and Figure 2-10 respectively.

Colloidal Coomassie is a very sensitive stain and can detect as little as 100ng protein. SYPRO dyes have sensitivity between colloidal Coomassie and silver staining, are compatible with MS and show a wide dynamic range for quantification. In this analysis, colloidal Coomassie stained the proteins more intensely than SYPRO Ruby, was more reproducible and is a much more cost effective stain.

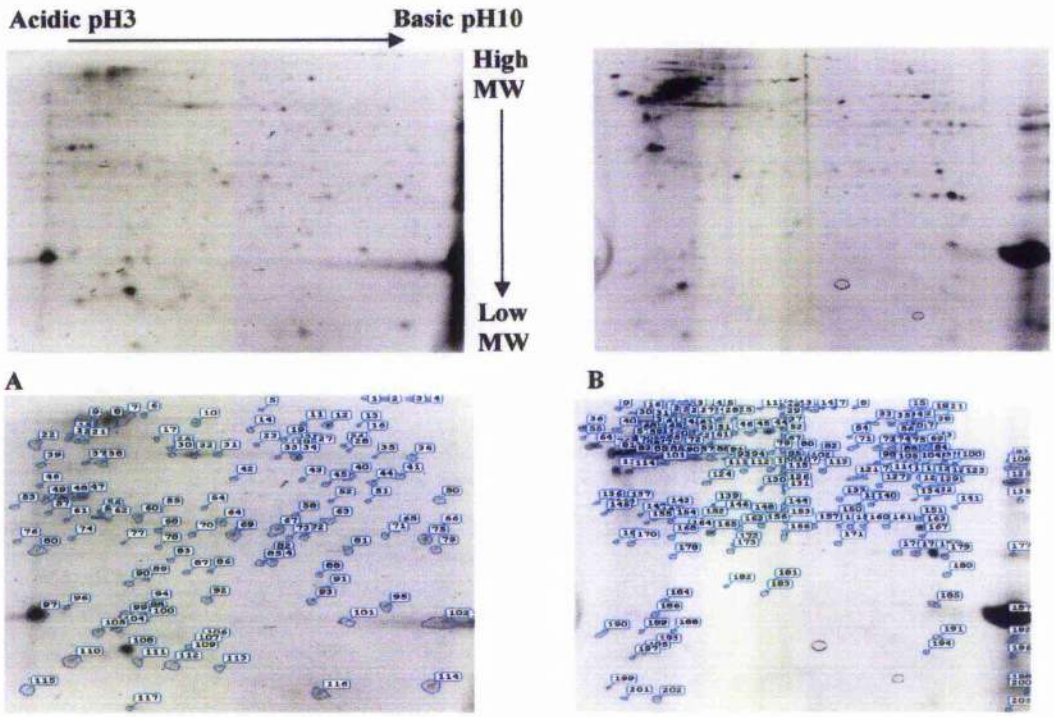


Figure 2-8. Assessment of the reproducibility of sample preparation method 3. Another two 7cm mini-gels (A and B) were run. 55µg of HFF protein was separated by 2-DE using pH3-10NL IPG strips. Gels were stained with Coomassie Brilliant Blue to visualise the protein spots and the images analysed and number of spots detected using the 2DElite software.

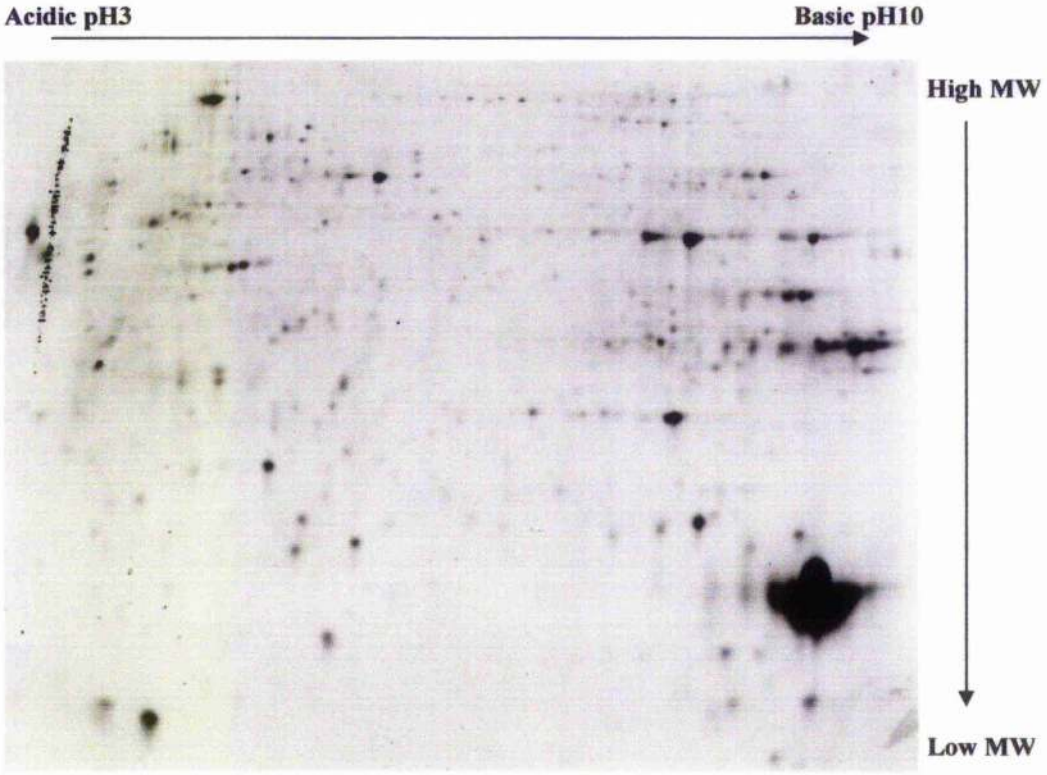


Figure 2-9 A whole proteome map of HFF cells. 300µg of HFF protein prepared using sample preparation method 3 was separated by 2-DE using pH3-10 NL IPG strips. The gels were stained with colloidal Coomassie to visualise the protein spots. This gel shows good spot resolution from acidic to basic pH and high to low molecular weight.

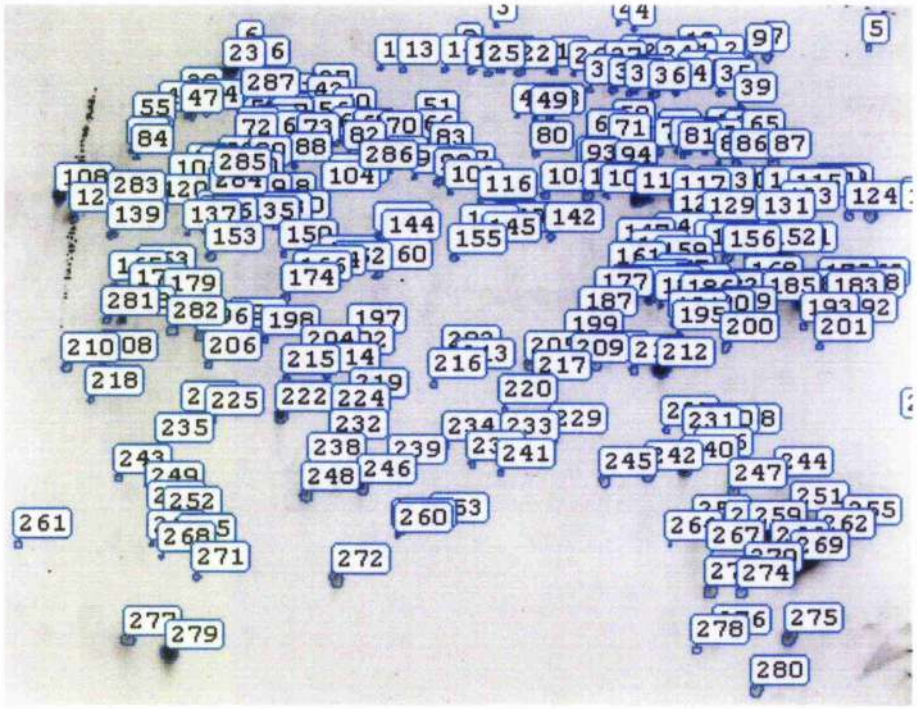


Figure 2-9 B Image analysed using 2D Elite software. The image in A was analysed using the 2DElite software and 280 spots detected.

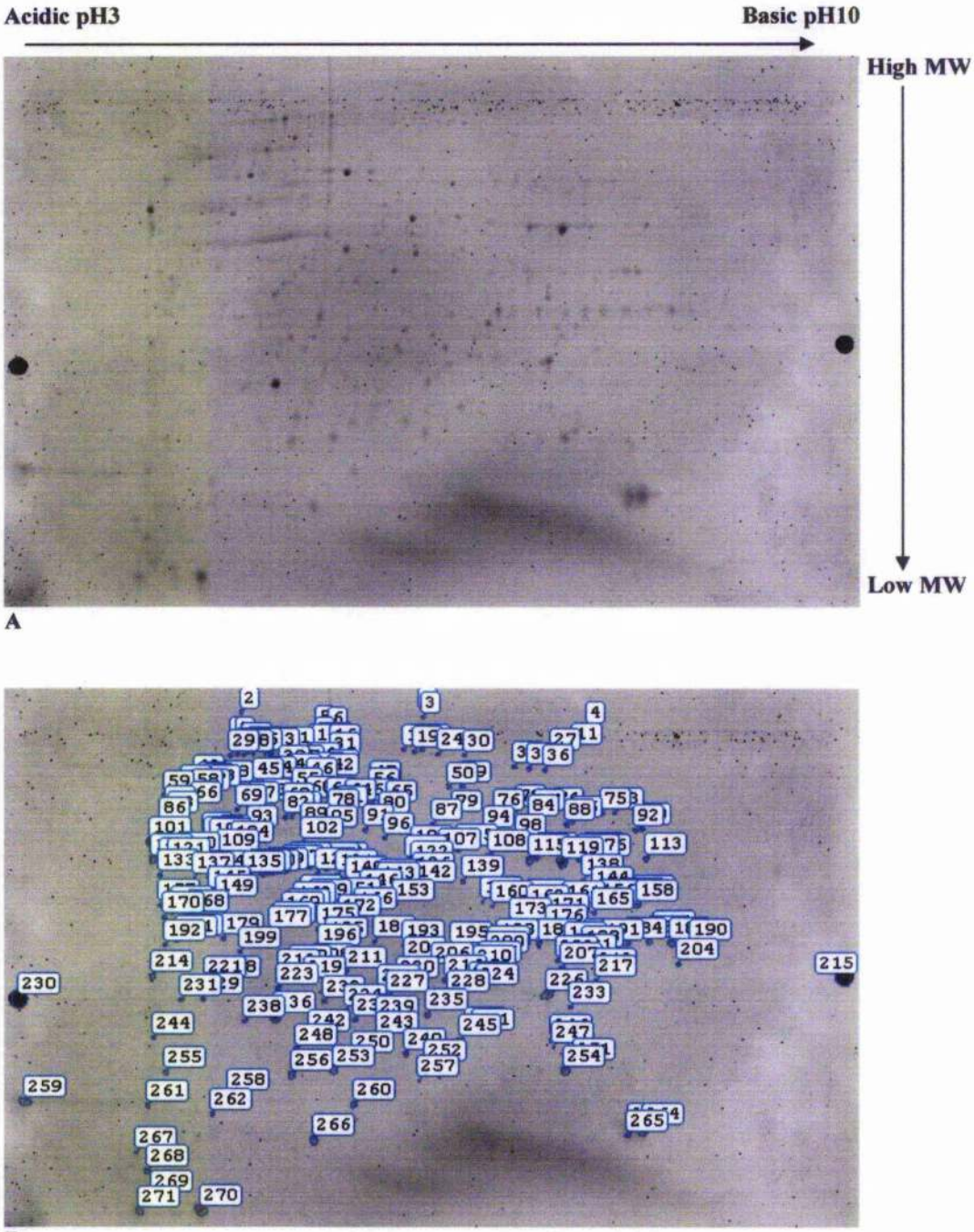


Figure 2-10 Whole proteome map of HFFs visualised using SYPRO Ruby. 300µg of HFF protein was separated by 2DE on pH3-10NL IPG strips and 24cm pre-cast gels. Gels were stained with SYPRO Ruby to visualise the protein spots (A). The gel shows good resolution of spots, very little streaking present, however spot intensity is not high relative to background. Using the 2DElite software, 270 protein spots were detected on this gel (B).

2.3.5.1 Narrow range pII strips

HFF protein extracts were separated on narrow pH range strips to increase the range of proteins seen within that pH gradient (Figure 2-11). pH 3.5-4.5 and 4.5-5.6 gives increased resolution of proteins which are acidic. pH 6-9 strips focused on the alkaline proteins. In this experiment in the pH range 3.5-4.5 no spots were visible, this could be due to the fact that not many proteins lie within this pH range. pH 4-7 gave good coverage of proteins and pH 4.5-5.5 and 5.5-6.7 gave increased resolution of the proteins present within these pH ranges that would maybe not normally be seen in such detail on a pH 3-10 or pH 4-7 gel.

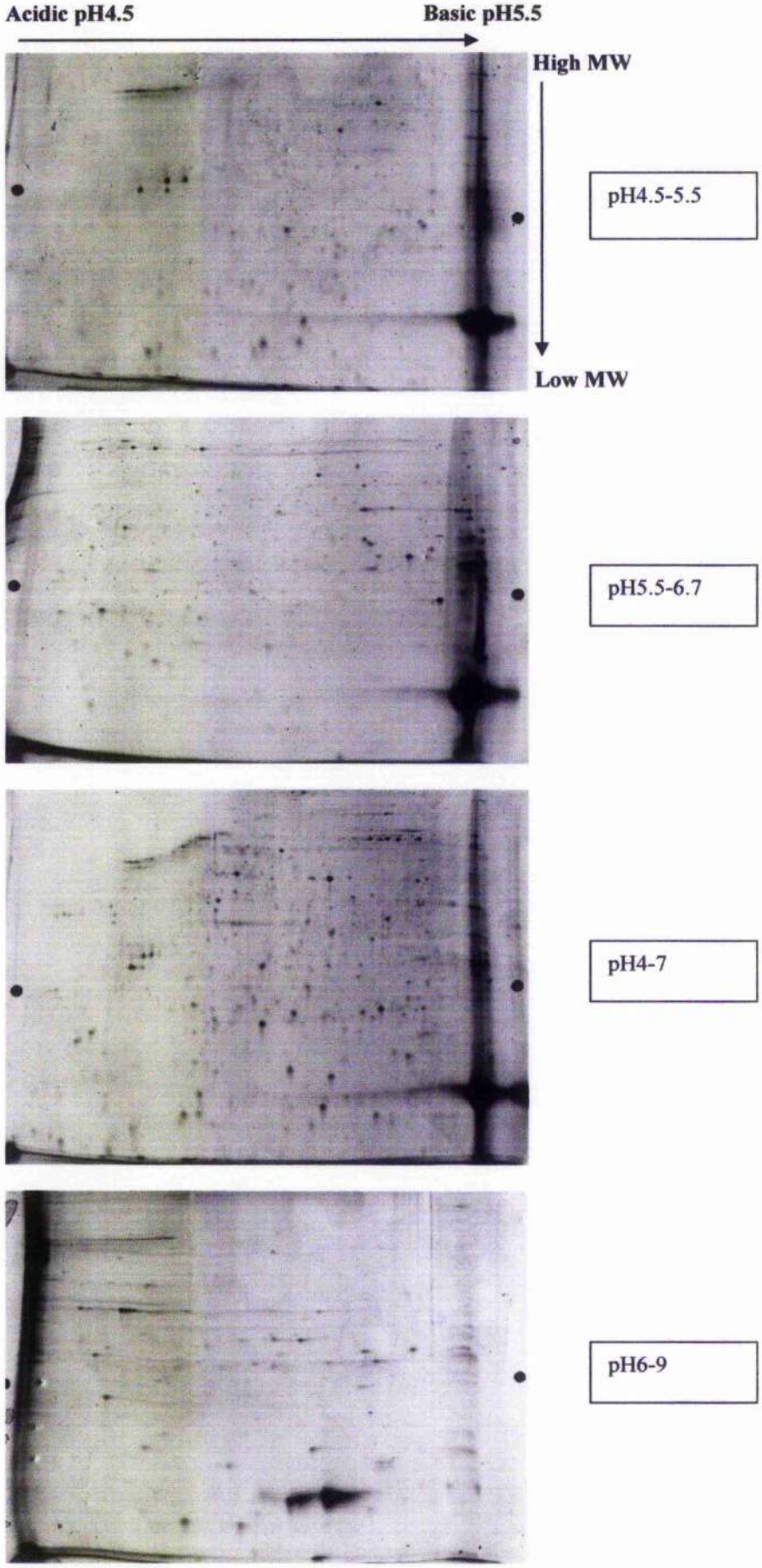


Figure 2-11 Comparison of proteome maps using narrow pH ranges. 300µg of protein was separated on different pH range strips (pH4.5-5.5, pH5.5-6.7, pH 4-7, pH 6-9) and 24cm pre-cast gels. Gels were stained with colloidal Coomassie to visualise the protein spots. pH5.5-6.7 and particularly pH 4-7 give good resolution of spots.

2.3.6 Preliminary HFF proteome map

An HFF proteome map was generated using method 3 for sample preparation and colloidal Coomassie stain. Figure 2-9 shows the 2-DE pattern of HFFs. HFF protein (300μg) was separated on a pH3-10NL IPG strip and 24cm pre-cast gel. The gel was stained with colloidal Coomassie and the spots detected using the 2DElite software. Spots circled in blue were picked using the Amersham Biosciences robotic spotpicker and analysed using MALDI. Mass spectra were acquired for each spot and a PMF produced, an example of which is shown in Figure 2-12 . The monoisotopic mass list obtained from the PMF was used to search the computer database producing a histogram (Figure 2-13) from which a significant score can be distinguished from an insignificant score. Protein scores greater than 64 are significant ($p<0.05$). The MOWSE cut off for a significant score differs between databases and per MS method. Figure 2-13 shows a histogram of the protein score distribution. The fifty best matching proteins are divided into sixteen groups according to their score, and the heights of the bars show the number of matches in each group. In the case of a peptide mass fingerprint, the protein score is a measure of the statistical significance of a match. The region in which random matches may be expected is shaded green. This region extends up to the significance threshold, which has a default setting of 5%. If a score falls in the green shaded area, there is greater than a 5% probability that the match was a random event and is of no significance. Conversely, a match in the unshaded part of the histogram has less than a 5% probability of being a random event. It is quite common to see several proteins getting the same high score. Even if the protein sequences in the database are non-identical, the same group of matched peptides may occur in multiple proteins (MASCOT web page). Proteins which had scores close to the significance cut off point were re-checked to determine whether it was an actual result. The mass and pI of the protein was checked against the spot location on the 2-DE gel and the number of peptides matched and the percentage sequence coverage was compared. The MS data was searched against the homosapiens NBCI database. The NCBInr database contains the translated protein sequences from the entire collection of annotated DNA sequences kept at GenBank, and also protein sequences in the Protein Data Bank, SWISS-PROT and Protein Information Resource databases, covering most of the publicly available data.

Identified proteins were annotated with the protein name and as can be seen from the proteome map in Figure 2-14, there is a good coverage of identified proteins across the gel showing that the MS technique used was able to successfully identify highly acidic and

basic proteins and also low and high molecular weight proteins. Unidentified spots circled in blue were subjected to MALDI analysis however were not identified, perhaps due to insufficient concentration of protein in the sample or more than one protein present in the spot. Table 2-8 contains the information about the identified proteins.

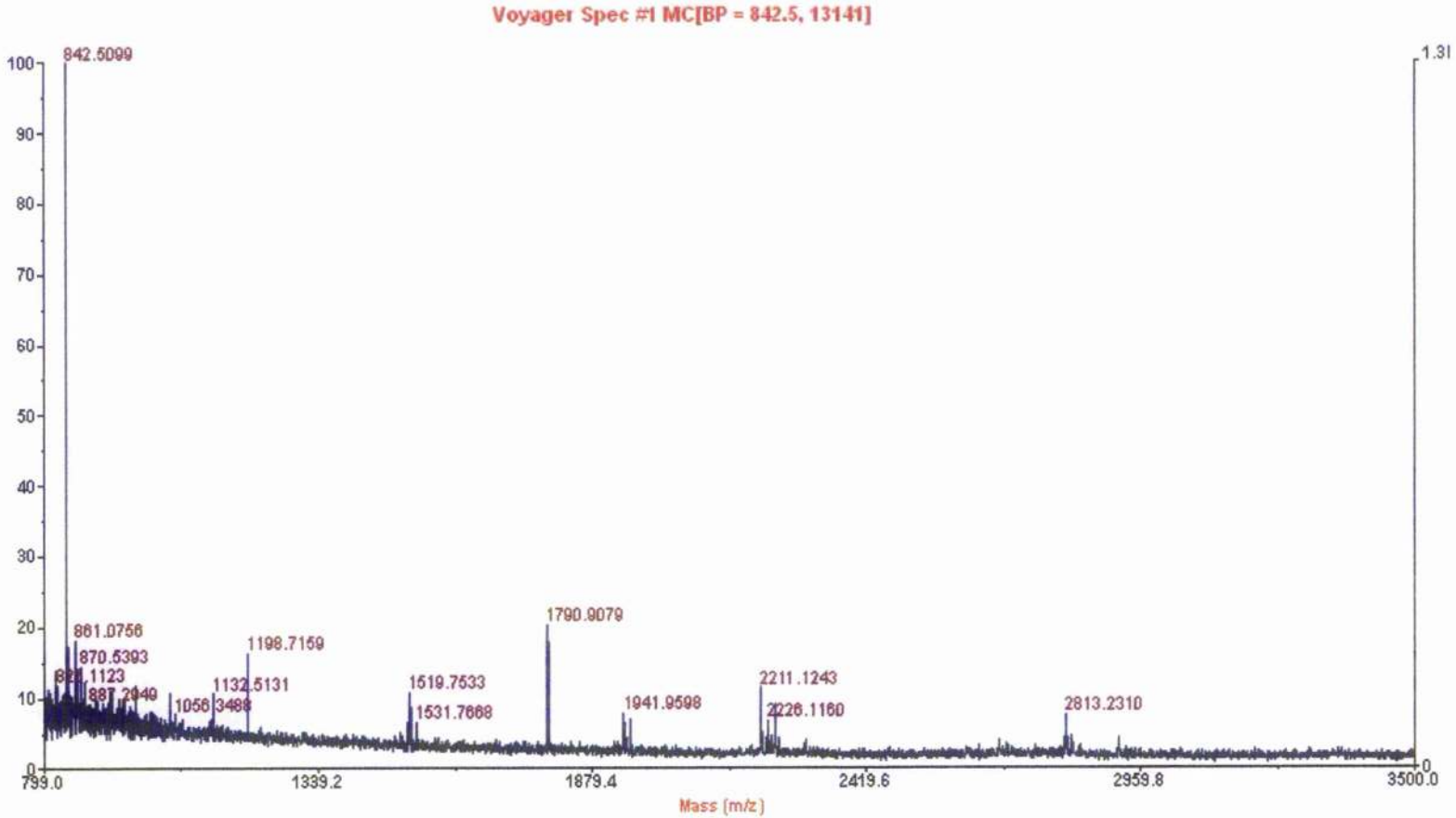


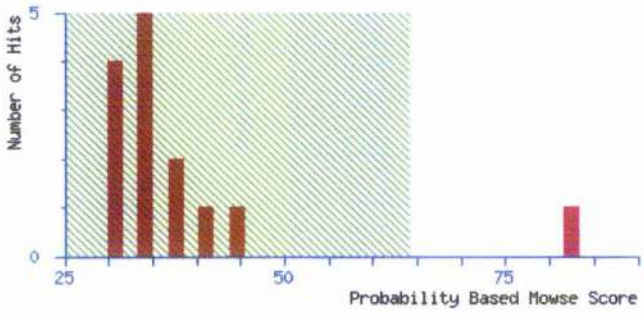
Figure 2-12 Example of a MALDI mass spectrum. The peptide mass fingerprint was derived from the in-gel digestion of a spot of interest. The typical spectrum was obtained by MALDI-MS in reflectron mode using α-cyano-4-hydroxycinnamic acid as the matrix. One hundred and fifty laser shots were acquired and the resultant peptide mass fingerprint shown above was internally calibrated using the porcine trypsin autolysis peptide peaks as highlighted at 842.5009 and 2211.1243 Da. This protein was identified as actin beta.

Mascot Search Results

```
User            : Morag
Email           : 9606340n@student.gla.ac.uk
Search title    : 11_11_03Plate1_spot89
MS data file    : C:\morag\11_11_03plate1\11_11_03plat31_89_0001.dat
Database        : NCBInr 20060331 (3525863 sequences; 1211011241 residues)
Taxonomy        : Homo sapiens (human) (145946 sequences)
Timestamp       : 13 Apr 2006 at 10:12:07 GMT
Top Score       : 82 for gi|15277503, ACTB protein [Homo sapiens]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 64 are significant (p<0.05).



Concise Protein Summary Report

Format As | Concise Protein Summary | Help

Significance threshold p< 0.05 Max. number of hits 20

Re-Search All | Search Unmatched

1. gi|15277503 **Mass:** 40536 **Score:** 82 **Expect:** 0.00082
ACTB protein [Homo sapiens]

Figure 2-13 Score histogram for a typical peptide mass fingerprint. Mascot incorporates a probability based implementation of the MOWSE algorithm (Perkins et al, 1999). Only scores out with the green region are significant. Reproduced from http://www.mascot.matrixscience.com/

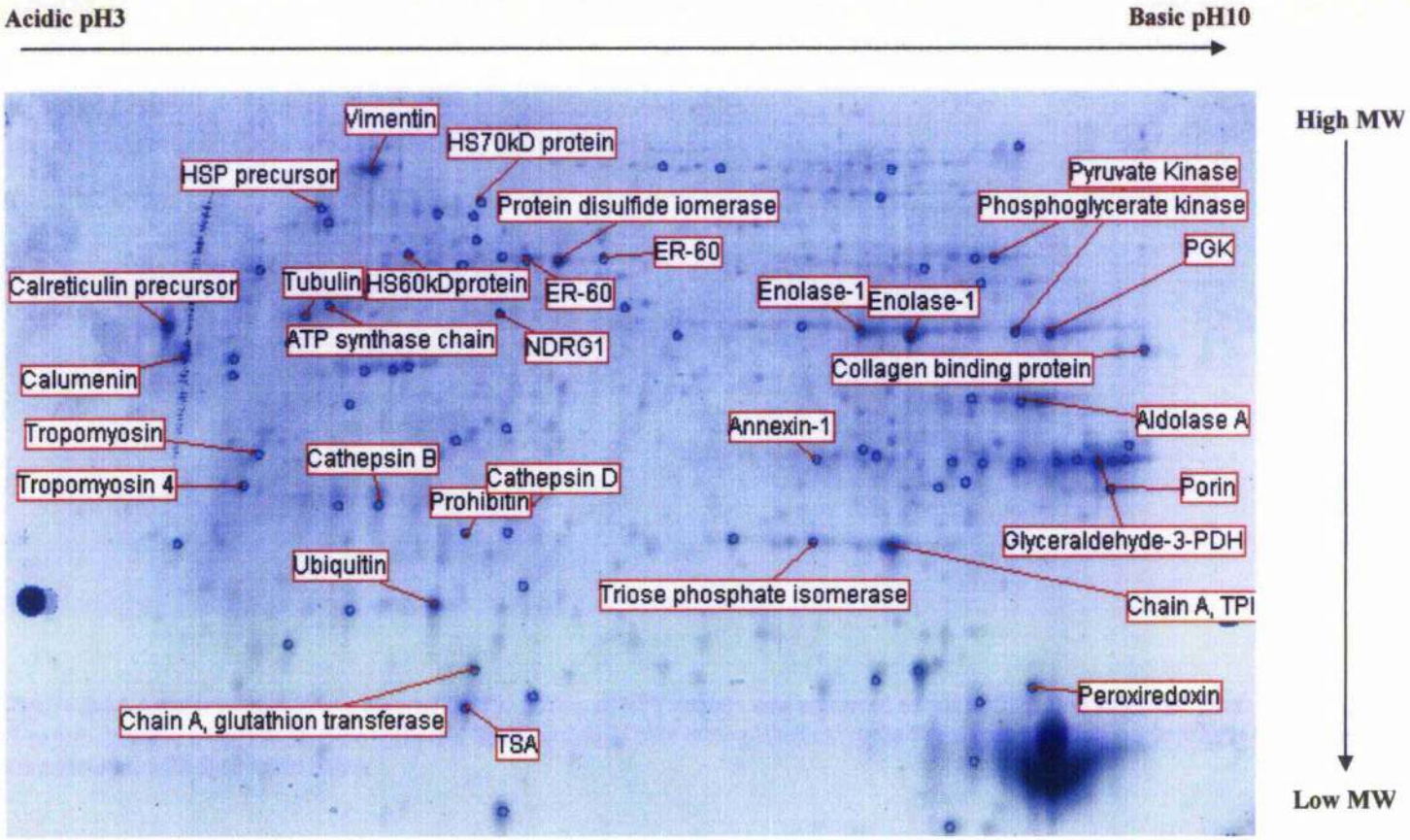


Figure 2-14 Annotated proteome map of HFFs. 300µg of HFF protein was separated on pH3-10NL IPG strips and 24cm pre-cast gels. The gel was stained with colloidal Coomassie and the spots detected using the 2DElite software. Spots circled in blue were picked using the Ettan spot handling workstation (Amersham) and analysed using MALDI MS. Identified proteins are annotated with the protein name.

Low MW

Spot number	Score	Mass (kDa)/pI	PM/s.c	Protein
4	348	53711/5.06	30/61%	Vimentin
7	101	74019/5.97	9/16%	Heat Shock 70kD protein (MTHSP75)
8	199	71002/5.23	12/24%	dnaK-type molecular chaperone HSPAS precursor
9	144	53598/5.62	13/35%	HSP 70kD protein 8 isoform1
10	133	41973/6.69	8/24%	HS70kDa protein. 42kd Alpha N-terminal domain
11	212	72185/5.03	19/32%	dnaK type molecular chaperone HSPAS precursor
13	143	59946/5.7	9/23%	HS60kDa protein 1 (chaperonin)
15	97	57043/6.1	7/18%	Protein-disulfide isomerase
17	123	58411/7.58	12/26%	Pyruvate kinase
18	152	57043/6.1	12/23%	Protein-disulfide isomerase
19	173	57043/6.1	15/28%	Protein-disulfide isomerase
26	124	56525/5.26	9/26%	ATP synthase β chain, mitochondrial precursor
27	84	49484/5.89	7/24%	Heterogeneous nuclear ribonucleoprotein H1
28	90	43264/5.49	8/27%	n-myc downstream regulated gene 1 protein.
29	113	50096/4.75	15/32%	Tubulin, β5
30	123	48283/4.29	11/42%	Calreticulin precursor; sicca syndrome antigen A
32	81	47481/7.01	10/26%	Enolase-1
33	100	44985/8.30	6/22%	Phosphoglycerate kinase 1
34	79	44985/8.30	6/22%	Phosphoglycerate kinase 1
35	118	47481/7.01	13/36%	Enolase-1
38	96	46525/8.75	7/22%	Proliferation inducing gene
39	69	37918/4.47	6/31%	Calumenin precursor
41	73	406536/5.55	6/21%	Actin
42	73	406536/5.55	6/21%	Actin
43	73	406536/5.55	6/21%	Actin
46	92	39720/8.39	9/29%	Aldolase A
50	77	36041/8.67	7/32%	Heterogeneous nuclear ribonucleoprotein As/B1
51	99	34166/7.8	6/25%	4and a half LIM domains 2
52	65	29109/4.79	5/14%	TPMsk3 (tropomyosin)
54	71	38918/6.57	6/26%	Annexin1 (lipocortin1)
55	64	35698/8.22	4/13%	Uracil DNA glycosylase
57	74	36202/8.26	6/23%	Glyceraldehyde-3-phophate dehydrogenase
58	74	36202/8.26	5/18%	Glyceraldehyde-3-phophate dehydrogenase
61	68	35698/8.22	5/19%	Uracil DNA glycosylase
62	67	36202/8.26	6/21%	Glyceraldehyde-3-phophate dehydrogenase
64	104	28619/4.67	10/35%	Tropomyosin 4
66	114	30737/8.63	7/40%	Porin 31HM
67	67	38722/5.77	8/30%	Cathepsin B
69	108	26457/-	8/-	Chain B, Cathepsin D
70	64	30952/5.69	4/11%	F-actin capping protein b unit; CapZ
71	93	29843/5.57	6/28%	Prohibitin
73	144	26876.51	8/44%	Triose phosphate isomerase
75	220	26807/6.51	14/79%	Chain A, Triose phosphate isomerase
78	74	25151/5.33	6/38%	Ubiquitin carboxyl terminal esterase L1
81	108	23394/5.74	6/48%	Chain A, glutathione transferase P1-1
85	85	22324/8.27	6/35%	Peroxiredoxin 1
87	82	16068/8.81	4/39%	Neuropolypeptide ti3 brain fragment
88	74	18486/5.19	4/32%	TSA
91	68	-/-	-/-	Peptidylprolyl isomerase A (cyclophilin A)

Table 2-8 Results of proteins identified on HFF map. Protein spots were analysed by MALDI. This table shows the spot number, the MOWSE score from MASCOT, the mass and pI of the protein, the number of peptides matched and sequence coverage (PM/s.c) and the name of the protein identified.

2.4 Discussion

Optimisation of sample preparation and staining led to a robust, reproducible method to profile the HFF proteome using 2-DE. Sample preparation is one of the key steps in 2-DE. Due to the presence of hydrophobic proteins, proteins with differing pI and MW and proteins present in low copy numbers that are masked by house keeping proteins, profiling an entire proteome using 2-DE is as yet not possible. Complete cell lysis should result in solubilisation, denaturation, reduction and disaggregation of all the proteins in the sample. The sample is solubilised and the proteins denatured into their polypeptide subunits using a cocktail of solubilising reagents, nucleases and protease inhibitors. The method of sample preparation and cell lysis must be optimised for the sample of interest. Inclusion of surfactants, reductants and chaotropes are essential in order to obtain a high level of protein extraction in a format compatible with 2-DE. However, some proteins are inherently difficult to solubilise, e.g. membrane proteins, and specialised solubilisation techniques are required.

The main aim of the work described in this chapter was to develop a reproducible and sensitive method of sample preparation for HFFs for use with 2-DE. The method of 2-DE used in this chapter resolved 280 spots on a 24cm gel.

In order to prepare samples suitable for 2-DE, four methods of protein extraction were analysed in regard to their ability to produce proteins of high resolution and solubility, and to have good representation of acidic through to basic proteins, and high to low molecular weight proteins. The method found to be most successful was sample preparation method 3 which contained urea (chaotrope), CHAPS (surfactant), DTT (reducing agent), protease inhibitors and nucleases. This method was found to be highly reproducible. The method determining protein concentration that was found to be most compatible with this sample preparation method was the Bio-Rad Protein Assay. This assay gave consistent reproducibility and is very easy to use. Colloidal Coomassie was found to be the most suitable stain due to its high sensitivity (sensitive down to 100ng/protein spot), compatibility with MS, linearity and low cost. pH3-10NL and pH4-7 gave the best coverage of proteins, and the narrower pH ranges could be useful to magnify into a pH of interest.

A proteome map of HFF cells was produced. Of the 91 spots picked, 49 were identified using MALDI. The map showed good coverage of proteins from high to low molecular weight and acidic to basic pH. MALDI allowed for identification of 54% of these spots and

several isoforms of the same protein were identified at different locations on the gel e.g. endoplasmic reticulum protein ER-60. Comparing this HFF proteome map to other published fibroblast proteome maps highlighted several similarities in protein location on the 2-DE gel. An example of this is the TIG3 human fibroblast line (http://proteome.tmig.or.jp/2D/Fibro/humfb_tags.html). Even though a different type of fibroblast was profiled, many of the proteins identified were similar between the TIG3 line and this HFF study, including calreticulin, HSPs and tropomyosin.

Further analysis of the unidentified spots using MS/MS would have led to an increased percentage of identified proteins. However, due to the time and cost necessary for this type of analysis it was decided not to carry out any MS/MS work on this experiment. MALDI enabled us to build a preliminary picture of the protein map of HFF cells and proved that this type of analysis would allow us to identify proteins of acidic to basic pH and high to low molecular weight. The preparation techniques optimised in this chapter were then used for all subsequent 2-DE work discussed in the rest of the thesis.

Several thousand proteins can be resolved on a 2-DE gel (Wu *et al.*, 2002). The estimated number of proteins in a human cell is thought to be around 30,000 (Jenkins and Pennington, 2001). Clearly 2-DE is only able to profile a small percentage of the total proteins present. However, this technique provides a platform to carry out hypothesis forming research instead of investigating a single protein of interest. Using 2-DE it is now possible to investigate the behaviour of several hundred or thousand protein species under a certain condition (e.g. disease vs. healthy).

In order to profile a greater percentage of the proteins present in HFFs, several approaches are possible. The use of narrow range pH strips would help to increase protein coverage. Often spots on a pH3-10 or pH4-7 gel are mixed, e.g. more than one protein species is present due to similar mass and charge. Increased pH resolution would help to separate these mixes and profile more proteins. Subcellular prefractionation or purification (e.g. SIGMA membrane protein purification kit) prior to 2-DE would also help to increase the number of proteins profiled. 2-DE is a poor technique to resolve membrane proteins, hydrophobic or high molecular weight proteins (Rabilloud, 2002). A modified 2-DE method termed benzyl hexadecyl ammonium chloride (16-Bac) has recently been shown to be very successful in profiling membrane proteins. 16-Bac separates proteins based on MW in the 1st dimension and the anionic detergent, SDS, separates proteins based on MW in the 2nd dimension. The acidic pH of the cationic detergent (16-Bac) has a denaturing

effect on the proteins and is therefore particularly suited for membrane proteins (Kramer, 2006).

The generation of a reproducible and high quality HFF proteome map showed that the sample preparation method was optimal for this cell type. The next step was to investigate what changes occur in the HFF proteome under a different biological condition. This dynamic model system presents a unique opportunity to investigate in detail the protein expression of parasite infected cells.

Chapter III Proteomic analysis of *Toxoplasma gondii* infected cells

3.1 Introduction

The aim of the work in this chapter was to profile the changes in protein expression that occur in cells during infection with *Toxoplasma gondii.* This in turn will lead to a better understanding of the host response to infection. Proteins from infected and non-infected cells were separated and visualized using 2-DE and the differentially expressed proteins identified using a combination of MALDI MS and MS/MS. Infection ratios and time-points were based on the study by Blader *et al* (2001) allowing for close comparison of gene and protein expression results which is discussed in Chapter V. Several time points post infection were investigated with the emphasis on 24 hours post infection (hpi). This infection time point was also studied extensively in the transcriptional studies (Blader *et al.*, 2001).

The gene expression of host cells is modulated significantly during infection with *Toxoplasma gondii.* Microarray analysis suggests that several host cell pathways are altered due to the presence of the parasite, in particular the up regulation of the glycolytic enzymes and modulation of the apoptotic machinery (Blader *et al.*, 2001). To what extent these transcriptional changes in the *Toxoplasma gondii* infected host cell relate to protein expression and therefore the actual cell behaviour remain unknown as the gene to protein correlation is still uncertain. Several studies claim there to be a very poor correlation between mRNA and protein expression (Gygi *et al.*, 1999; Lehr *et al.*, 2002) whereas others (Sharabiani *et al.*, 2005) conclude there is a strong level of correlation between protein and gene expression. The use of proteomic techniques will allow for direct comparison of gene and protein expression and reveal what level of correlation between mRNA and protein expression exists. Also possible is the identification of post translational modifications (PTM) in the infected cell which cannot be investigated using microarrays. The proteomic techniques will therefore potentially highlight novel proteins or pathways involved in the host infection response.

Several studies have used 2-DE for comparative proteomic analysis (Jungblut *et al.*, 1999; Lian *et al.*, 2002; Ruiz-Romero *et al.*, 2005; Zhang *et al.*, 2005) in different cell types looking at a variety of types of comparisons i.e. disease/stress/aging/strains. To date, there

are no published studies which use proteomics to investigate the host response to infection with an apicomplexan parasite.

Human foetal foreskin fibroblasts (HFF) were chosen as the cell model to investigate the host-pathogen response due to the fact they are representative of many cell types, easy to culture and infect with *T.gondii* and are the same cell type as used in the transcriptional studies allowing for ease of comparison of gene to protein (Chapter V).

Alongside MALDI MS, a second mass spectrometric technique was used in this Chapter; MS/MS. This type of MS fragments peptides into their actual peptide ions therefore providing sequence data. The fragment ions obtained by this method are derived from the amino or carboxyl terminus of the peptide and are designated b or y ions respectively. The set of fragment ions generated by MS/MS act as a fingerprint for an individual peptide. The matching of observed b or y ions against the theoretical (expected) fragment ion masses using the MASCOT® database allows for amino acid identification by mass. This technique is a great deal more sensitive than MALDI but is more time consuming and therefore lower throughput.

In this chapter the host response to infection with *Toxoplasma gondii* was investigated using 2-DE combined with MS. Several hundred differentially expressed proteins were identified at several time points post infection. Clustering these changes and assigning function (Chapter VIII) will lead to a better understanding of the host response to this pathogen and highlight novel proteins and pathways involved in the host response to infection.

3.2 Materials and Methods

3.2.1 Passage of Vero cells

Vero cells were maintained in continuous culture in IMDM with L-glutamine and 25mM HEPES +5%FCS + 1%Pen/Strep at 37°C, 5% CO_2 and passaged every seven days.

3.2.2 Parasites

Toxoplasma gondii tachyzoites (isolate SRH) were cloned in Sweden by Jens Mattsson. Parasites were maintained in Vero cells for use in the HFF infection experiments. One day old Vero cell cultures were infected with *T.gondii* ($8x10^4$parasites/cm^2) and the parasites left to replicate for 3-4 days. The parasites were then harvested and used to infect fresh 1 day old Vero cells.

3.2.3 HFF infection with Toxoplasma gondii

T.gondii tachyzoites were grown in HFFs (passage < 20). Parasites were harvested from infected HFF monolayers when lysis by infection was almost complete and few intact host cells remained. The flasks were then scraped, and the entire material harvested and passed through a 3µm filter to rupture intact HFFs and release any remaining intracellular parasites. The resulting material was pelleted at 2,900 g, washed in DMEM, and parasites counted by light microscopy.

For the HFF infection experiments parasites were added to confluent HFF monolayers (7-10 days old, passage ≤20) at various infection ratios (Table 3-1). The infection experiments were carried out in $75cm^2$ or $175cm^2$ flasks with 15 or 45ml of DMEM respectively for 6, 12, 16 or 24h. The cells were then harvested as described in Chapter II. The infected cells were harvested first and the number of cells and extracellular parasites counted. The material was then pelleted by centrifugation and washed three times in PBS. After the third wash and spin, all PBS was removed and the pellet frozen at -70°C.

The non-infected cells were then harvested and counted and an equal number of extracellular parasites as counted in the infected cell sample were added to the non-

infected control. The material was then pelleted by centrifugation and washed three times in PBS. After the third wash and spin, all PBS was removed and the pellet frozen at -70°C.

3.2.4 Giemsa staining of infected cells

To investigate the optimal ratio of infection, HFFs (passage ≤20) were grown in 1cm² chamber slides (Lab-tek, Nalge Nunc International) in 500µl DMEM + 10% FCS, Pen/Strep and L-glutamine) and once confluent, infected with *T.gondii* tachyzoites at various infection ratios (2, 5 and 10 parasites: cell) and left at 37°C, 5% CO_2 for 24h. 24hpi the cells were stained with Giemsa (as described in Chapter II). The fixed cells and parasites were then viewed and the images captured using a Zeiss Axiovert 200M light microscope at 200x and 400x magnification.

24h infection experiment	Infection ratio	pH range	used for analysis
1	10 parasites: 1 cell	pH 3-10NL	Yes. Comparative analysis and protein identification
2	10 parasites: 1 cell	pH 3-10NL	Yes. Comparative analysis and protein identification
3	1 parasite: 1 cell	pH 3-10NL	Yes. Comparative analysis and protein identification
4	1 parasite: 1 cell	pH 3-10NL	Yes. Comparative analysis and protein identification
5	1 parasite: 1 cell	pH 3-10NL	Yes. Comparative analysis and protein identification
6	1 parasite: 1 cell	pH 3-10NL	Yes. Comparative analysis and protein identification
7	1 parasite: 1 cell	pH 3-10NL	Yes. Comparative analysis and protein identification
8	2 parasites: 1 cell	pH 3-10NL	Used for comparison, not picked from
9	2 parasites: 1 cell	pH 3-10NL	Used for comparison, not picked from
10	2 parasites: 1 cell	pH 4-7	Yes. Comparative analysis but no proteins identified
11	2 parasites: 1 cell	pH 4-7	Yes. Comparative analysis and protein identification

16h infection experiment	Infection ratio	pH range	used for analysis
1	10 parasites: 1 cell	pH 4-7	Yes. Comparative analysis and protein identification
2	10 parasites: 1 cell	pH 4-7	used for phosphorylation (PP) experiment, chapter 7
3	10 parasites: 1 cell	pH 4-7	Used for PP experiment, chapter 7
4	10 parasites: 1 cell	pH 4-7	Used for PP experiment, chapter 7
5	10 parasites: 1 cell	pH 4-7	Used for PP experiment, chapter 7

12h infection experiment	Infection ratio	pH range	used for analysis
1	5 parasites: 1 cell	pH4-7	Difficult to compare between gels
2	5 parasites: 1 cell	pH4-7	Difficult to compare between gels
3	10 parasites: 1 cell	pH4-7	Difficult to compare between gels
4	10 parasites: 1 cell	pH4-7	Difficult to compare between gels
5	10 parasites: 1 cell	pH4-7	Yes. Comparative analysis and protein identification
6	10 parasites: 1 cell	pH4-7	Yes. Comparative analysis and protein identification

6h infection experiment	Infection ratio	pH range	used for analysis
1	10 parasites: 1 cell	pH4-7	Difficult to compare between gels
2	10 parasites: 1 cell	pH4-7	Yes. Comparative analysis and protein identification
3	10 parasites: 1 cell	pH4-7	Difficult to compare between gels

Table 3-1 Time course details and infection ratios in the *T.gondii* infection experiments. This table shows which infection ratio was used at which infection time point. It also lists the pH range of the IPG strip used in each 2-DE experiment. The final column 'used for analysis' indicates whether spots from the gel were analysed using MS (yes), whether the gels were used for comparison and therefore a replicate (for comparison), or whether it was not possible to compare the opposite gels due to poor quality of gel matching (difficult to compare between gels). Time courses were carried out to investigate the protein expression response of the host cell during the course of infection. Several infection ratios were investigated to find the optimum infection ratio of parasite: cell.

3.2.5 Time course infection experiments

Eleven replicates of a 24h infection experiment, five replicates of a 16h infection experiment, six replicates of a 12h infection experiment and three replicates of a 6h invasion experiment were carried out. 24h replicate 11 will be discussed in detail in this chapter as an example of the experimental and analytical procedures used.

3.2.5.1 24h Infection experiment 11

In the 24h infection experiment replicate 11, cells were infected at a ratio of 2 parasites: 1 host cell (HFFs 9 days old, passage number 20). The infected cells were harvested, the cells and parasites counted and the parasite: cell ratio calculated. The non-infected cells were then harvested and the same percentage of parasites as seen in the infected sample was added to the non-infected cell sample.

To each sample 80µl lysis buffer 1, 10µl nuclease mix and 5µl protease inhibitors were added and lysis method 3 (Chapter II) was followed. The Bradford protein assay determined the protein concentration of the non-infected cells to be 6.72µg/µl (591µg in total) and the infected sample, 2.88µg/µl (254µg in total). Perhaps the lower amount of protein recovered from the infected cell flask reflects loss of host material due to infection with and lysis by *Toxoplasma*.

Equal amounts of protein (254µg) were run on pH4-7 IPG strips following the 2-DE protocol detailed in Chapter II. The IPG strips were run to 77950VhrT. The gels were stained with colloidal Coomassie and the gel images attained using the AmershamPharmacia Biotech Labscan. The gels were analysed using the Image Master 2DElite software.

3.2.6 Image Master 2DElite analysis

Replicates 1-7 were analysed by eye by comparing the infected and non-infected gel images. Any spots present on one gel and not on the other were chosen for MS analysis. In replicate 11, all spots on the two gels were detected using the Image Master 2DElite software and the volume of each spot calculated, subject to background subtraction and normalisation. Spots present on both gels were then matched manually and the volume difference calculated. Any spots present on one gel and not on the other were chosen for

MS analysis. Spots present on both gels with a volume difference of 30% or more were chosen for MS analysis.

3.2.7 Protein identification

Spots of interest were excised from the infected and non-infected gels as 1.4mm plugs and subjected to in-gel trypsin digestion in an Ettan Spot Handling Workstation (Amersham Biosciences) following the standard protocol. The gel plugs were washed and the proteins digested into their component peptides using the enzyme trypsin. Tryptic peptides were extracted from gel plugs and an aliquot (10%) was mixed with α-cyano hydroxyl cinnamic acid (CHCA) and then analysed by MALDI. In replicate 11 any spots that were not identified using MALDI were then analysed by MS/MS. For MS/MS the tryptic peptides were solubilised in 0.5% formic acid and were fractionated by nanoflow high-performance liquid chromatography on a C18 reverse phase column (Amersham Biosciences) and elution was performed with a continuous linear gradient of 40% acetonitrile for 20min. The eluates were analysed by online electrospray MS/MS by use of a QStar Pulsar mass spectrometer (Applied Biosystems). A 3sec survey scan preceded each MS/MS data-collection cycle of 4 product ion scans of 3sec each, and this gave a duty cycle of 115sec.

The set of fragment ions generated by MS/MS act as a fingerprint for an individual peptide. Two or more peptides identified in this way are usually sufficient to unambiguously identify a protein. In the collision induced dissociation (CID) process, peptides fragment in a semi-predictable manner; thus, sequences from the database can be used to predict an expected fragmentation pattern and match the expected pattern to that observed in the acquired data.

The resulting MS/MS spectra were interpreted using the MASCOT® software by querying the present non-redundant National Centre for Biotechnology Information (NCBI) and a locally installed version of the 10x *Toxoplasma* database (ToxoDB, www.toxodb.org/). Partial carboxymethylation and oxidation of methionine residues were considered in the search.

3.3 Results

3.3.1 Optimal Infection Ratio

Determining the optimal infection ratio was extremely important. If the infection ratio was too low (i.e. few cells infected) then any changes in protein expression would be too insignificant to be detected using the proteomic techniques. By contrast, too high a rate of infection would result in cell lysis and loss of host cell material.

To determine the optimal ratio of infection, HFFs (passage ≤20) were grown in chamber slides in 500µl DMEM and once confluent, infected with *T.gondii* tachyzoites at various infection ratios and left at 37°C, 5% CO_2 for 24h. After 24h the cells were stained with Giemsa. The fixed cells and parasites were then viewed using a Zeiss Axiovert 200M light microscope at 200x and 400x magnification and the images captured.

Figure 3-1 A-I shows images of non-infected HFFs and HFFs infected with different infection ratios of *Toxoplasma gondii*.

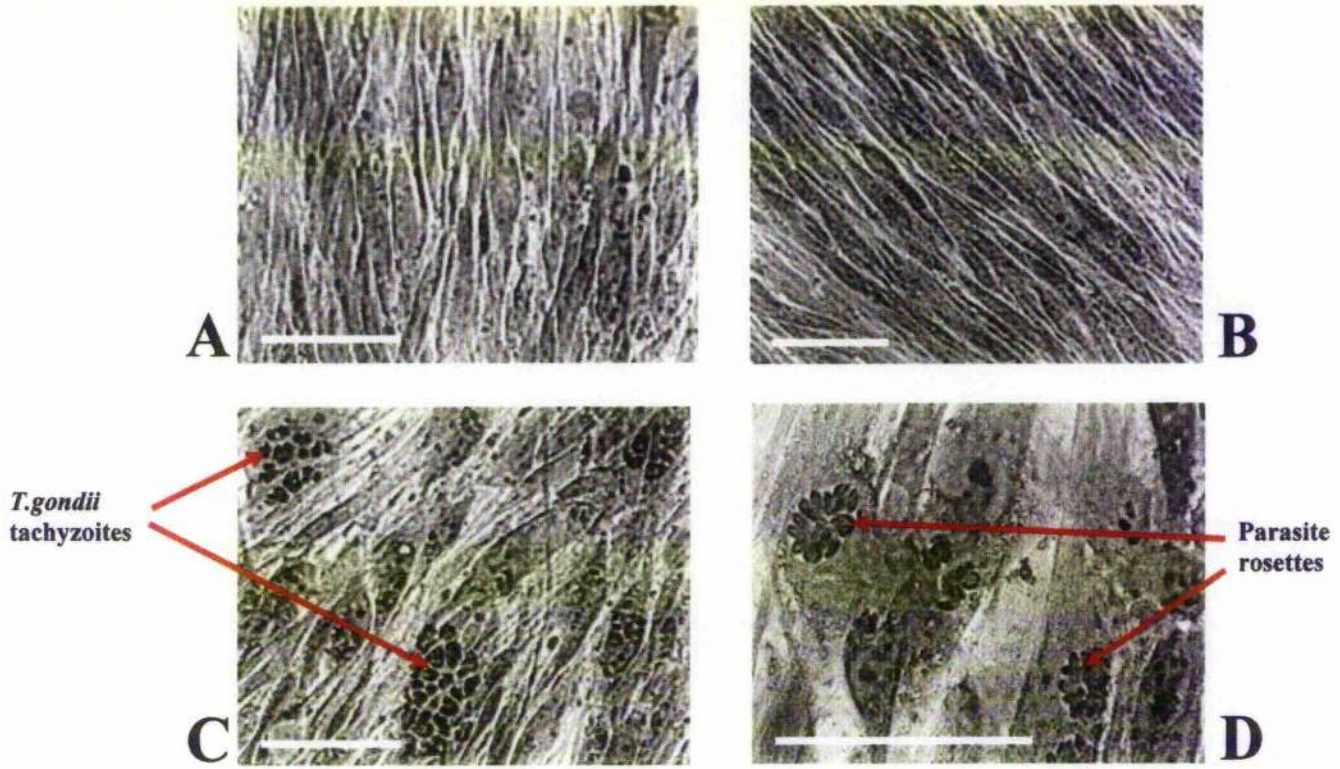


Figure 3-3 A. HFF cells grown in 1cm² chamber slides in DMEM and infected with different ratios of *Toxoplasma gondii* to determine optimal infection ratio. A. Non-infected HFFs at 200x magnification. Cells grow parallel to each other and are thin and tapered towards the ends. **B**. 2:1 Infection ratio at 200x magnification. HFF cells infected with *Toxoplasma gondii* at an infection ratio of 2parasites:1cell. Cells appear darker in colour and cell shape is still similar to that of non-infected HFFs. Cells still growing in the one direction. **C**. 5:1 Infection ratio at 200x magnification. Parasite bodies can be seen within the cells clearly (indicated by the arrows). Cells appear swollen and not growing in a uniform direction. **D**. 5:1 Infection ratio at 400x magnification. A rosette of parasites can be seen in the cell, this rosette shape is characteristic of replicating *Toxoplasma gondii.* Cells are less densely packed than the non-infected cells seen in A. Bar equals 50μm.

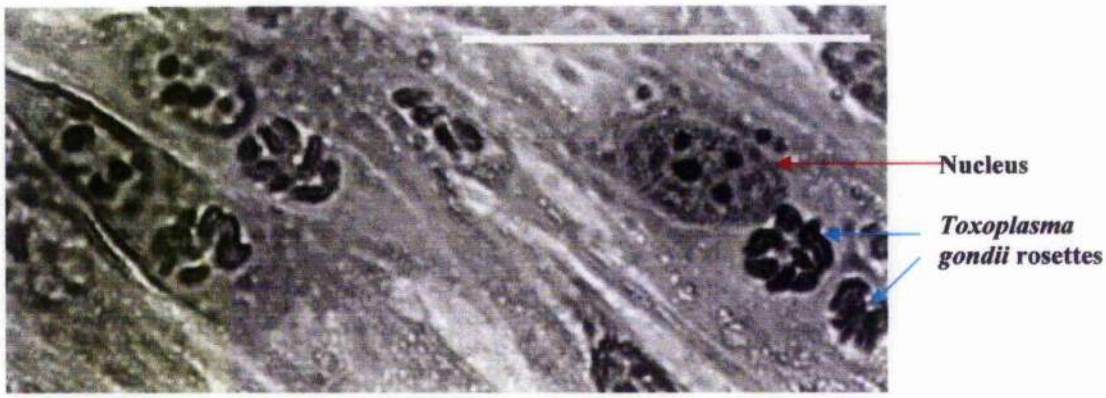


Figure 3-1 E. HFF cells grown in 1cm² chamber slides in DMEM and infected at a 5:1 ratio of parasite: cell. Photographed at 400x magnification. Both the nucleus and parasite rosettes can be seen clearly within the one cell. Bar equals 50μm.

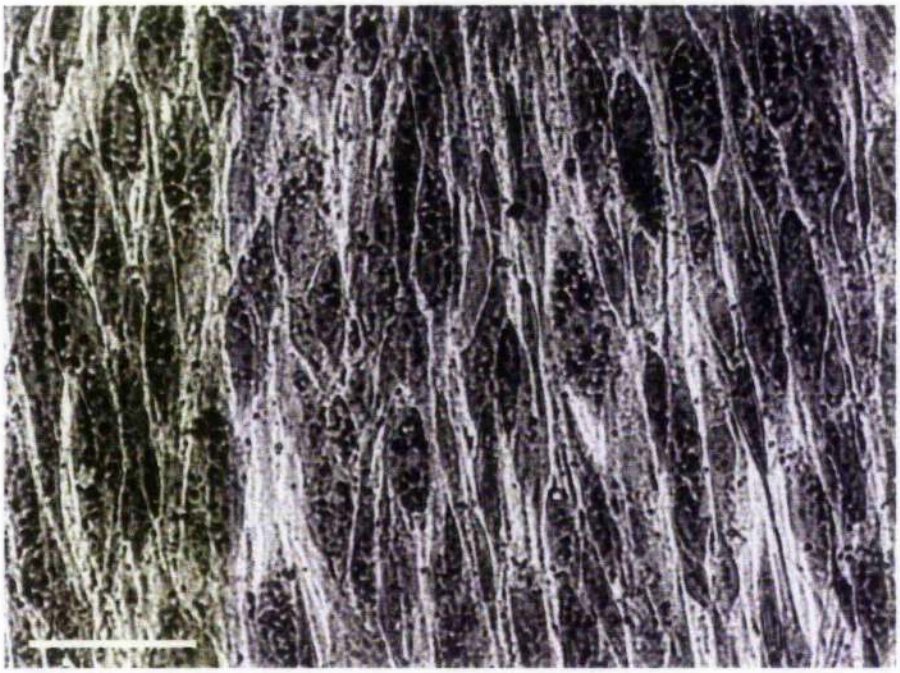


Figure 3-1 F. HFF cells grown in 1cm² chamber slides in DMEM and infected at a 10:1 ratio of parasite: cell. Photographed at 200x magnification. Cells are heavily infected; almost all cells appear to be parasitised. The cells appear swollen and darker than the cells in Figure 3-1 A (non-infected HFFs).

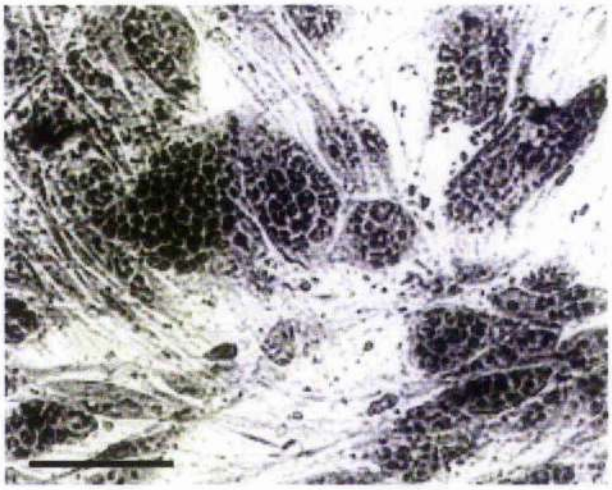


Figure 3-1 G. HFF cells grown in 1cm² chamber slides in DMEM and infected at a 10:1 ratio of parasite: cell. Photographed at 200x magnification. All cells appear to be infected and full of parasites, lysis would probably occur soon. Cells are not densely packed together which may be indicative that cell lysis has occurred perhaps due to the high ratio of infection. Bar equals 50μM.

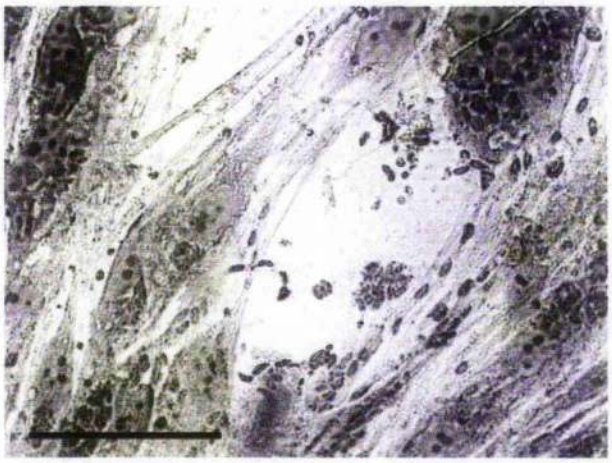


Figure 3-1 H. HFF cells grown in 1cm² chamber slides in DMEM and infected at a 10:1 ratio of parasite: cell. Photographed at 400x magnification. An infected cell has just lysed and free parasites can be seen in the intercellular space. Bar equals 50μM.

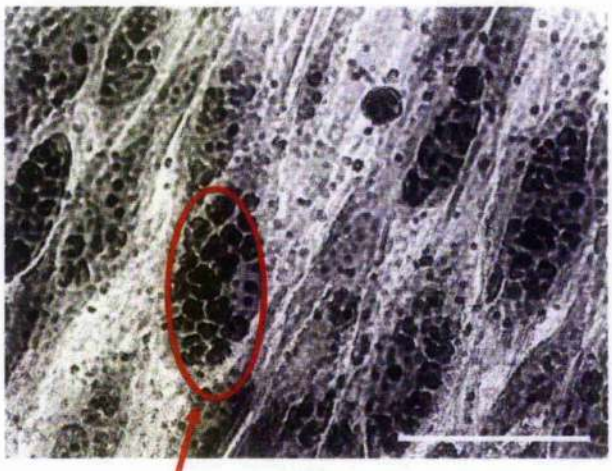


Figure 3-1 I. 10:1 HFF cells grown in 1cm² chamber slides in DMEM and infected at a 10:1 ratio of parasite: cell. Photographed at 400x magnification. Parasites can be seen clearly. The entire cell is full of replicating parasites. Bar equals 50μM.

Three infection ratios were investigated; 2:1, 5:1 and 10:1 parasites: cell. The 10:1 infection ratio was determined to be too high as cells were beginning to lyse at the end of the 24h infection resulting in loss of host material. There was also an increased amount of *Toxoplasma* protein appearing on the gels. At an infection ratio of 2:1 approximately 70% of the cells were determined to be infected 24hpi. This infection ratio ensured the majority of cells were infected but also that there was not too much parasite protein carry over to the 2-DE gels. Earlier time points (6h, 12h and 16h) were infected at ratios of 5:1 and 10:1 parasites: cell.

3.3.2 HFF time course infection experiments

Several hundred proteins were identified as modulated in the host cell during infection with *T.gondii* using the gel-to-gel technique.

Confluent HFFs (passage <20) were infected with *T.gondii* tachyzoites at various infection ratios and left for a set time of 6, 12, 16 or 24h at 37°C, 5% CO_2. The cells were then harvested and the protein expression of the cells analysed using 2-DE. This protein expression profile from the infected cells was then compared with that of non-infected HFFs, of the same age, and harvested at the same time as the infected cells.

The *T.gondii* tachyzoites used to infect the HFFs were maintained in HFFs prior to the infection experiments. The parasites were harvested, passed through a 3μm filter and washed three times in PBS to ensure no HFF cell debris or medium remained.

There is no method available to completely remove all parasites from infected cells during harvesting that would not result in host cell lysis and release of the proteins. Therefore parasites (previously grown in HFFs) were added to the non-infected controls during harvesting to provide equal quantities of parasite proteins in both the infected and non-infected cells. The actual quantity of parasite proteins present in the sample is very small compared to the amount of host cell protein. The proteins were separated using 2-DE with the aim that any parasite proteins present in the infected cells and in sufficient quantities to appear on the 2-DE gel, would also be present in the non-infected controls and would therefore not appear to be a differentially expressed host cell protein. There is a possibility that parasite proteins which appear on the 2-DE gel are in fact proteins which are induced or repressed when *Toxoplasma* infects, replicates within or is released from the host cell. These proteins would therefore be of interest in determining which parasite proteins are

involved in the host-pathogen interaction. Further analysis of the *T.gondii* proteins identified from the 2-DE gels was not within with the scope of this project.

Both pH3-10NL and pH4-7 IPG strips were used to separate the infected and non-infected cell extracts. In the wide range pH3-10NL two-dimensional maps, there is often a loss of resolution in the region pH4-7, most probably due to the fact that the pI values of many proteins occur in this range and therefore protein spots overlap on the gel, therefore electrophoresis was also performed using pH 4-7 strips to improve protein separation.

24h replicates 1-7 were analysed using the 2DElite software and by manual gel to gel comparison. Spots found to be present in one gel and not in the corresponding gel were marked as differentially expressed, picked and analysed using MALDI MS. The 24h replicates 8, 9 and 10 were not picked for MS, but were included as replicates in the comparative analysis of protein expression.

In 24h replicate 11, differential spot analysis was carried out using the 2DElite software (Amersham Biosciences). Figure 3-2 shows the proteome map of non-infected HFFs and Figure 3-3 shows the proteome map of *Toxoplasma gondii* infected HFFs from 24h infection replicate 11. Figure 3-2 and Figure 3-3 show good resolution of spots from high to low molecular weight and from acidic to basic pH. There is little streaking aside from some minor vertical streaking on the non-infected gel, and some horizontal streaking at the basic end of each gel.

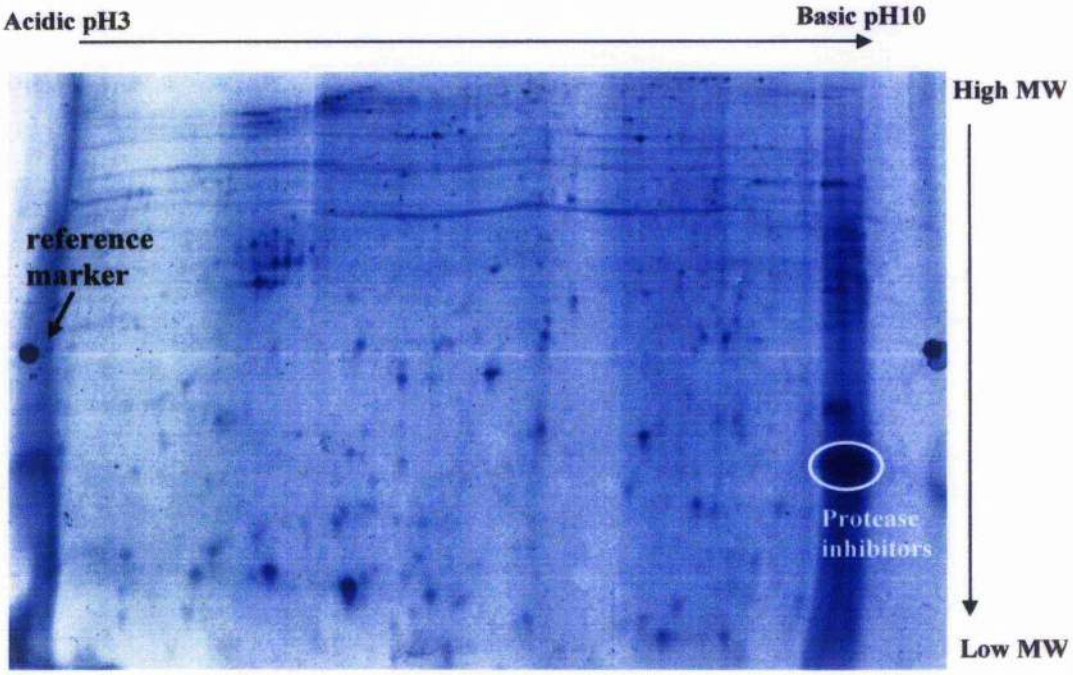


Figure 3-2 Proteome map of non-infected HFFs from 24h infection experiment replicate 11. 254µg of protein were separated by 2-DE on a pH4-7 IPG strip and the gel stained with colloidal Coomassie. Protease inhibitors added during cell lysis can be seen on the bottom right of the image. There is good separation of proteins from acidic to basic pH and high to low molecular weight.

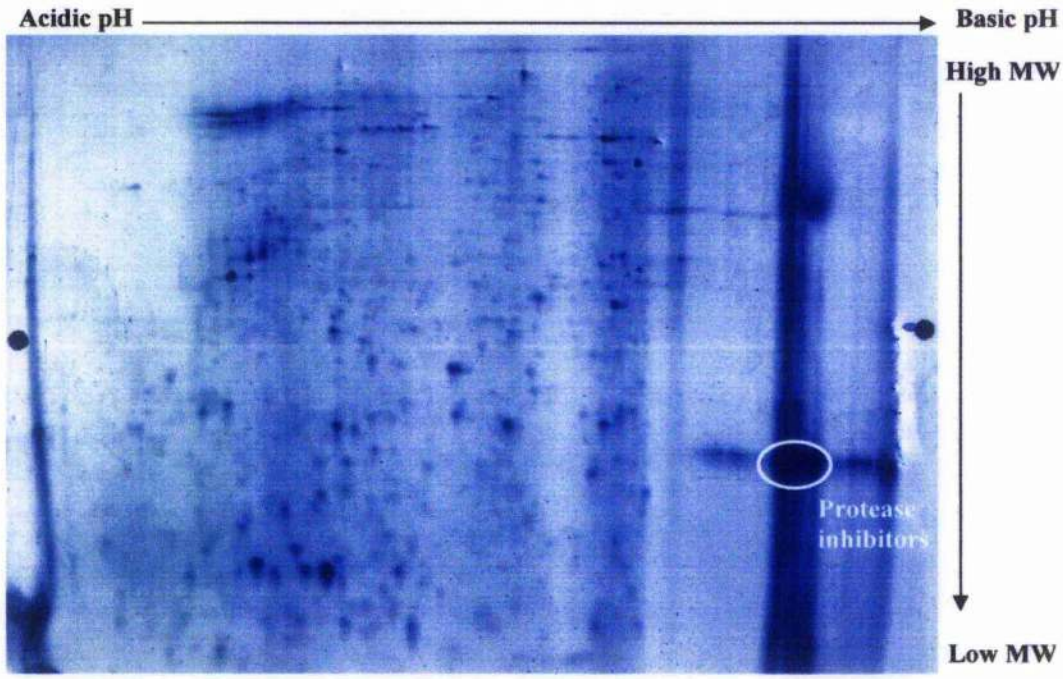


Figure 3-3 Proteome map of *Toxoplasma gondii* infected HFFs from 24h infection experiment replicate 11. 254µg of protein were separated by 2-DE on a pH4-7 IPG strip and the gel stained with colloidal Coomassie. Protease inhibitors added during cell lysis can be seen on the bottom right of the image. There is good separation of proteins from acidic to basic pH and high to low molecular weight.

Figure 3-4 and Figure 3-5 show all the detected spots on the non-infected (299 spots) and infected (254 spots) gels respectively after analysis with the 2DElite Image Master software. Dust and scratches on the gel detected as spots were removed manually. The software overlays the two images and background subtraction and normalization (lowest on boundary) were carried out to gain accurate spot volumes. Spots common to both gels were matched manually. Any spots whose volume differed between gels by 30% or more, or a spot which was present on one gel and not the other, was marked of interest. A cut-off of 30% was chosen to ensure that the changes seen were indeed biological changes and not experimental artefacts. For example, differences in the IPG strips or gels can mean that the same protein on two gels can separate differently during IEF or SDS-PAGE resulting in that protein spot appearing differentially expressed whereas in actual fact it is due to experimental differences.

Figure 3-6 and Figure 3-7 show the spots marked for picking from non-infected and infected cells respectively. The pick list was then exported from 2DElite for use with the Spot Handling Work Station (Amersham Biosciences). From the two gels, 313 spots were picked, 144 from the non-infected gel and 169 from the infected gel. Each of these spots was analysed by MS (MALDI or MS/MS). The PMF and the peptide fragment data were searched against the NCBI homosapiens database and ToxoDB using MASCOT®. Protein identification of 143 individual spots on the infected gel and 125 individual spots on the non infected gel was possible. Of the 143 spots identified on the infected gel 45 spots were *Toxoplasma gondii* proteins. Of the 223 human spots identified, 89 human protein species were identified, as often several proteins were present in one spot.

Figure 3-8 and Figure 3-9 show the gel images annotated with the protein names. Figure 3-8 shows a non-infected cell lysate; all annotated proteins on this image are down regulated during infection with *T.gondii*. Figure 3-9 is of the infected cell lysate therefore all proteins annotated on this image are up regulated during infection with *T.gondii*.

An example of a MALDI spectrum and results page was shown in Chapter II. Figure 3-10 shows the MS/MS trace and Figure 3-11 the MASCOT® output. In the case of MS/MS data, it is the ion scores for individual peptide matches that are statistically meaningful. A peptide summary report groups peptide matches into protein hits, and derives a protein score from the combined ions scores. A score of >33 was deemed significant in this study and this spot was identified as protein disulfide isomerase (PDI). Figure 3-11B shows the actual ion series of one of the peptides, the y ion series is almost complete therefore it is possible to be very confident that this peptide match and protein identification is correct.

To determine protein function, each protein was searched against the Harvester search engine at EMBL (www.harvester.embl.de). This provided information about the function of the protein, its location in the cell, the corresponding gene name, PubMed (www.pubmed.co.uk) references and information about the gene ontology of the protein. Chapter VIII discusses the functions of the proteins identified as modulated during the infection process.

Figure 3-12 shows some areas of the infected (A) and non-infected (B) areas of the gel images magnified to highlight the differences in spot expression. As seen in this figure, several heat shock proteins are up regulated in infected cells (HSPA9B, HSPA8, MTHSP75 and HSPA6). Down regulated in infected cells are several structural proteins; vimentin and three tropomyosin isoforms (VIM, TPM2, TPM3, TPM4).

Table 3-2 contains the information about the spots that matched between the infected and non-infected gels and were subsequently identified by MS. The table contains information about the spot number, the spot co-ordinates, spot volume and the normalised volume of each spot and the volume difference between the matched spots. Figure 3-13 shows the spots that matched across the 2 gels whose volume differed by 30% or more. This figure shows the non-infected gel with the matched spots from the infected gel superimposed. Red spots indicate decreased spot volume in infected cells, and green spots indicate increased spot volume in infected cells.

Table 3-3 is a section of the results table from 24h infection experiment replicate 11 and the full results of all modified proteins will be discussed in greater detail in Chapter VIII. The full results table can be seen in Appendix I. This table contains information about each spot that had a significant hit in MASCOT®. The table includes the picklist number, the number of proteins identified from the one spot, whether the protein is human or parasite, whether it was identified using MALDI or MS/MS, the score from the MASCOT® search, the protein name and gene name, the mass and pI of the protein, the number of peptides matched, the sequence coverage of the protein and the gene identification number (gI number). The table contains the URL links to the MASCOT® results page, allowing others to re-query the data. It also gives the volume of any matched proteins and some information on the protein function.

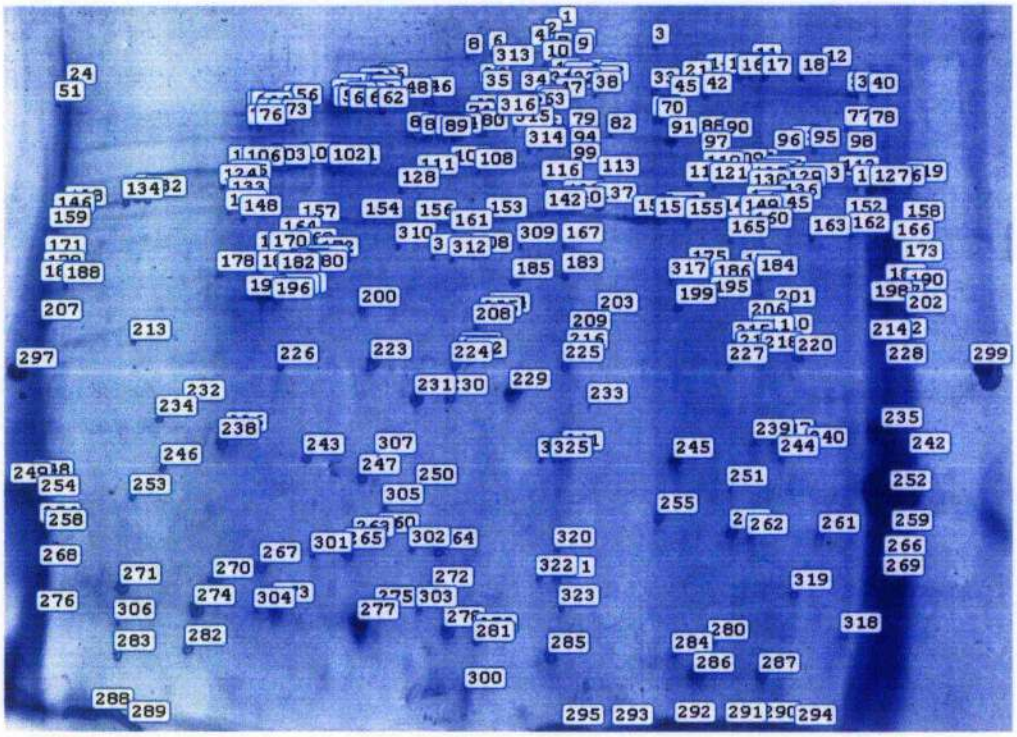


Figure 3-4 Non-infected gel image with all spots detected using the 2DElite software. 299 spots were detected on this gel. The volume of each spot was calculated relative to background and normalisation.

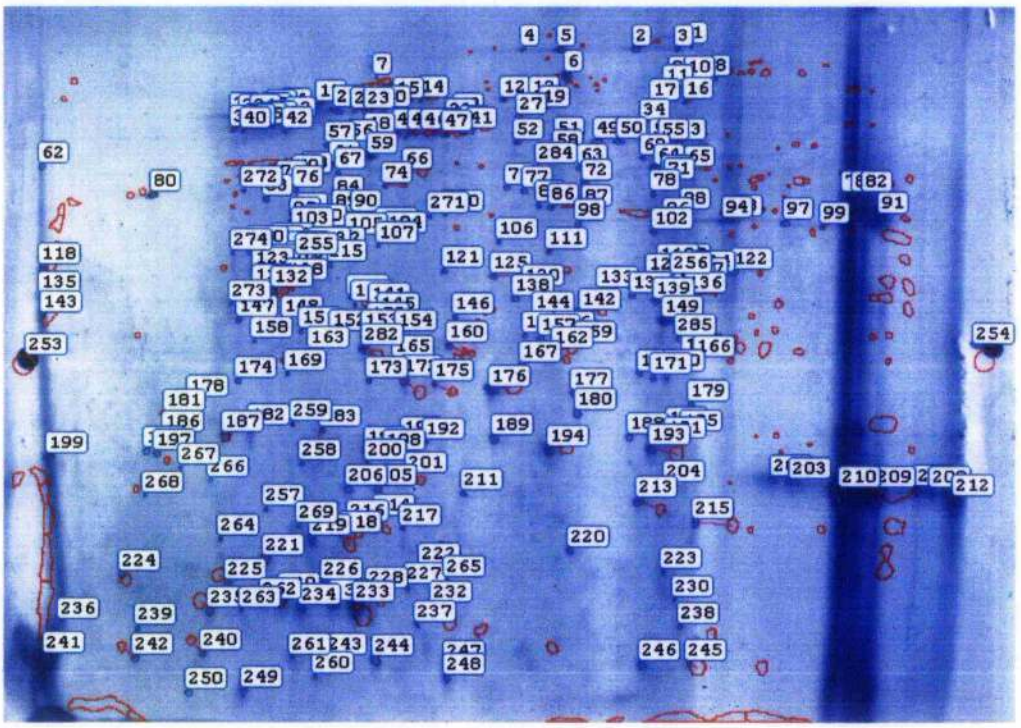


Figure 3-5 Infected gel image with all spots detected using the 2DElite software. The volume of each spot was calculated relative to background and normalisation. The blue outlines show the spot boundaries of infected proteins. The red outlines show the corresponding spot from the non-infected image. The two images are overlaid and common spots matched between gels (linked by a purple line).

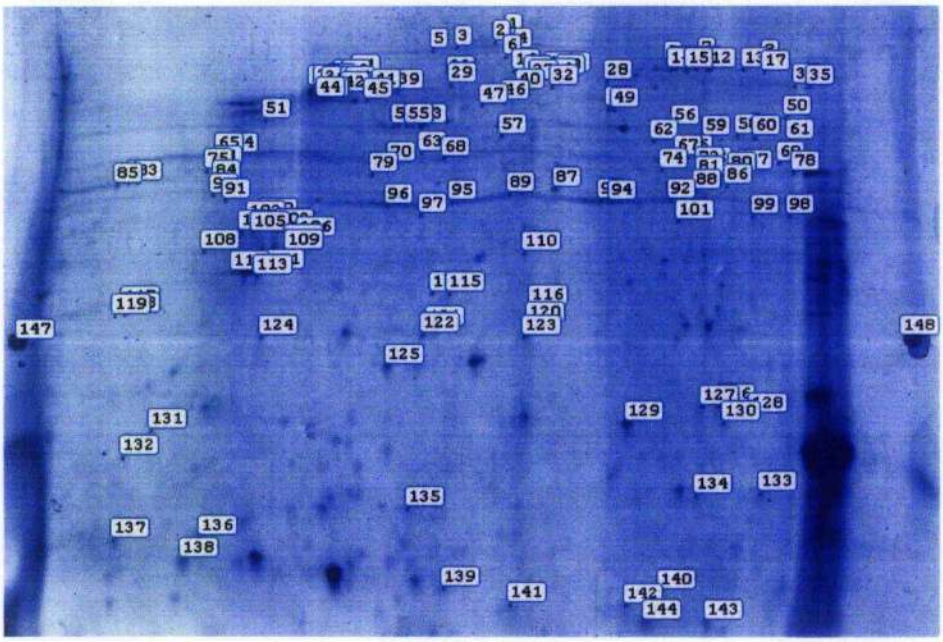


Figure 3-6 The non-infected gel image with all proteins of interest numbered for spot picking. These spots are either not present on the infected gel, or are present but the volume difference is 30% or greater. 144 spots were marked of interest and subsequently analysed by MS.

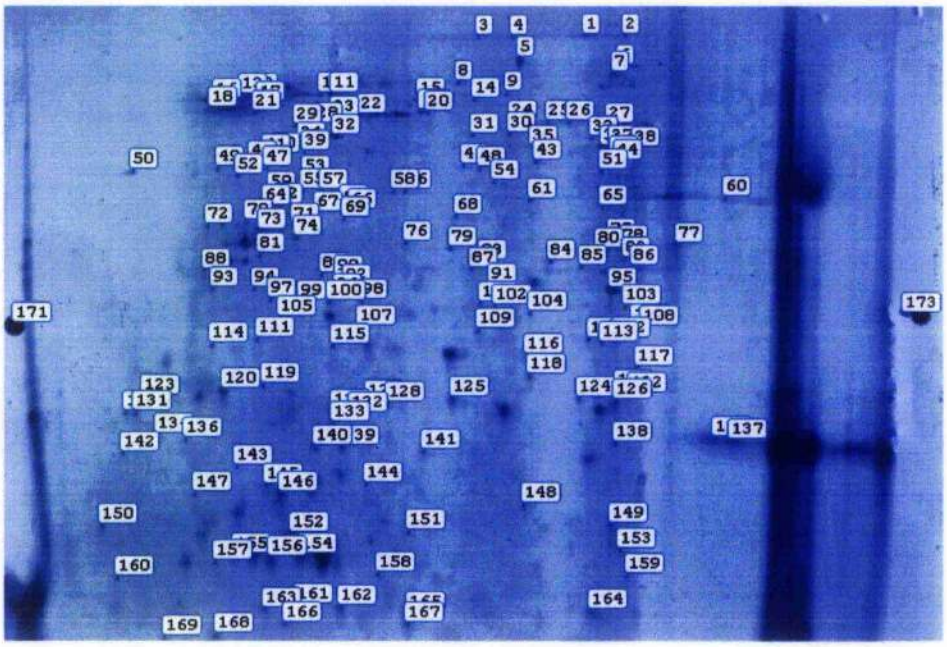


Figure 3-7 The infected gel image with all proteins of interest numbered for spot picking. These spots are either not present on the non-infected gel, or are present but the volume difference is 30% or greater. 169 spots were marked of interest and subsequently analysed by MS.

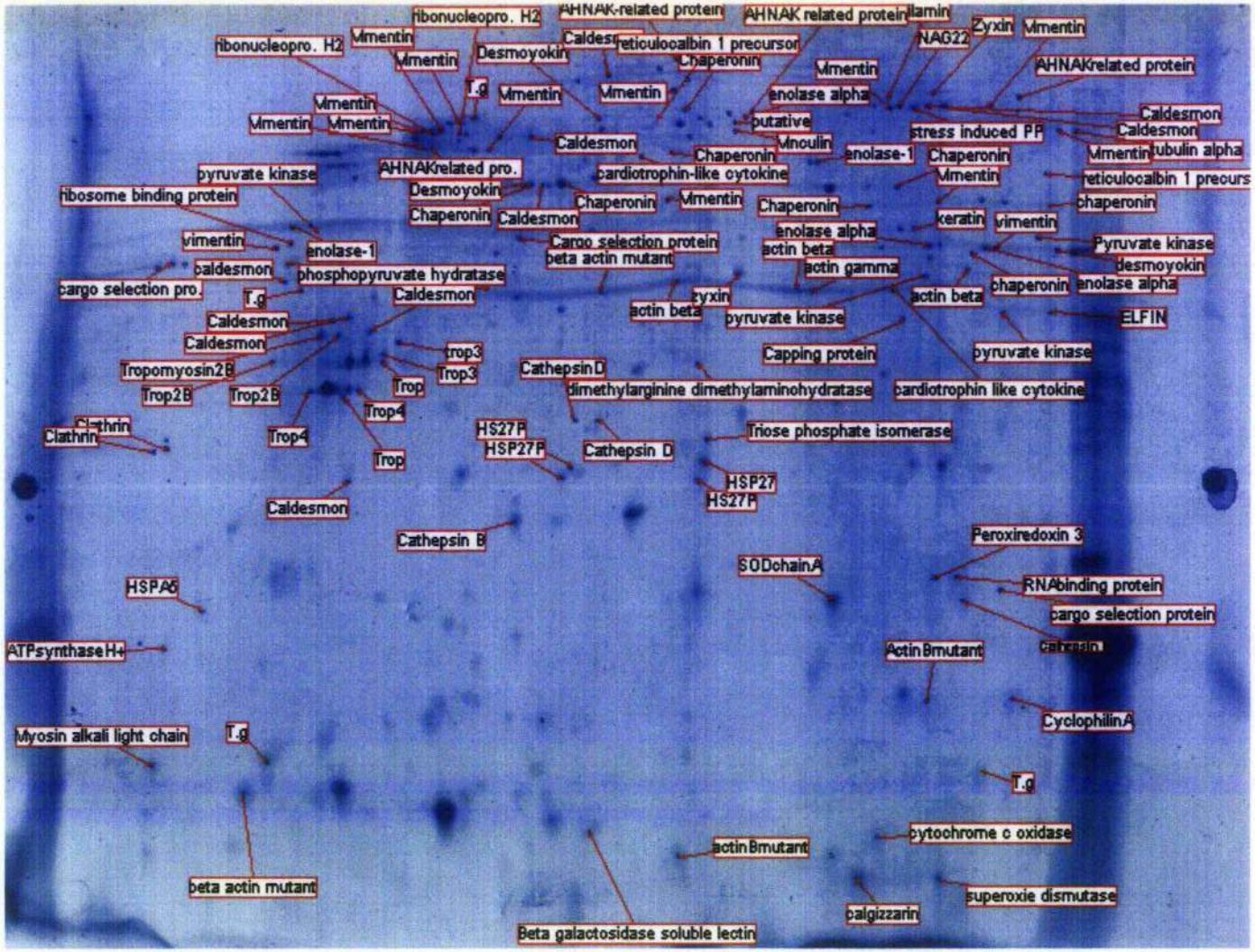


Figure 3-1 Annotated 2-DE gel of non-infected HFFs. All of the annotated proteins were identified using MALDI or MS/MS. All of the annotated proteins in this figure are down regulated in the host cell during infection with *Toxoplasma gondii* 24hpi.

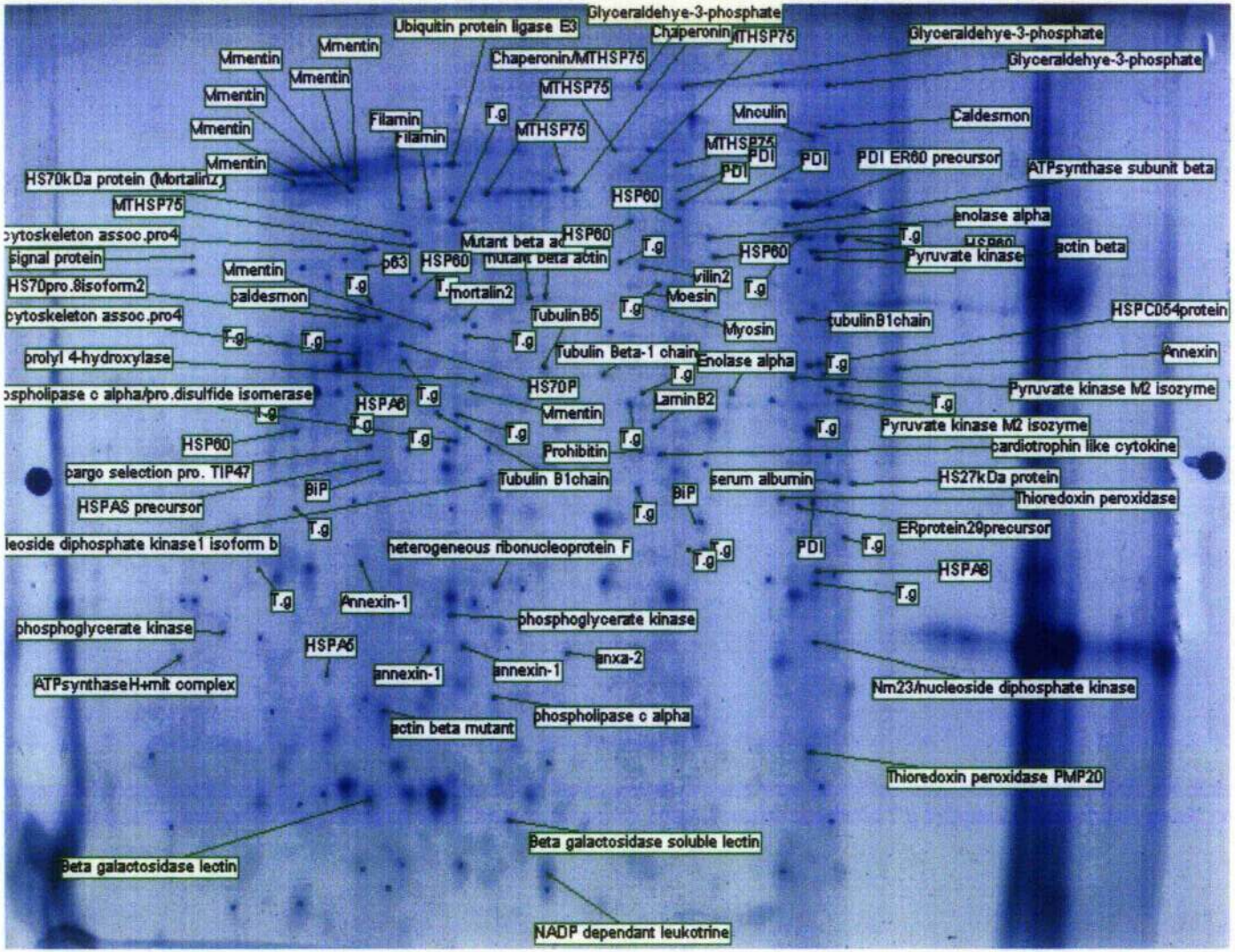


Figure 3-2 Annotated 2-DE gel of infected HFFs. All of the annotated proteins were identified using MALDI or MS/MS. All of the annotated proteins in this figure are up regulated in the host cell during infection with *Toxoplasma gondii* 24hpi. Proteins annotated T.g indicates a *Toxoplasma gondii* protein

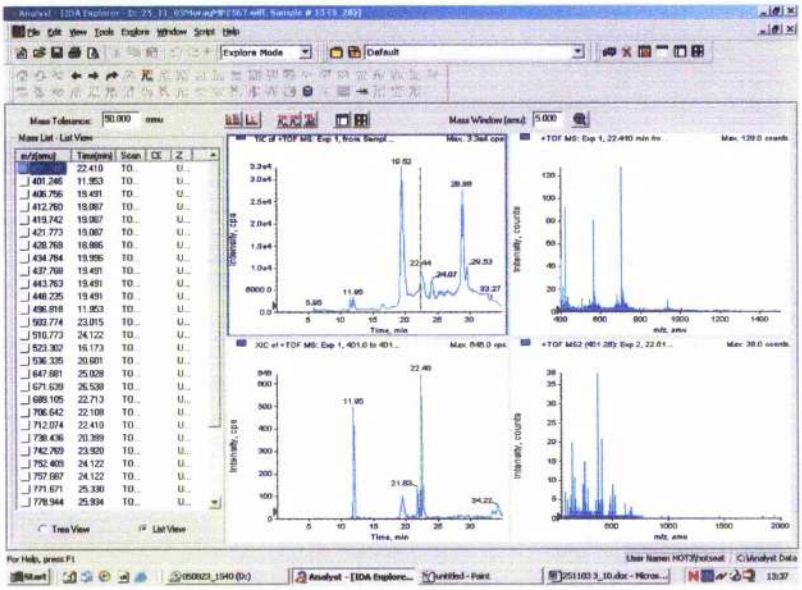
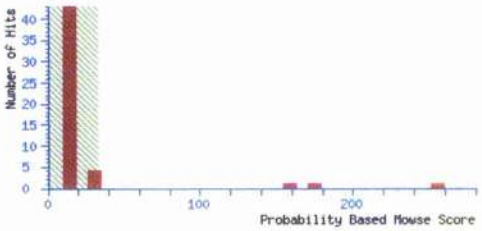


Figure 3-10 Screenshot of an MS/MS trace on the Analyst software. The MS trace was submitted via MASCOT® for database searching. The highlighted panel (top left) shows the time in minutes of the run (x-axis) versus the intensity of the signal of peptides (y-axis). This example shows that peptides started to be eluted from the column at around 20min. The other 3 panels show actual peptide fragment mass data at a specific time point.

User : **Morag**
Email : **9606340n@student.gla.ac.uk**
Search title : **24hexpt11MSMSplate2_68**
MS data file : **C:\morag\10_01_04Morag2566.wiff**
Database : **NCBInr 20031122 (1531464 sequences; 497479537 residues)**
Taxonomy : **Homo sapiens (human) (93722 sequences)**
Timestamp : **15 Jan 2004 at 10:48:03 GMT**
Significant hits: gi|1085373 **protein disulfide-isomerase (EC 5.3.4.1) ER60 precursor - human**
gi|27805905 **glucose regulated protein 58kD [Bos taurus]**
gi|136066 **TRIOSEPHOSPHATE ISOMERASE (TIM)**

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 33 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

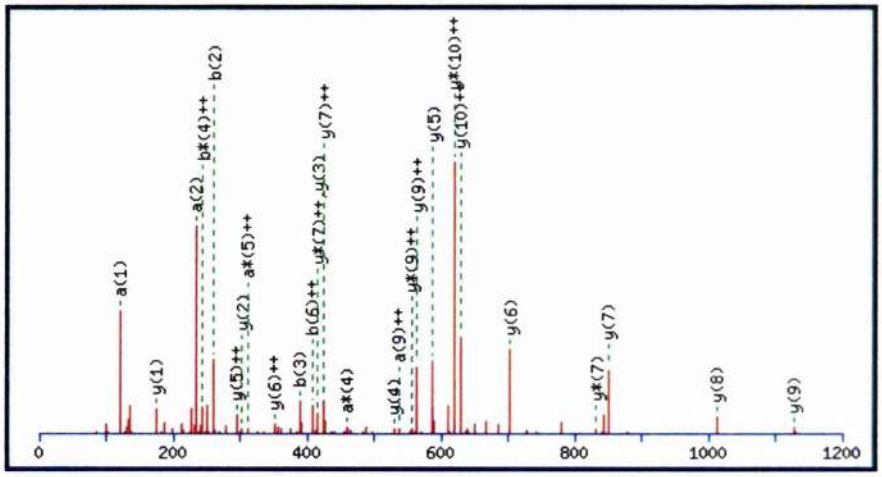


Peptide Summary Report

Format As | Peptide Summary | Help
Significance threshold p< 0.05 | Max. number of hits 20
Standard scoring ◉ MudPIT scoring ○ | Ions score cut-off 0 | Show sub-sets ☐
Show pop-ups ◉ Suppress pop-ups ○ | Sort unassigned Decreasing Score | Require bold red ☐
Select All | Select None | Search Selected | ☐ Error tolerant | Archive Report

1. gi|1085373 **Mass:** 57043 **Score:** 255 **Queries matched:** 8
protein disulfide-isomerase (EC 5.3.4.1) ER60 precursor - human
☐ Check to include this hit in error tolerant search or archive report

A

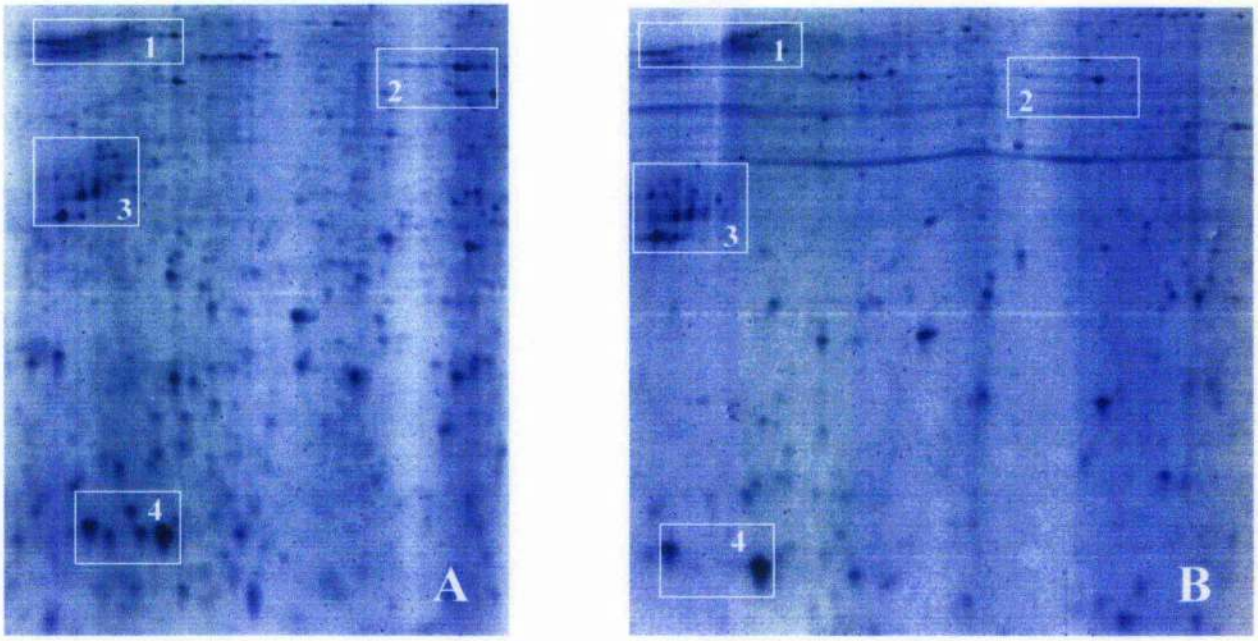


Monoisotopic mass of neutral peptide Mr(calc): 1514.75
Fixed modifications: Carbamidomethyl (C)
Ions Score: 44 **Expect:** 0.0044
Matches (Bold Red): 27/124 fragment ions using 46 most intense peaks

#	a	a^{++}	a^*	a^{*++}	b	b^{++}	b^*	b^{*++}	Seq.	y	y^{++}	y^*	y^{*++}	#
1	**120.08**	60.54			148.08	74.54			F					12
2	**233.16**	117.09			**261.16**	131.08			L	1368.69	684.85	1351.66	676.34	11
3	361.22	181.12	344.20	172.60	**389.22**	195.11	372.19	186.60	Q	1255.61	**628.31**	1238.58	**619.79**	10
4	476.25	238.63	**459.22**	230.12	504.25	252.63	487.22	**244.11**	D	**1127.55**	**564.28**	1110.52	**555.76**	9
5	639.31	320.16	622.29	**311.65**	667.31	334.16	650.28	325.64	Y	**1012.52**	506.76	995.49	498.25	8
6	786.38	393.69	769.36	385.18	814.38	**407.69**	797.35	399.18	F	**849.46**	**425.23**	**832.43**	**416.72**	7
7	901.41	451.21	884.38	442.69	929.40	465.21	912.38	456.69	D	**702.39**	**351.70**	685.36	343.19	6
8	958.43	479.72	941.40	471.21	986.43	493.72	969.40	485.20	G	**587.36**	**294.18**	570.34	285.67	5
9	1072.47	**536.74**	1055.45	528.23	1100.47	550.74	1083.44	542.22	N	**530.34**	265.67	513.31	257.16	4
10	1185.56	593.28	1168.53	584.77	1213.55	607.28	1196.53	598.77	L	**416.30**	208.65	399.27	200.14	3
11	1313.65	657.33	1296.63	648.82	1341.65	671.33	1324.62	662.81	K	**303.21**	152.11	286.19	143.60	2
12									R	**175.12**	88.06	158.09	79.55	1

B

Figure 3-11 MS/MS spectra MASCOT® results page. A. This protein was identified as protein disulfide isomerase. The MASCOT® results page gives information on the protein gene identifier (gI), the protein score (>33 is significant in this case), the observed and expected peptide mass and the peptide sequence. **B.** Eight peptides from the experimental MS data matched peptides in the theoretical database. One of these peptides is shown in B. MASCOT® shows the ion series, the masses of the ions and the identified amino acids. The y ion series for this peptide is almost complete. The difference in mass between two adjacent y ions is the equivalent mass of an amino acid, therefore actual sequence data can be generated from these results leading to unambiguous identification of proteins.



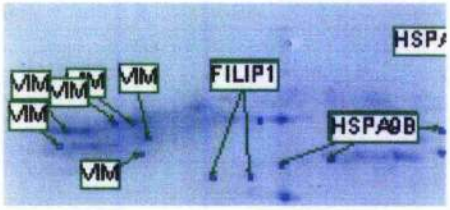
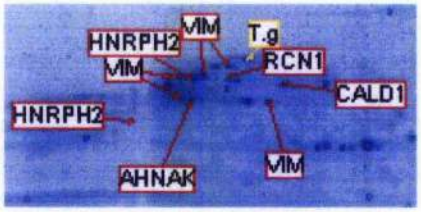
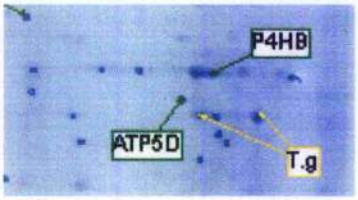
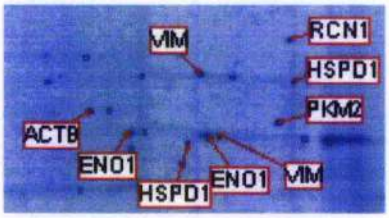
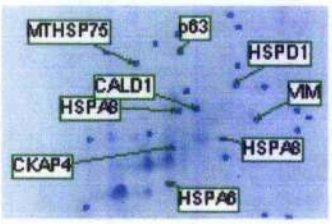
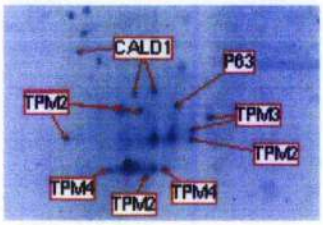
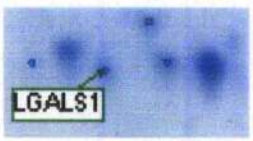
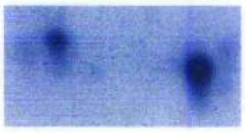
	A	B
1		
2		
3		
4		

Figure 3-12 Proteomic differences between *Toxoplasma gondii* infected cells (A) and non-infected cells (B). 2-DE maps obtained from both samples are shown at the top (IPG strips pH 4-7 colloidal Coomassie stained). Areas of differences between the two samples are enlarged and shown in the table. The proteins are annotated by gene name. A full list of gene and protein names can be seen in Chapter VIII.

	Gel	Spot number	Score	Protein	Volume INF	Normalised volume INF	Volume NON	Normalised volume NON	% Difference
1	11inf	8	218	MTHSP75	308,258	46.880	169.2	31.743	32
2	11inf	11	39	Ubiquitin protein ligase E3	2,809,021	427.180	1,222,354	210.561	51
3	11inf	13	709	Vimentin	505,971	76.948	316,665	59.408	36
4	11inf	22	448	MTHSP75	1,035,501	157.478	240,430	45.106	71
5	11inf	22	278	Chaperonin HSP60	1,035,501	157.478	240,430	45.106	71
6	11inf	24	255	Protein disulfide isomerase	328597	49.973	21,672	4.066	92
7	11inf	79	288	Tubulin beta-1 chain	218,489	33.228	966,243	181.274	82
8	11inf	86	327	Pyruvate kinase M2 isozyme	286,837	43.622	38,598	7.241	83
9	11non	52	760	Chaperonin HSP60	1172219	178.270	688421	129.152	28
10	11non	53	997	Chaperonin HSP60	1,949,914	296.542	4,300,463	807	63
11	11non	70	67	Cargo selection protein	239318	36.395	1451165	272.250	87
12	11non	79	99	Dynactin 2	135621	20.625	212960	39.953	48
13	11non	87	43	Zyxin	1659562	57.718	682002	103.731	80
14	11non	104	496	Tropomyosin 2beta	379,651	57.731	2,013,718	377.787	85
15	11non	107	208	Tropomyosin 3	59276	9.015	168059	31.529	71
16	11non	116	339	Triose phosphate isomerase, chain A	79048	12.076	630702	118.324	90
17	11non	120	301	HS27kDa protein 1	718,268	109.234	2,085,338	319	66
18	11non	125	149	Cathepsin B	836465	127.209	2322288	435.677	71
19	11non	129	120	Superoxide dismutase Chain A	2,323,244	353.317	5,040,928	945.712	63
20	11non	138	156	Actin beta, mutant	673,296	102.394	1,973,275	370.200	72
21	11non	144	221	Calgizzarin	138,025	20.991	2,617,147	499.994	96

Table 3-2 Volume differences of the protein spots that matched across the two gels that were subsequently identified using MS. This table shows the spot numbers, the MASCOT® score, the protein identified, the spot volume, normalised volume and the difference in volume of the matched spots between the two gels (far right column).

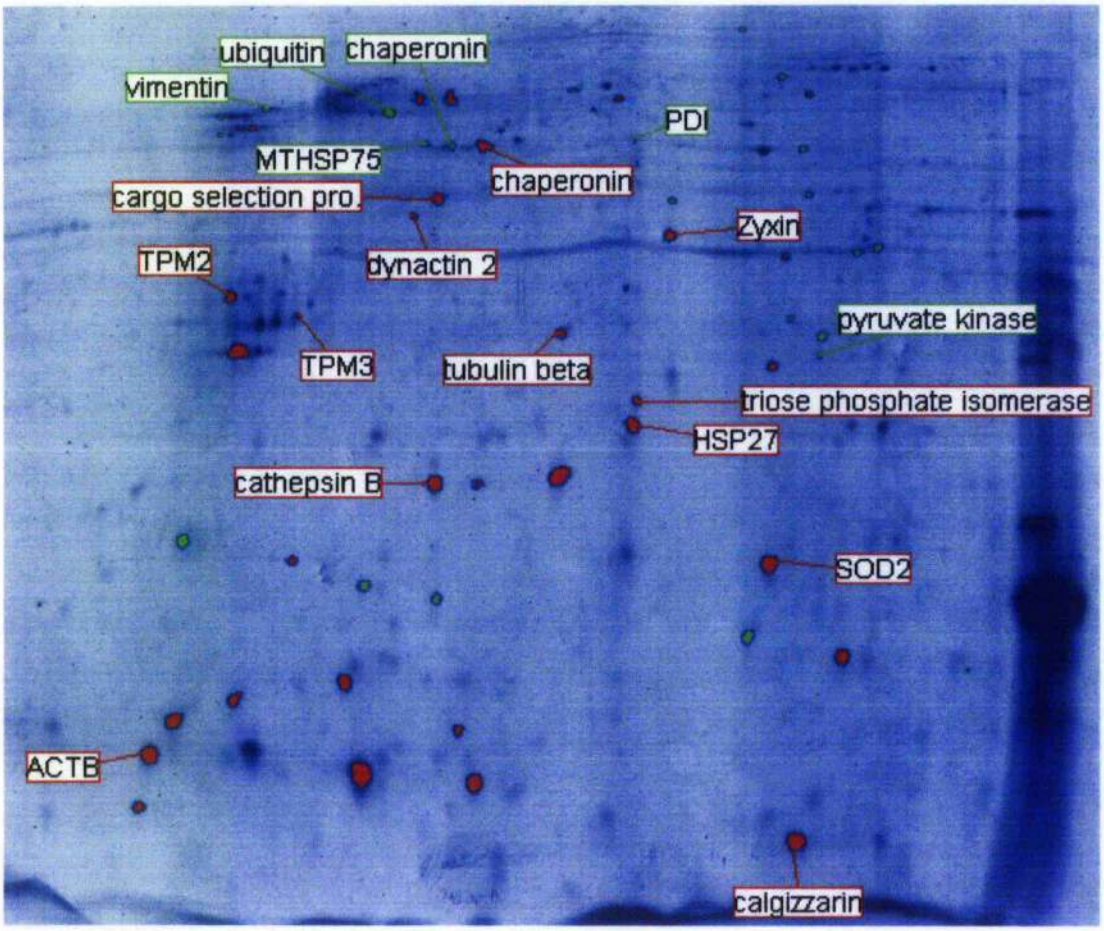


Figure 3-13 2-DE gel of non-infected cells showing the identified matched spots. This image shows the proteins which matched across the two gels and differed in volume by 30% or more and were identified using MS. This gel profiles non infected cells. Protein names boxed in green indicate spots which were up regulated during infection and protein names boxed in red, indicate spots whose volume decreased in infected cells. Red spot = volume less in infected image, green spot = volume greater in infected image.

1	2	3	4	5	6	7	8	9	10	11
U/D	Gel	Spot No.	Hits	MALDI/MS/MS	T.g/H	Score	Id	Gene name	Mass	pI
DOWN	11non	125	1 of 3	MS/MS	H	149	Cathepsin B	CTSB	21907	5.27
DOWN	11non	114	1 of 1	MS/MS	H	137	Cathepsin D preprotein	CTSD	45037	6.1
DOWN	11non	115	1 of 1	MS/MS	H	94	Cathepsin D, Chain B	CTSD	26457	5.31
UP	11inf	141	1 of 1	MS/MS	H	148	ANXA2 protein	ANXA2	38780	7.57
UP	11inf	77	1 of 1	MS/MS	H	381	ANXA2 protein	ANXA2	387801	7.57
UP	11inf	2	1 of 1	MS/MS	H	92	Gastric associated differentially expressed protein	GAPD	14858	6.84
UP	11inf	3	1 of 1	MS/MS	H	161	Gastric associated differentially expressed protein	GAPD	14858	6.84
UP	11inf	4	1 of 1	MS/MS	H	247	Gastric associated differentially expressed protein	GAPD	14858	6.84
DOWN	11non	143	1 of 1	MS/MS	H	131	Superoxide dismutase [Mn], mitochondrial precursor	SOD2	24878	8.35
DOWN	11non	129	1 of 1	MS/MS	H	120	Superoxide dismutase Chain A	SOD2	15972	5.54

12	13	14	15	16	17	18	19	20	21	22	23	24
P.M	s.c (%)	gI number	Mascoturl	gInumber	EMBL	Function	GO	GO biological process	vol inf cell	vol non-inf cell	FunCat	
4	0	gi|3929733	http://fun-g	http://fun-g	http://harve	thiol protea	http://goda	proteolysis and peptidolysis	836,465	2,322,288	1	metabolism
3	6	gi|4503143	http://fun-g	http://fun-g	http://harve	acid protea	http://goda	aspartic-type endopeptidase			1	metabolism
4	18	gi|4503143	http://fun-g	http://fun-g	http://harve	acid protea	http://goda	aspartic-type endopeptidase			1	metabolism
4	13	gi|18645167	http://fun-g	http://fun-g	http://harve	play a role	http://goda	anti-inflammatory			10	signal transduction
8	24	gi|18645167	http://fun-g	http://fun-g	http://harve	play a role	http://goda	anti-inflammatory			10	signal transduction
4	30	gi|6970062	http://fun-g	http://fun-g	http://harve	2nd phase	http://www	glycolysis/oxidoreductase			2	glycolysis
4	41	gi|6970062	http://fun-g	http://fun-g	http://harve	2nd phase	http://www	glycolysis/oxidoreductase			2	glycolysis
6	41	gi|6970062	http://fun-g	http://fun-g	http://harve	2nd phase	http://www	glycolysis/oxidoreductase			2	glycolysis
3	16	gi|134665	http://fun-g	http://fun-g	http://harve	destroys r	http://goda	response to oxidative stress			11	stress response
3	0	gi|349904	http://fun-g	http://fun-g	http://harve	destroys r	http://goda	response to oxidative stress	2,323,244	5,040,928	11	stress response

Table 3-3 An excerpt from the 24h infection experiment replicate 11 results table. Figure legend: Column 1= spot is up or down regulated on infected gel, 2= spot picked from infected or non-infected gel, 3= spot number, 4= number of significant hits from that spot, 5= protein identified using MALDI or MS/MS, 6= protein is human or from *Toxoplasma gondii*, 7= MASCOT® score, 8= Protein id., 9= gene name, 10= Mass, 11= pI, 12= peptides matched (P.M), 13= sequence coverage (s.c), 14= gI number, 15= MASCOT® url, 16= gI number url, 17= EMBL link, 18= protein function, 19= gene ontology url, 20= GO biological process, 21= volume of spot on infected gel, 22= volume of spot on non-infected gel, 23= FunCat number, 24= FunCat category.

3.3.3 24h infection experiment replicates

Eleven replicates of a 24h infection were carried out in total. Of these 11 replicates, 8 gel sets were analysed and differentially expressed spots picked. The number of spots chosen for MS analysis and the percentage of spots identified can be seen in Table 3-4. All of the data from the 24h experiments will be discussed in further detail in Chapter VIII.

A limitation of the techniques discussed in this chapter is that it is not always possible to identify all the differentially expressed spots. This can often be due to equipment error e.g. loss of the excised spot plug within the Ettan Spot Handling workstation or loss of sample in a faulty MS run. Other reasons for non-identification can be that not enough protein was present in the spot on the gel or in the case of MALDI analysis; more than one protein was present in the spot.

In total, it was possible to get protein identifications for 38% of spots picked from the 24h infected gels and 44% of spots picked from the non-infected gels. Combining all the data from all of the 24h replicates shows that overall in cells infected with *T.gondii* for 24h, 59 individual proteins were up regulated in expression and 63 proteins were down regulated in expression in the infected cell. However, often the same protein was identified in both the infected and non infected gel and was therefore labelled modulated. This is most likely due to the fact that there were often several isoforms of the same protein present on the gel due to PTM. All the differentially expressed proteins were combined and assigned status of up regulated, down regulated or modulated in the parasite infected cell. This gave the result of 85 individual proteins being differentially expressed in the host pathogen response, 46% were up regulated, 39% were down regulated and 17% were modulated upon infection.

∴ overall, 85 unique pro's modulated 24h p.i

24h infection experiment	Gel	Total no. spots picked	Total no. of spots id.	no. T.g spots id.	no. human spots id.	no of individual human proteins id.	Infection ratio
1	1inf	91	34	20	14	13	10 to 1
	1non	58	30	0	30	16	
2	2inf	135	22	19	3	3	10 to 1
	2non	13	0	0	0	0	
3	3inf	132	22	14	8	7	10 to 1
	3non	73	25	0	25	20	
4	4Ainf	46	4	1	3	3	1 to 1
	5Bnon	135	10	0	10	10	
6	6inf	33	4	0	4	4	1 to 1
	6non	27	15	0	15	12	
7	7inf	7	5	2	3	3	1 to 1
	7non	12	11	0	11	8	
11	11inf	169	143	45	98	43	2 to 1
	11non	144	132	7	125	46	
Total			457	108	349		

	Total no of spots picked	approx no. of T.g spots	approx. no. of human spots	no. human proteins id.	Human % id.
infected	613 up	264	349	121	38%
non-infected	462 down	0	462	216	44%

Table 3-4 Number of spots chosen for MS analysis in the 24h infection experiments and the number of proteins identified. In general, more spots were seen as up regulated than down regulated in infected cells (57% had increased volume, 43% decreased volume). However, this is partly due to the presence of *T.gondii* spots on the infected gel. Of the 613 infected differentially expressed spots that were analysed by MS, 38% were identified. Of this 38%, 43% were proteins from *Toxoplasma* and 57% were human proteins. If it is then assumed that 43% of the total number (613) of up regulated spots of interest were from *Toxoplasma* then 349 of the 613 are likely to be human proteins. Of this 331, 121 (35%) individual spots were identified as human proteins. These 121 human spots consisted of 59 unique proteins. 462 spots were picked from non-infected gels and 202 were identified (47%). Of these 202 spots there were 85 unique proteins. Therefore these 24h experiments identified 59 individual proteins up regulated due to infection and 63 proteins down regulated due to infection in the host cell.

3.3.3.1 Proteins modulated 16h post invasion by *Toxoplasma gondii*

HFFs were infected at a ratio of 10 parasites: 1 cell and left for 16h at 5%CO_2, 37°C. The cells were harvested and the proteins separated by 2-DE on pH4-7 IPG strips. In the 16h infection experiment replicate 1, only spots that matched between the two gels and had a volume difference of 30% or more were chosen for MS analysis. 23 spots had a decreased volume in infected cells and 23 spots had an increased volume in infected cells. The gels were analysed using the 2DElite software and the proteins were identified using a combination of MALDI and MS/MS. Of the 46 differentially expressed spots, 43 were identified by MS. Figure 3-14 shows the annotated 16h non-infected gel image. Spots were picked which were present on both the infected and non-infected gel therefore any proteins boxed in green were up regulated in 16h infected cells. Proteins boxed in red were down regulated during infection with *T.gondii*. Proteins identified as from *T.gondii* are boxed in yellow. An excerpt from the MS results from the 16h infection is in Table 3-5. The full results table can be seen in Appendix II.

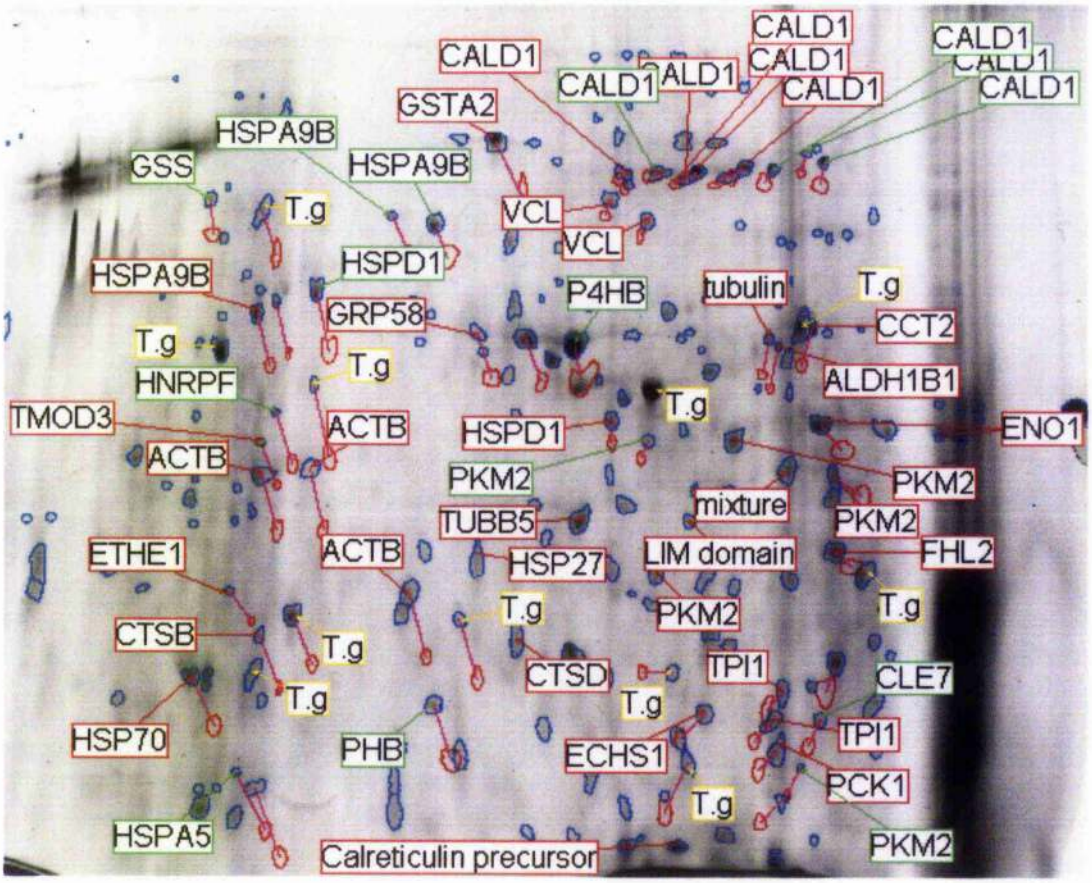


Figure 3-14 Annotated proteome map of 16h non-infected cells. The protein names boxed in red were down regulated 16h post infection and the proteins boxed in green were up regulated 16h post infection. A blue spot represents the outline of the spot in non-infected cells. The red outline represents the spot on the infected gel. Matching spots are connected by a purple line. Proteins identified as from *T.gondii* are boxed in yellow.

1	2	3	4	5	6	7	8	9	10	11	12	13
UP/DOWN	16hnon	no of hits	MS	T.g/H	Score	Protein	Gene name	Mass	pI	PM	s.c	gI
DOWN	131	1 of 1	MALDI	H	75	Actin Beta	ACTB	41321	5.56	7	24	14250401
DOWN	174	1 of 1	MALDI	H	65	Actin Beta	ACTB	41321	5.56	5	16	14250401
UP	42	1 of 1	MS/MS	H	253	MTHSP75	HSPA9B	74019	5.97	7	14	292059
UP	45	1 of 1	MALDI	H	76	MTHSP75	HSPA9B	74019	5.97	6	11	292059
UP	123	1 of 1	MS/MS	H	243	Pyruvate kinase	PKM2	58534	7.95	8	19	338827
UP	132	1 of 3	MS/MS	H	510	Pyruvate kinase	PKM2	58411	7.58	12	27	35505
UP	219	1 of 1	MALDI	H	82	Pyruvate kinase	PKM2	25454	5.77	4	28	9954651
DOWN	208	1 of 1	MALDI	H	112	TPI chain A	TPI1	26807	6.51	19	43	999892
DOWN	200	1 of 1	MALDI	H	83	Triose phospate isomerase Chain A	TPI1	26894	7.01	6	27	136066

14	15	16	17	18	19	20	21	22	23	24
MASCOT URL	gInumber	EMBL	Function	GO	GO biological process	vol inf	vol non	volume diff	Funcat	
http://fun-gen1	http://fun-g	http://harve	actins are highl	http://goda	cell motility/cytoskeleton	2,881,110	4,460,296	54.8	16	structural
http://fun-gen1	http://fun-g	http://harve	actins are highl	http://goda	cell motility/cytoskeleton	1929095	5483375	184.2	16	structural
http://fun-gen1	http://fun-g	http://harve	implicated in th	http://goda	ATP binding/chaperone activity	4281143	505932	-88.2	6	protein fate
http://fun-gen1	http://fun-g	http://harve	implicated in th	http://goda	ATP binding/chaperone activity	23,597,447	8,196,362	-65.3	6	protein fate
http://fun-gen1	http://fun-g	http://harve	that catalyzes	http://goda	glycolysis	1624649	277331	-21.8	2	glycolysis
http://fun-gen1	http://fun-g	http://harve	that catalyzes	http://goda	glycolysis	3516622	2173399	-38.2	2	glycolysis
http://fun-gen1	http://fun-g	http://harve	that catalyzes	http://goda	glycolysis	774,467	360,303	-53.5	2	glycolysis
http://fun-gen1	http://fun-g	http://harve	Plays an impor	http://goda	triose phosphate isomerase activity	21640998	4241396	96	2	glycolysis
http://fun-gen1	http://fun-g	http://harve	Plays an impor	http://goda	triose phosphate isomerase activity	2068991	3681813	78	2	glycolysis

Table 3-5 An exert from the 16h infection experiment replicate 1 results table. Figure legend: Column 1= spot is up or down regulated on infected gel, 2= spot picklist number, 3= number of significant hits from that spot, 4= protein identified using MALDI or MS/MS, 5= protein is human or from *Toxoplasma gondii*, 6= MASCOT® score, 7= Protein id., 8= gene name, 9= Mass, 10= pI, 11= peptides matched (P.M), 12= sequence coverage (S.c), 13= gI number, 14= MASCOT® url, 15= gI number url, 16= EMBL link, 17= protein function, 18= gene ontology url, 19= GO biological process, 20= volume of spot on infected gel, 21= volume of spot on non-infected gel, 22= FunCat number, 23= FunCat category.

3.3.3.2 Proteins modulated 12h post invasion by *Toxoplasma gondii*

HFFs were infected at a ratio of 10 parasites: 1 cell and left for 12h at 5%CO_2, 37°C. The cells were harvested and the proteins separated by 2-DE on pH4-7 IPG strips. Figure 3-15 and Figure 3-16 show the non infected and infected gels respectively.

The spots were matched between the two gels using the 2DElite software. Background subtraction and spot normalisation was carried out. Any spots whose volume differed by 30% or more were marked of interest. 89 spots had a volume increase of 30% or more in the infected cells, and 118 spots had a decreased volume of 30% or more, in comparison to the non-infected gel. These spots were excised from the gel, digested using the workstation and analysed by MALDI. Unfortunately it was only possible to identify 8 of the spots picked. This was due to a fault with the Ettan Spot Handling workstation. The results can be seen in Table 3-6. As so few spot were identified, an annotated image has not been included in this Chapter.

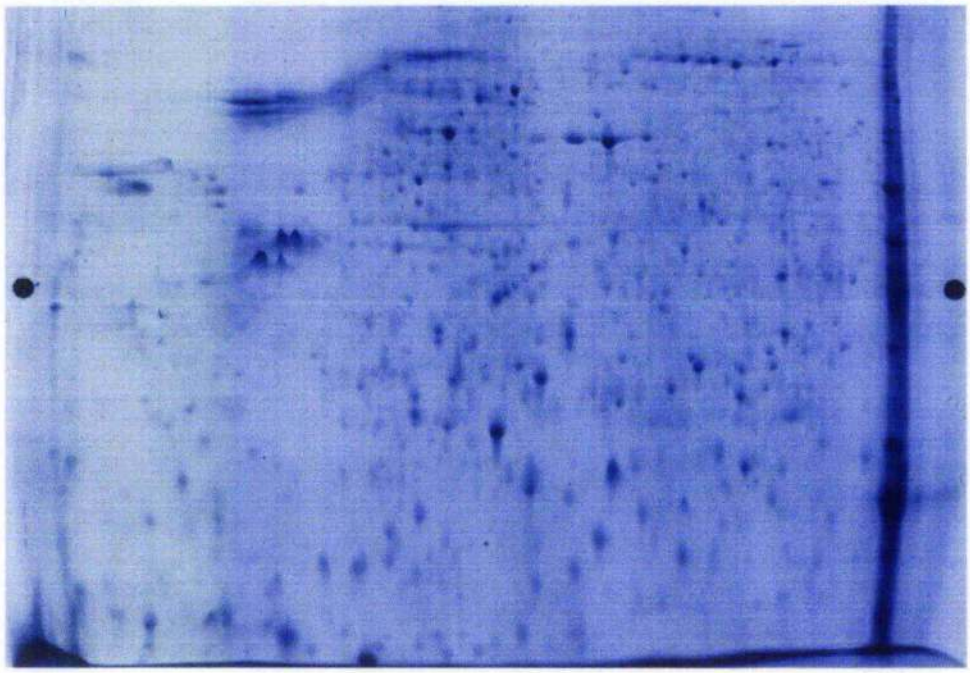


Figure 3-15 Proteome map of non infected HFFs from 12h infection experiment replicate 2. Colloidal Coomassie stained gel of whole cell extract from non-infected HFFs.

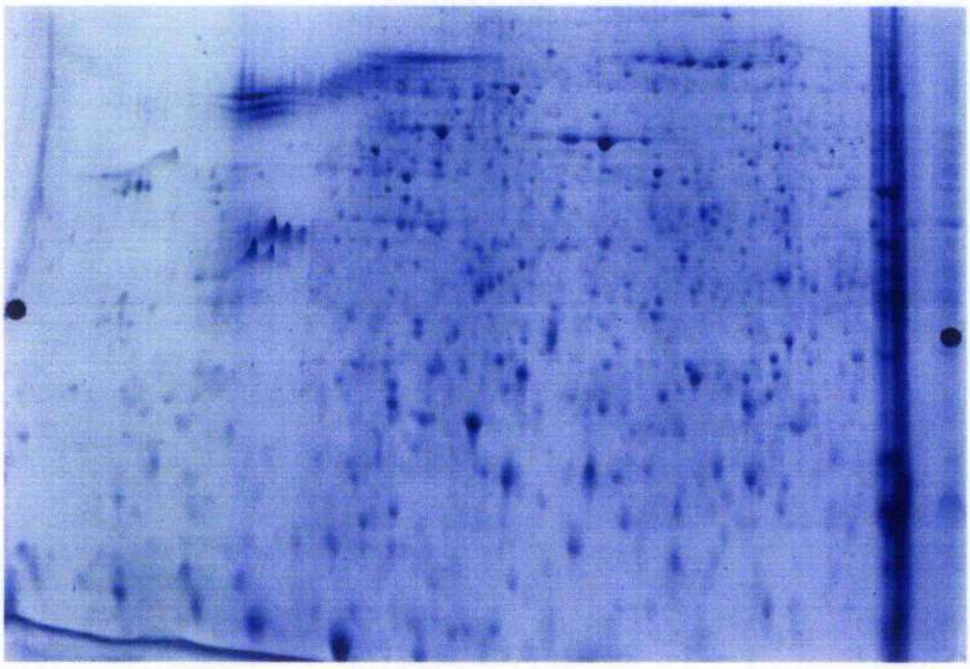


Figure 3-16 Proteome map of infected HFFs from 12h infection experiment replicate 2. Colloidal Coomassie stained gel of whole cell extract from infected HFFs.

	Gel	Spot no.	Score	Protein	Volume INF	Normalised volume INF	Volume NON	Normalised volume NON	Difference
1	morag12h6_191005	57	80	l-caldesmon II	6,526,989.00	1,471.19	4292229	1,202.48	18
2	morag12h6_191005	122	114	Vimentin	905,130.00	204.017	2109911	591.099	65
3	morag12h6_191005	128	90	Heat shock cognate 71	2,840,501.00	640.252	3941843	1,104.32	42
4	morag12h6_191005	129	137	Vimentin	10,644,562.00	2,399.30	13730931	3,846.77	38
5	morag12h6_191005	143	144	Vimentin	6,673,351.00	1,504.18	4200768	1,006.86	33
6	morag12h6_191005	269	72	HNRPH1	1,582,797.00	356.764	971767	24472.244	32
7	morag12h6_191005	289	74	HNRPF	1,798,369.00	405.355	1230004	244.59	40
8	morag12h6_191005	324	65	ACTG1	1,293,076.00	291.461	2212006	619.702	53

Table 3-6 Volume differences of the protein spots that matched across the two gels in the 12h infection experiment that were subsequently identified using MS. This table shows the spot number, the MASCOT® score, the protein identified, the spot volume, normalised volume and the difference in volume of the matched spots between the two gels (far right column).

3.3.3.3 Proteins modulated 6h post invasion by *Toxoplasma gondii*

HFFs were infected at an infection ratio of 10 parasites: 1 cell and left for 6h at 5%CO_2, 37°C. After 6h cells did not appear to be very heavily infected and many single parasites were observed within the cells. The cells were harvested and the proteins separated by 2-DE on pH4-7 IPG strips. The gels were stained with colloidal Coomassie and analysed using the 2DElite software. In comparison to the 24h, 16h and 12h infection experiments, not many changes were seen. This was to be expected as there were not many changes in the gene expression of the infected host cells by MA analysis at 4hpi.

The spots were matched between the two gels using the 2DElite software. Background subtraction and spot normalisation was carried out. Any spots whose volume differed by 30% or more were marked of interest. 59 spots had a volume increase of 30% or more in the infected cells, and 23 spots had a decreased volume of 30% or more, in comparison to the non-infected gel. These spots were excised from the gel and digested using the workstation, and analysed by MALDI. Unfortunately it was not possible to identify any of the spots picked, perhaps again due to a technical error during spot picking.

3.3.4 Analysis of differential protein expression across time

Table 3-7 compares the differentially expressed proteins modulated across time in two or more of the time points. The table contains information on protein and gene name, whether the protein is up regulated, down regulated or modulated (e.g. was identified as up and down regulated in the same experiment, therefore an ambiguous result) in infected cells and whether the expression response correlates across the time course. Proteins in green correlate, protein in red do not correlate and proteins in orange it is not possible to determine. Of the proteins which do not correlate across time it is interesting to note that caldesmon is down regulated 12hpi, modulated 16hpi and then up regulated 24hpi. The expression of this protein may therefore be induced upon parasitisation as part of the host or parasite survival response.

Protein name	Gene name	12h	16h	24h	correlates across time course
Actin	ACTB		down	down	yes
Aldolase A	ALDOA		up	up	yes
BiP protein	HSPA5		up	up	yes
Cathepsin B	CTSB		down	down	yes
Chaperonin HSP60	HSPD1		modulated	modulated	yes
Glutathione-S-Transferase chain A	GSTA2		down	down	yes
Heterogeneous ribonucleoprotein H2	HNRPH2	down		down	yes
Protein disulfide isomerase	P4HB		up	up	yes
Caldesmon	CALD1	down	modulated	up	no
Carboxyl terminal LIM domain protein	PDLIM1		up	down	no
Chaperonin containing TCP1 subunit 2beta	CCT2		up	down	no
Heterogeneous nuclear ribonucleoprotein F	HNRPF	down	up	down	no
Caldesmon	CALD1	down	modulated	modulated	not determined
Enolase 1	ENO1		down	modulated	not determined
HS70kd pro 42kd ATPase N-terminal domain	HSPA8	up		modulated	not determined
HSP -MTHSP75	HSPA9B		modulated	up	not determined
Prohibitin	PHB		up	modulated	not determined
Pyruvate kinase M2 isozyme	PKM2		up	modulated	not determined
Triose phosphate isomerase	TPI1		down	modulated	not determined
Vimentin	VIM	modulated		down	not determined
Vinculin	VCL		down	modulated	not determined

Table 3-7. Analysis of protein modulation across time. This table compares the differentially expressed proteins modulated in one more of the time points post infection; 12h, 16h and 24h. The table contains information on protein and gene name, whether the protein is up regulated, down regulated or modulated (e.g. was identified as up and down regulated in the same experiment) in infected cells and whether the expression response correlates across the time course. Proteins in green correlate, proteins in red do not correlate and proteins in orange it was not possible to determine.

3.4 Discussion

The method of conventional gel to gel 2-DE discussed in this chapter proved a powerful tool for analysing the protein expression of pathogen infected cells. As the gene to protein correlation remains unknown, a proteomic analysis of the *Toxoplasma gondii* infected cells will provide a more in depth understanding of the functional proteins and pathways involved in the host response. The wide range of time-points was chosen so as to be comparable to the gene expression studies (Blader *et al.*, 2001). Ideally all time points post invasion through to lysis would have been investigated but this was not possible in the time available. Infection experiments of 24h or less were carried out. It was not possible to carry out a time course of greater than 24h due to loss of host material through infected cell lysis.

Determining the optimal infection ratio was extremely important as too low a rate of infection resulted in few changes in protein expression and too high a rate of infection resulted in cell lysis and loss of host cell material. A study of synchronised cell populations showed that when cells were moving through from G_0 phase, parasite attachment increased 3-fold. Although fibroblasts divide asynchronously, when confluent, most cells would be in G_0 (Black and Boothroyd, 2000). Three infection ratios were investigated; 2:1, 5:1 and 10:1 parasites: cell. At 24h the 10:1 infection ratio was determined to possibly be too high as cells were beginning to lyse at the end of the 24h infection resulting in loss of host material. There was also an increased amount of *Toxoplasma* protein appearing on the gels. At an infection ratio of 2:1 approximately 70% of the cells were determined to be infected 24hpi. This infection ratio ensured the majority of cells were infected but also that there was not too much parasite protein carry over to the 2-DE gels. Despite adding parasite material to the non-infected cells as a control measure, parasite proteins were present on the infected gel. These spots were discounted from the analysis.

To ensure reliable and reproducible results and to account for gel-to-gel variation, eleven replicates of the 24h time point were run. The inclusion of eleven replicates of the 24h infection experiment ensured that the protein expression changes detailed in this chapter are indeed due to true biological variation due to the presence of the intracellular parasite in the host cell and not due to experimental variation. Manual verification of all identified modulated proteins was carried out. The spot position of the modulated proteins on the gel was compared to the same position on all the other 24h gel images to confirm that the same change was seen and the modulation was not due to experimental error. The 24h replicate 11 was analysed most thoroughly. All spots of interest were analysed firstly by MALDI

and if it was not possible to identify the protein then the peptides were analysed by MS/MS. MS/MS was used only in replicate 11 due to time and financial constraints.

The differential protein expression of 85 individual proteins was identified 24hpi using the gel-to-gel technique. Of these 85 proteins, 46% were identified as down regulated, 39% as up regulated and 17% as modulated in the infected cell. Modulated means that the protein was identified as both up and down regulated in the parasite infected cell. Modulation is most likely due to the fact that there were often several isoforms of the same protein present on the gel due to PTMs. In general, MALDI is not sensitive enough to detect changes in isoform therefore different forms may be arbitrarily classified as the one protein leading to an ambiguous result. MS/MS can detect different protein isoforms if the modified peptide (e.g. a phosphopeptide) is one of the peptides identified. The function of all the differentially expressed proteins identified in the time course experiments and how their altered levels of expression are thought to be involved in the host pathogen response will be discussed in Chapter VIII.

The techniques employed in this chapter, which did not involve any pre-fractionation or enrichment of the sample before 2-DE, should hypothetically profile all proteins present in the HFF proteome. Clearly this is not possible as there are an estimated 30,000 genes present in the human genome which is hypothesised to produce about 90,000 proteins due to PTMs, and 24cm 2-DE generally profile around three thousand spots (Jenkins and Pennington, 2001). Despite this, gel-to-gel electrophoresis allows for the study of several thousand proteins at one time under different conditions and is a hypothesis generating technique. Disadvantages of this technique are the poor ability to separate membrane proteins, proteins with very small or large molecular weight, proteins with pI values less than 4 or greater than 9 and proteins present in small copy numbers which may be masked by more abundant proteins.

Some of the time course infection experiments gave very few protein identifications. This was most likely due to insufficient protein in the spot on the gel. Or in the case of 24h replicates 1-7 which were analysed only by MALDI, there may have been more than one protein species present in each spot and MALDI is not sensitive enough to process and identify data from a mixed spot. No proteins were identified from any of the spots picked from the 6h infection experiment. This was due to an error with the SpotHandling workstation and the samples were lost.

Proteins with similar molecular weight and pI can co-migrate on a 2-DE gel and often several proteins were identified from the one gel spot, therefore making it difficult to know which protein is responsible for the volume change seen during infection. To overcome this problem, fractionation of the sample or fractionation of the gels (e.g. using narrow pH ranges) would allow for increased spot resolution and accurate quantitation of individual protein spots. However this approach is not very practical and as yet, no groups have attempted this. The problem of mixed spots in 2-DE analysis is very under represented in the literature.

Several disadvantages are associated with gel-to-gel electrophoresis, including variability between 2-DE gels and IPGs strips. This can cause difficulties in the comparison between two samples of interest as much of the variation seen may be due to experimental variation and not true biological differences. This can result in many replicates having to be done. The software available to analyse the 2-DE gel images and calculate spot volume is not very sophisticated and is often open to user intervention. A more quantitative and reproducible 2-DE approach to investigate the host pathogen response would help elucidate more significant results.

Chapter IV Analysis of differential protein expression in *Toxoplasma gondii* infected cells by Difference Gel Electrophoresis

4.1 Introduction

Accurate quantitation of changes in protein expression remains one of the most important aspects of proteomics. It is thought that many of the biologically significant changes in protein expression in the infected cell will be small. Therefore quantitation of protein expression is a vital tool in understanding the mechanisms of the host cell response to invasion and infection (Guerrera and Kleiner, 2005).

Conventional gel-to-gel 2-DE faces several limitations including differences in IPG strips and second dimension gels which can result in difficulties in comparing samples run on separate gels. Moreover, the currently available staining methods including Coomassie and silver staining have either limited sensitivity or linearity respectively (Van den Bergh and Arckens, 2004). A technique that compensates for or avoids these problems is therefore highly desirable. The recent development of Difference Gel Electrophoresis (DIGE) is able to overcome many of the problems of conventional 2-DE and is now beginning to be used widely to analyse differential protein expression (Karp *et al.*, 2004; Kernec *et al.*, 2001; Van den Bergh and Arckens, 2004). A study by Alfonso *et al* in 2005 investigated the proteomic expression analysis of colorectal cancer by DIGE. The premise behind this study was that using this technique would allow the authors to detect specific protein markers for this type of cancer and gain a better understanding into the mechanisms behind colorectal cancer progression. DIGE proved an extremely useful technique for this type of analysis as the study identified 41 proteins with statistically significant volume differences in cancerous cells. An ontology analysis of the modulated proteins showed that proteins that differed in expression in the cancerous cell were involved in structural, signal transduction, cell communication and protein synthesis functions. With further studies the authors hope to define a set of proteins as biomarkers for colorectal cancer (Alfonso *et al.*, 2005).

Variation in the context of 2-DE can be classed into three groups; (1) experimental, e.g. differences in IPG strips or second dimension gels when comparing two samples, DIGE removes this type of variation as samples are run on the same gel, (2) inherent/sample

variation, not all cells will behave similarly under the same conditions, this is unavoidable and (3) biological, this is the variation of interest, e.g. which proteins are up or down regulated in the infected cell.

DIGE was first described in 1997 by Unlu *et al* (Unlu *et al.*, 1997). The premise of DIGE is that the two samples of interest are labelled with different fluorescent dyes, mixed together and run on the same 2-DE gel. This removes the problems associated with gel-to-gel variation and allows for the detection and quantitation of differences in protein abundance between samples of interest. A third sample, the internal standard, is run on every gel which allows for accurate gel to gel matching and detection of minor differences in protein expression with statistical confidence. This method of comparing the protein expression of two samples of interest on the one gel using DIGE provides an equivalent technique to microarray analysis (Alfonso *et al.*, 2005).

A standard DIGE workflow is illustrated in Figure 4-1. To date, no studies have used DIGE to investigate the host response to infection with an apicomplexan parasite. A study by Jiang *et al* (2005) compared DIGE and ICAT in investigating the host cell response to infection with severe acute respiratory syndrome (SARS). Using these two methods they found 186 proteins were differentially expressed by 1.5 fold or more. One third of the proteins were down regulated and two thirds up regulated during infection. Interestingly the up regulated proteins were located mainly in the nuclei, whereas the down regulated proteins were distributed within the cell. The proteins were classified by function, more than half of the down regulated proteins were involved in the enzymatic reactions, signal transduction, immune responses and actin networks (Jiang *et al.*, 2005).

In this chapter HFFs were infected with *Toxoplasma gondii* and differential protein expression was investigated using DIGE at two time points after invasion. Investigation of the protein response 24hpi using DIGE allowed for a direct comparison of the DIGE technique with the traditional gel-to-gel 2-DE method used in Chapter III. This data was then compared with the gene expression data from the microarray experiments of Blader *et al.* (2001) (Chapter V).

The two samples of interest are labelled with a CyDye (Cy3 or Cy5) and the internal standard labelled with Cy2. DIGE fluor minimal dyes have an NHS-ester group which can form a covalent bond with the epsilon amino group of lysine residues in proteins via an amide linkage. All three CyDyes are size and charge matched ensuring equal migration on the 2-DE gel. The amount of dye used in the protein labelling reaction is limiting, leading

to the labelling of approximately 1-2% of lysine residues. This results in the labelling of only a small proportion of the total protein (~3-5% (Westermeier and Marouga, 2005)) in a sample and is termed minimal labelling. CyDyes are very sensitive and able to achieve detection down to 125 pg of a single protein. In addition, they show a linear response to protein concentration of up to five orders of magnitude. By contrast, silver stain detects only 1-60 ng of protein with a dynamic range of less than two orders of magnitude (Ettan DIGE System User Manual Edition AA). The three labelled samples are mixed together and separated by 2-DE and the gel is then scanned at the three different excitation and emission wavelengths unique to each CyDye. The fluorescence imaging system is capable of detecting proteins at a sensitivity at least equal to silver staining (Unlu *et al*, 1997). The DeCyder software matches the CyDye labelled samples on each gel, calculates the spot volume and also matches spot maps across gel sets. Differentially expressed protein spots are then excised from the gel and subjected to mass spectrometry for identification. Gel multiplexing where spots are compared in more than one gel is shown in Figure 4-2.

The inclusion of the internal standard is critical for a multiple gel DIGE experiment. It allows for accurate gel-to-gel matching and therefore increases statistical confidence. The standard is created by pooling an aliquot of all biological samples in the experiment which is then labelled with Cy2 and run on every single DIGE gel in the experiment. Linking every sample in-gel to a common internal standard offers several advantages; accurate spot quantification and accurate spot statistics between gels, increased confidence in inter-gel matching, separation of induced biological change (e.g. the protein expression differences between the cells of interest) from system variation (e.g. user intervention in data analysis and interpretation, differences between IPG strips and second dimension gels) and inherent biological variation (Ettan DIGE System User Manual Edition AA). Using appropriate experimental design and statistical analysis DIGE can help to differentiate between inherent and induced variation and will allow for accurate profiling of the host cell proteome during infection with *Toxoplasma gondii*.

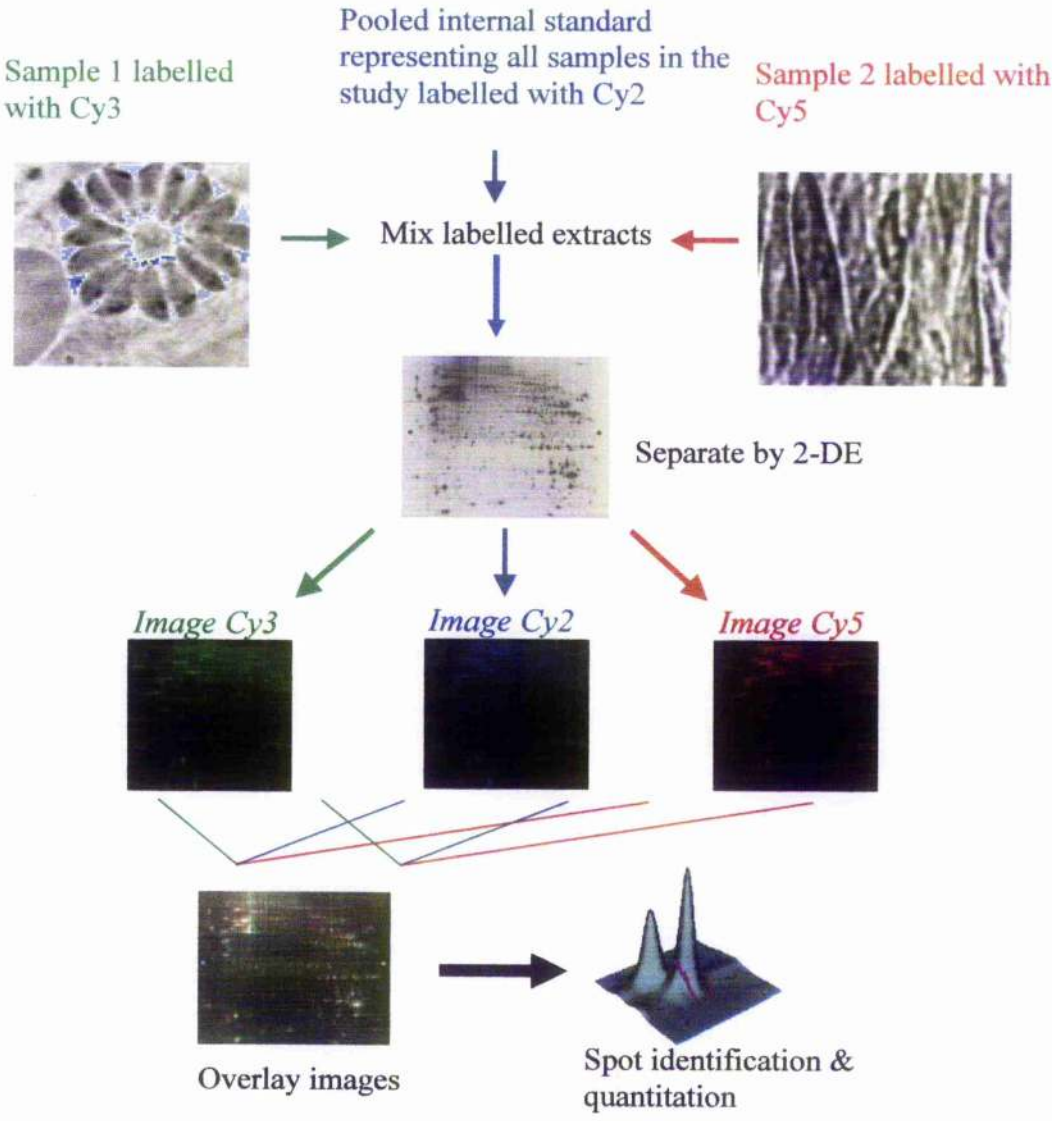


Figure 4-1 A schematic of the DIGE process. The two samples of interest (in this example, infected and non-infected samples) are labelled with Cy3 and Cy5 respectively. A third sample, the pooled internal standard containing every protein in the experiment is labelled with Cy2. The three labelled samples are mixed together and separated using 2-DE on the one gel. The gel is then scanned at the three wavelengths corresponding to the CyDye wavelengths. The spot maps are then analysed using the DeCyder software and differentially expressed spots determined.

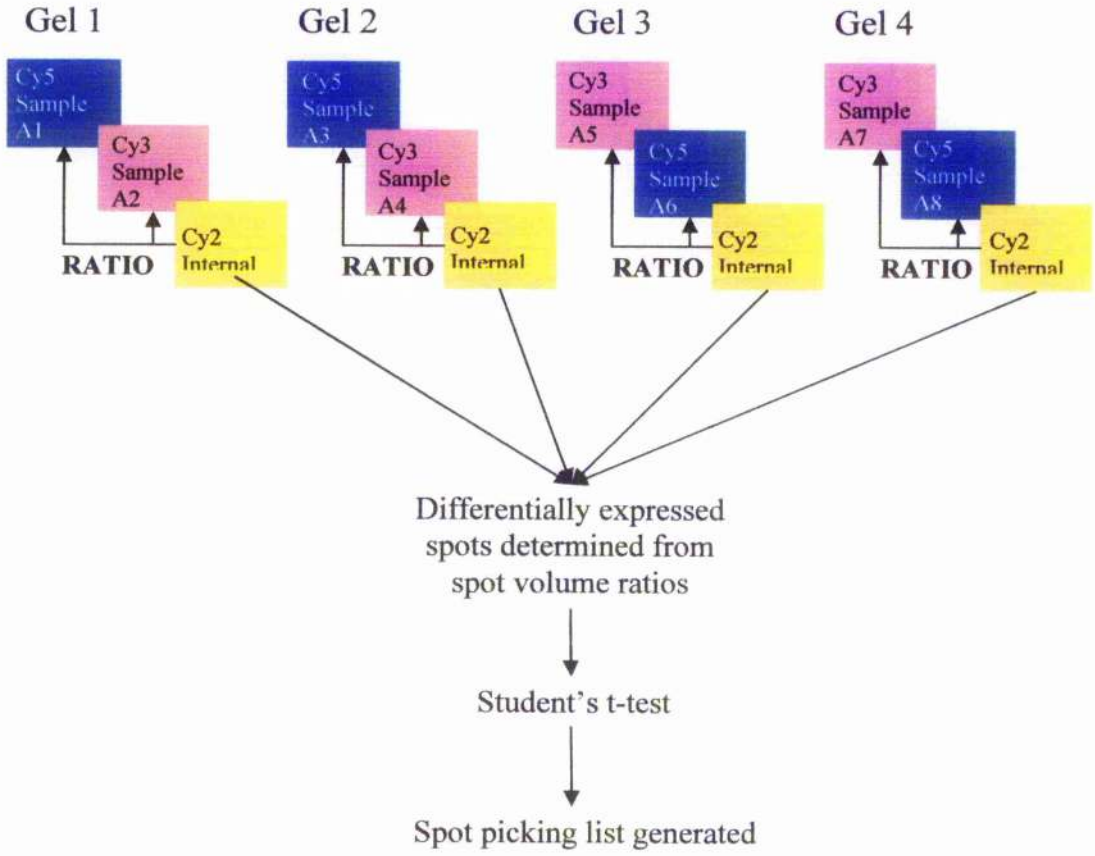


Figure 4-2 Intra gel matching with DIGE. The spot volume ratio between matched spots from the Cy5 or Cy3 gel versus the Cy2 labelled gel are calculated. The spot ratios from all of the gels are then compared and the spot volume differences calculated. Using the Students t-test, spots with a p-value of <0.05 are deemed significant and a picklist of significant spots is generated.

4.2 Materials and Methods

4.2.1 Infection of HFFs with Toxoplasma gondii

Infection experiments were performed as detailed in Chapter III. Protein expression of infected host cells at 24hpi and 12hpi was investigated by five individual 24h infection and four sets of 12h infection experiments. An infection ratio of 2:1 parasites: HFF was used. To all samples 200µl DIGE lysis buffer (25mM Tris, 8M Urea, 4% CHAPS), 10µl DN/RNases and 5µl protease inhibitors were added. The samples were then frozen in liquid nitrogen, thawed at 29°C and mixed by vortexing. This was repeated five times in total. The cell lysate was then centrifuged at 12,000 g, 4°C for 10min. As with conventional gel-to-gel electrophoresis, DIGE is incompatible with contaminants such as excess salts or lipids therefore the sample must be desalted and delipidated prior to CyDye labelling as is described below.

4.2.2 PlusOne™ 2-D Clean-Up Kit

PlusOne™ 2-D Clean-Up Kit (Amersham) is designed to prepare samples for 2-DE that otherwise would produce poor 2-DE results due to high conductivity, high levels of interfering substances or low concentration of protein. The kit precipitates the proteins and leaves behind in solution interfering substances such as detergents, salts, lipids, phenolics and nucleic acids. The proteins are then resuspended in DIGE lysis buffer ready for IEF.

The clean-up kit is compatible with up to 100µg protein in 100µl solution and all steps were carried out on ice unless otherwise stated. To the sample, 300µl of precipitation solution were added and the sample vortexed and incubated on ice for 15min. Following this, 300µl of co-precipitant were added and the sample centrifuged at 12,000 g for 5min. The supernatant was removed and the sample centrifuged briefly to collect any remaining supernatant. Without disturbing the pellet, 40µl of co-precipitant were added on top of the pellet and the tube left on ice for 5min. Sample was centrifuged at 12,000 g for 5min and the liquid discarded. ddH_2O (25µl) was added to the pellet and vortexed to disperse the pellet. Wash buffer (1ml, pre-chilled at -20°C for 1h) and 5µl wash additive were added to the pellet and the sample left for 30min at -20°C with vortexing for 20-30sec every 10min. The sample was then centrifuged at 12,000 g for 5min. The supernatant was discarded and the pellet left to air-dry for a maximum of 5min. The pellet was resuspended in a small volume of DIGE lysis buffer (8M urea, 4%CHAPS, 25mM Tris Base) and the protein

concentration determined using the Bradford method. The optimal protein concentration for DIGE labelling is 5mg protein/ml DIGE lysis buffer and the protein concentration of the samples was adjusted accordingly using DIGE lysis buffer.

4.2.3 DIGE labelling

For optimal CyDye labelling the pH of the sample must be alkaline and lie between pH8 and 9. Samples were tested on pH strips and the pH was always found to lie between pH8 and 9. pH can be adjusted by adding 1M NaOH.

A pooled sample named the internal standard was generated for the 12h and 24h experiments. This standard contained proteins from all of the samples of interest in the study. Table 4-1 shows which samples were labelled with which CyDyes in the 24h experiment (A) and the 12h experiment (B). To 50µg of sample in 1µl DIGE lysis buffer, 1µl (400pmoles/µl) of appropriate working dye (Cy3/Cy5/Cy2) was added and vortexed to mix, centrifuged briefly and left on ice for 30min in the dark.

To stop the labelling reaction, 1µl 10mM lysine was added to the sample, centrifuged briefly and left on ice for 10min. The appropriate Cy2, Cy3 and Cy5 labelled samples were combined into a single microfuge tube and vortexed to mix. DIGE lysis buffer (66µl) was added and the sample was left on ice for 10min. The volume was made up to 460µl using rehydration buffer (containing 10mgDTT/ml rehydration solution, 5µl IPG buffer/ml rehydration solution) and the sample left on ice for 10min.

In the 12h and 24h experiments a preparative gel was run which contained 400µg of sample (pooled from the replicates) and also 50µg of Cy2 labelled internal standard. The inclusion of the standard on the preparative gel is to aid the matching of the preparative gel to the other spot maps.

Gel Number	Cy3	Cy5	Cy2
1	Infected 1	Non-infected 1	Pooled standard
2	Infected 2	Non-infected 2	Pooled standard
3	Non-infected 3	Infected 3	Pooled standard
4	Non-infected 4	Infected 4	Pooled standard
5	Infected 5	Non-infected 5	Pooled standard

Table 4-1 A. 24h infection experiment details. Five experimental gels were run, each containing a CyDye labelled infected and CyDye labelled non-infected sample and a Cy2 labelled pooled standard. Samples were dye swapped, i.e. three infected samples were labelled with Cy3 and two infected samples labelled with Cy5. This was to ensure that any differences seen between infected and non-infected samples were due to biological variation and not differences in the dye labelling process.

Gel Number	Cy3	Cy5	Cy2
1	Infected 1	Non-infected 1	Pooled standard
2	Infected 2	Non-infected 2	Pooled standard
3	Non-infected 3	Infected 3	Pooled standard
4	Non-infected 4	Infected 4	Pooled standard

Table 4-1 B. 12h infection experiment details. Four experimental gels were run, each containing a CyDye labelled infected and CyDye labelled non-infected sample and a Cy2 labelled pooled standard. Samples were dye swapped, i.e. two infected samples were labelled with Cy3 and two infected samples labelled with Cy5. This was to ensure that any differences seen between infected and non-infected samples were due to biological variation and not differences in the dye labelling process.

4.2.4 2-DE

First dimension IEF was performed as detailed in Chapter II. Following IEF the IPG strips (pH 4-7) were equilibrated for 15min in 10ml SDS-equilibration buffer containing 100mg DTT (65mM) then for 15min in 10ml SDS-equilibration buffer containing 250mg iodoacetamide.

Second dimension separation was modified from the protocol used in the earlier chapters as pre-cast plastic backed gels are incompatible with DIGE due to the plastic backing fluorescing during scanning. The gels were therefore poured manually. All glass plates were soaked in 1% decon and cleaned thoroughly. The back plates were treated with bind-silane solution (80%(v/v) EtOH, 18%(v/v) ddH_20, 2%(v/v) acetic acid, 0.1%(v/v) Ymethacryloxypropyltrimethoxysilane), which sticks the gel to the glass plate, and left to dry for several hours. Reference markers were applied to the back glass plate. Gels were 12.5% SDS-PAGE gels (40% (w/v) acryalmide/bis, 1.5M Tris pH8.8, 10%(w/v) SDS, TEMED, 10%(w/v) APS). The gel solution containing only tris, ddH_20 and acrylamide was filtered (Stericup-GP filter unit, 0.22 μm, catalogue# SCGPU05RE). The SDS, TEMED and APS were added to the filtered solution and poured immediately into the casting unit. Each gel was overlaid with 1ml of ddH_20-saturated butanol.

Strips were placed onto the top of the gel and sealed into place using agarose sealing solution (0.5%(w/v) agarose, trace bromophenol blue). Gels were run at 120V overnight in 1x SDS-running buffer (1x SDS-running buffer: 25mM Tris, 192mM glycine, 2%(w/v) SDS) in the bottom chamber and 2x SDS-running buffer in the top chamber.

4.2.5 Gel imaging

Gels were imaged using the Typhoon Variable Mode 9400 Imager using the settings recommended by Amersham in the Ettan DIGE user manual. The gels were scanned at the three wavelengths which correspond to the three CyDyes as seen in Table 4-2.

Fluorophore	Emission Filter (nm)	Laser
Cy2	520 BP 40	Blue (488)
Cy3	580 BP 30	Green (532)
Cy5	670 BP 30	Red (633)
SYPRO Orange	580 BP 30	Green (532)

Table 4-2 The appropriate laser and wavelength for each CyDye scan using the Typhoon imager.

To perform the analytical scan, a 100μm pixel size was selected for quantitative scans. Post scan the gel image was cropped so that only the area of interest of the gel was included in the analysis.

4.2.6 SYPRO staining

To detect the total protein present on the preparative gel, the total protein stain SYPRO Orange was used. The gel was fixed in 40%MeOH, 10% acetic acid for 1h, probed using 0.05% SDS for 1h then washed briefly in 7.5% acetic acid. The gel was stained with SYPRO Orange (1 in 10,000 dilution in 7.5% acetic acid) for 1–2h. The gels were then washed briefly in 7.5% acetic acid and scanned using the green laser Cy3 (532nm, 580 BP 30) (Table 4-2).

4.2.7 Gel analysis using the DeCyder software

DeCyder is a fully automated image analysis software suite developed for detection, quantitation, positional matching and differential protein expression analysis on DIGE generated images. The software consists of four modules; (1) Differential In-gel Analysis (DIA): protein spot detection and quantitation on a pair of images from the same gel, (2) Biological Variation Analysis (BVA): matches multiple images from different gels to provide statistical data on differential expression levels between multiple groups, (3) Batch Processor: automated image detection and matching of multiple gels without user intervention and (4) XML Toolbox: used to generate user reports.

The DIA module is used to first calculate how many spots are likely to be present on each DIGE gel. The three CyDye images from one gel in the 24h experiment were loaded into the DIA module and all spots present on the gel detected. All spots in every gel were then detected using the batch processor. Telling the software how many spots to detect is the only user intervention required at this stage. The output files from the batch processor were then opened in the BVA module. The BVA module was used to carry out gel-to-gel matching of spots using the match detection algorithm allowing quantitative comparisons of protein expression across multiple gels. One of the Cy2 images was assigned master gel status and all other images were then matched to it, identifying common protein spots across the gels.

4.2.8 Statistical Analysis

DeCyder BVA possesses several statistical analysis methods which can be used to determine whether changes in volume of specific spots (e.g. expression of specific proteins) are significant between samples from different experimental groups. The statistical test relevant to this study was a Student's *t*-test. The Student's *t*-test was used to test the hypothesis that a variable, in this case protein expression between infected and non-infected cells, differs between two groups. The Student's *t*-test is an equal variance two-tailed test, therefore, direction of change (i.e. increases and decreases) in the standardised abundance parameter was calculated. The degree of difference in the standardised abundance between a protein spot group was expressed as the average ratio. The average ratio value indicates the standardised volume ratio between the two groups or populations. Values are displayed in the range of -1 000 000 to –1 and +1 to +1 000 000.

The Student's *t*-test null hypothesis was that there was no change in the protein abundance between infected and non-infected cells (i.e. the average ratio between two groups was 1). Therefore the *t*-test *p* value represented the probability of obtaining the observed data if the two groups had the same protein abundance. For example, if the *t*-test value between two groups was 0.01, then the probability of obtaining the observed difference in protein abundance by stochastic variation alone was 1 in 100. Protein abundance differences were generally assumed to be statistically significant when $p \leq 0.05$, which is the standard p-value cut off for biological data.

4.3 Results

DIGE technology was used to compare the protein expression profiles of non-infected and infected HFFs at 12h and 24h post infection. Several hundred spots were modulated in volume in each experiment and in turn several hundred proteins were identified as modulated in expression following infection of the cells with *Toxoplasma gondii*.

In the 24h DIGE experiment, 362 differentially expressed matched spots were present in 9 or more of the 15 gel images and had a *p*-value of 0.05 or less. 149 spots had a decreased volume and 213 had an increase in volume. This means that these spots differed in volume in a statistically significant manner and it is possible to conclude that they were definitely modulated in expression in the infected cell. Of these 362, 149 had a *p*-value of less than or equal to 0.01. In the 12h DIGE experiment, 177 spots were present in 9 or more of the 14 gel images with a p-value of ≤ 0.05. Of these 177 spots, 39 had a decreased volume in *T.gondii* infected cells and 138 had an increased volume.

CyDye molecules have molecular weights of approximately 500Da and are mass and charge matched. This ensures that a protein species labelled with any of the CyDye Fluor minimal dyes will migrate to the same position on a 2-DE gel. The dyes are also positively charged to compensate for the positive charge that is lost during conjugation of the dye to lysine, ensuring that the pI of the protein does not alter after labelling. Lysine is highly abundant in proteins representing the most appropriate labelling strategy to detect proteins present in a complex sample. Of the total lysine residues 1-2% are labelled thus a labelled protein species is visualised as a single protein spot (Van den Bergh and Arckens, 2004).

As a consequence of the infected, non-infected and pooled standard being resolved on the same gel, the spot maps were overlaid and compared directly. Automated spot detection and quantitation were performed on overlaid Cy2 (standard sample)/Cy3 (sample of interest 1) and Cy2 (standard sample)/Cy5 (sample of interest 2) image pairs from each gel, followed by gel-to-gel matching and statistical analysis. For a given spot, the standardised abundance was expressed as a volume ratio between the pooled standard and a co-detected sample from the same gel. DeCyder Differential Analysis Software compares how the abundance ratio measurements of the same protein spot from different samples and gels relate to the standard sample, which makes further normalisation unnecessary. Since each sample spot map is co-detected with a standard spot map, all of the spots are compared internally to the same pooled standard.

Figure 4-3 shows one of the DIGE gels from the 24h infection experiment. The colour image on the left is the composite scan of the three CyDye labelled samples overlaid. On the right are the three black and white scans corresponding to the Cy3 labelled non-infected extract, the Cy5 labelled infected extract and the Cy2 labelled internal standard.

The 24h DIGE experiment is used as an example of DIGE analysis in sections 4.3.1 – 4.3.6.

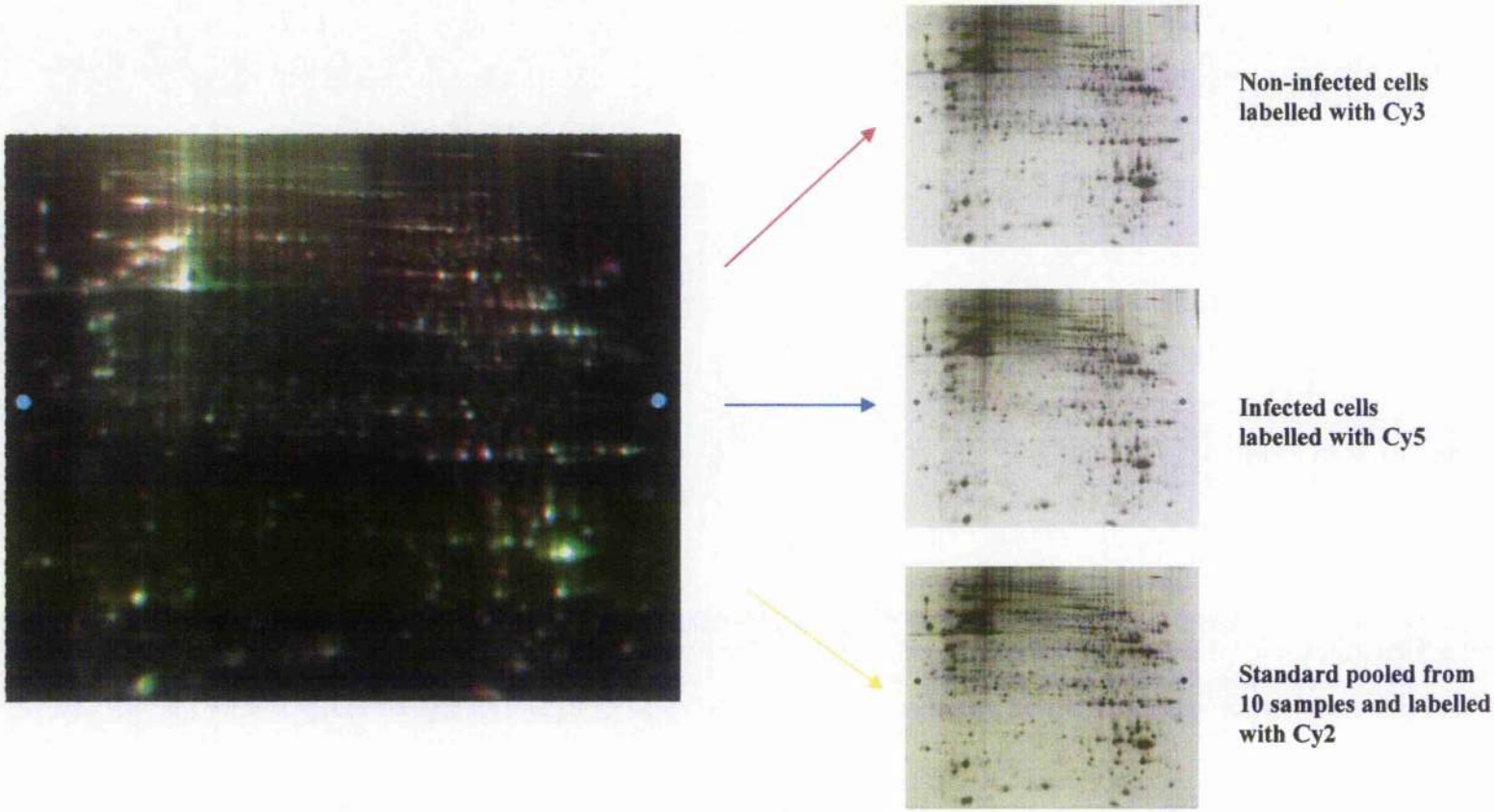


Figure 4-3 Differential protein expression analysis of *T.gondii* infected HFFs by DIGE. Protein extracts (50μg) from *T.gondii* infected and non-infected HFFs were differentially labelled with Cy5 and Cy3 fluorophores respectively. A mixed internal standard representing proteins from all 10 samples in this 24h infection experiment was labelled with Cy2 and included on all gels. Proteins were separated on pH3-10NL IPG strips in the 1st dimension. The colour image on the left is a composite picture of the three CyDye labelled samples overlaid. The three black and white images on the right are the Cy3, Cy5 and Cy2 labelled images separately. Several thousand spots were resolved on each image and with good resolution.

4.3.1 DIA module

As previously mentioned, the three CyDye gel images from one of the 24h infection experiments were loaded into the DIA module and the number of spots present on the gel detected. The volume, area, peak height and slope for each individual spot was calculated. Spot volumes (sum of pixel intensity within the spot boundary) were expressed with background subtraction. Background was subtracted on a spot specific basis, by excluding the lowest 10^{th} percentile pixel value on the spot boundary. Spot ratios were calculated (volume of secondary image spot/volume of primary image spot). This ratio indicated the change in spot volume between the two images. This ratio parameter was referred to as the volume ratio. The volume ratio is always expressed as 1 to 1 000 000 or –1 to –1 000 000. Values between –1 and 1 are not represented (Amersham Ettan DIGE user manual).

Figure 4-4 shows a screenshot of the DIA module. The gel images seen are a Cy3 labelled infected sample on the left, and a Cy5 labelled non-infected sample on the right. All spots were detected (2814) and any spots with a slope of 1.6 or greater were excluded (as these were dust particles). This left 2561 spots for analysis. A 2.0 fold difference in volume between the matching infected and non-infected spots was deemed to be significant in the DIA module. Spots shown in green had a volume ratio of less than 2.0 fold. Spots shown in red had a volume decrease of 2.0 fold or more in the infected sample. Spots shown in blue had an increased volume ratio of 2.0 fold or more in the infected sample. Using these parameters, 384 spots were decreased in volume and 54 were increased in volume in *Toxoplasma gondii* infected cells on this one gel. Therefore 17% of all the spots included for analysis changed in volume 2-fold or greater due to infection. Analysing one gel set with the DIA module showed how many spots were present on one of the gels. This information allowed for automatic detection of all spots on all the gels in the experiment using the batch processor. To apply the statistical tests to this data, all gels were then analysed in the BVA module.

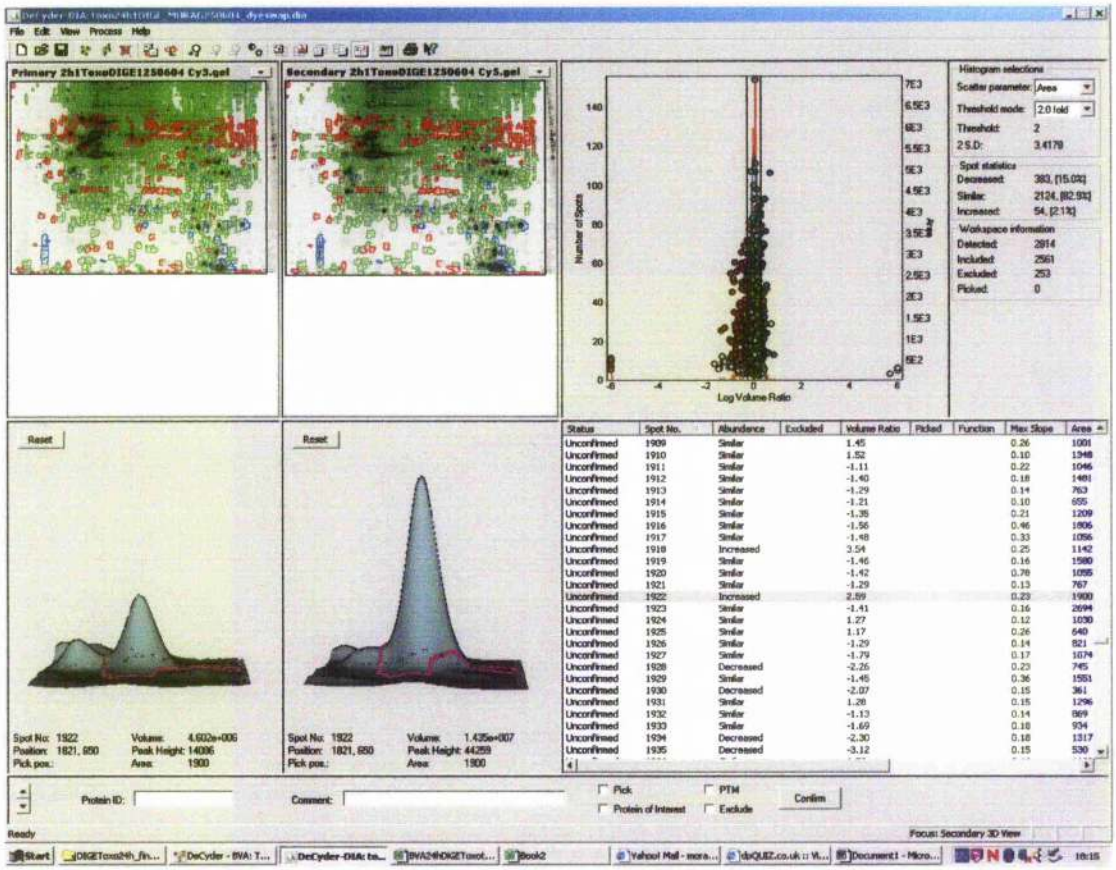


Figure 4-4 Screenshot of the DIA module using one of the 24h infection experiment gels as an example. The DIA window is split into four sections. Upper left is the image view which shows both the primary (Cy3, infected) and secondary (Cy5, non-infected) gel images. Spots shown in red are down regulated in infected cells; spots shown in blue are up regulated in infected cells. Spots shown in green are spots whose volume is not modulated more than 2-fold in the infected cell. Upper right is the histogram view which displays data associated with all detected spots in the primary and secondary images. The left y-axis displays the spot frequency. The red curve represents the frequency distribution of the log volume ratios. The blue curve represents a normalised model frequency fitted to the spot ratios so that the modal peak is zero. The right y-axis represents the scatter parameter selected in the histogram selection box (right of the histogram). A plotted single data point on the histogram represents an individual protein spot. Bottom left is the 3D view of a selected spot on the primary and secondary image. Bottom right is the table view which displays data associated with selected co-detected spots, including volume ratio and whether the spot volume is increased/decreased or unchanged.

4.3.2 Batch Processor

The Batch Processor links the DIA and BVA modules to perform all stages of the 2-DE DIGE analysis process. It is an automated process and therefore does not introduce further processing and analysis to the spot map data. All of the gel images in the experiment to be matched are loaded into the processor and the number of spots to be detected is entered. The processor then links each Cy3 and Cy5 image from a gel to the corresponding Cy2 image in that gel set. The output file is in XML format ready for use in BVA.

4.3.3 BVA module

The BVA module allows processing of multiple DIGE gels, performs spot matching between gels and allows quantitative comparison of protein expression across multiple gels. The software processes the gel image pairs that underwent spot detection in the Batch Processor module. One gel image is assigned master status and all the other CyDye images are then matched to the single master image, identifying common protein spots across the gels.

There are four main sections in BVA; (1) Spot Map Table, (2) Match Table, (3) Protein Table and (4) Appearance Table. The 24h infection experiment was analysed using these methods as follows.

The Spot Map table (Figure 4-5) was used to set up spot matching for matching and statistical analysis. The table lists all data related to the Spot Maps in the experiment imported from DIA and can be seen in detail in Table 4-3, including which group each gel belongs to, the number of spots detected on the gel and the number of spots matched to the master gel. The Match Table (Figure 4-6) was used for the processes associated with inter gel matching. To aid the matching algorithm, a few spots were matched manually, a process called landmarking, and the automated matching algorithms were carried out which matched all spots maps in the workspace. The Protein Table (Figure 4-7) displays and processes data associated with the protein spots identified across the gels. Each row of the table corresponds to one protein spot, which may be present in several Spot Maps. The Graph View (top right) shows the average ratio trend of the volume of this spot in the 5 gel sets in the 24h experiment. The Appearance Table (Figure 4-8) was used to display data associated with a single selected spot across the gels.

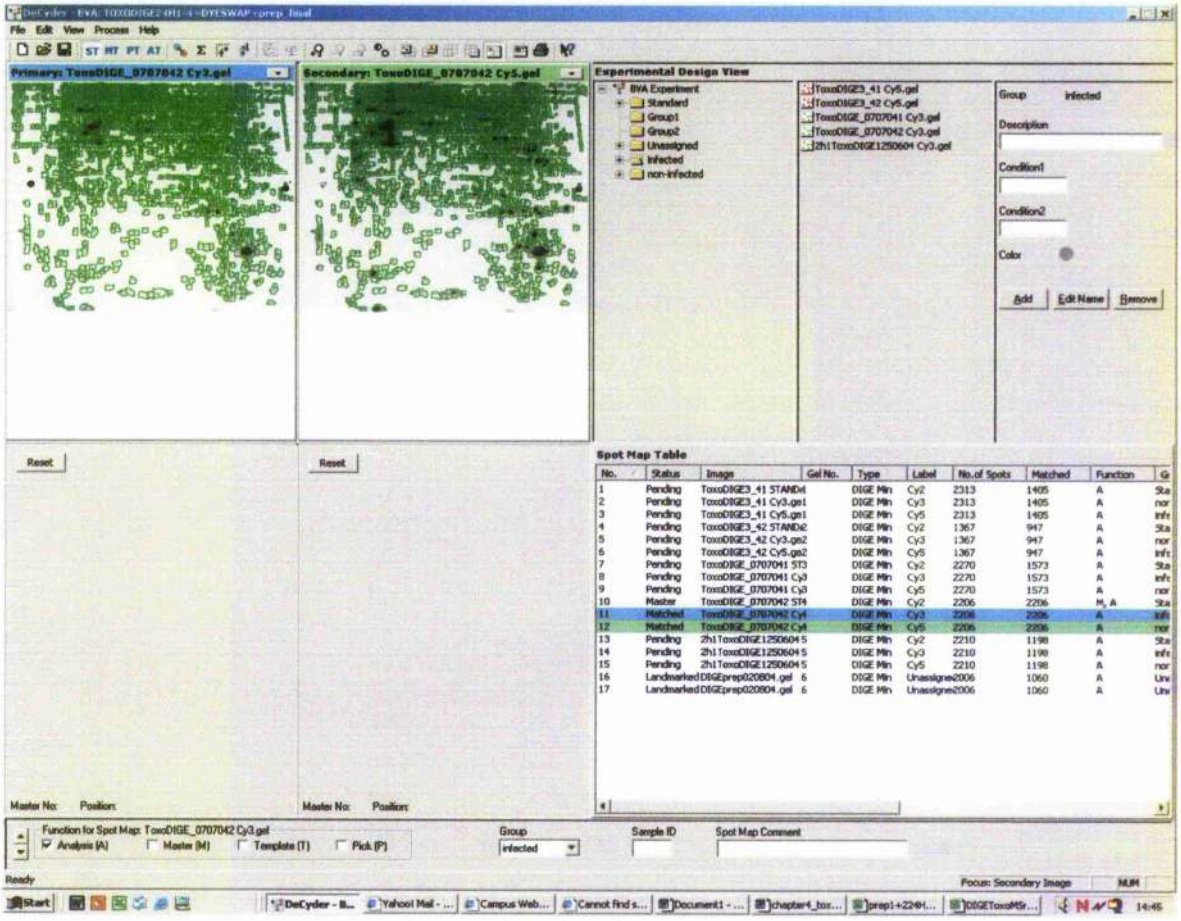


Figure 4-5 Screenshot of the BVA analysis software in Spot Map Table mode used to analyse the 24h infection experiment. The screen is split into 4 sections. Upper left is the image view which shows an infected Cy3 labelled image on the left and the corresponding non-infected Cy5 labelled image on the right. The spots detected using the DIA/Batch Processor module are shown in green. The second section of the screen (top right) contains information on all of the gels in the experiment, including whether the images are of infected or non-infected cells and which CyDye they are labelled with. The bottom right hand section is the Table View which contains more information about all of the gels in the experiment, including the file names of all the gel images, CyDye information, how many spots were detected in DIA, how many spots were matched to the Master gel, and which group (infected, non-infected or standard) the image belongs to.

Spot Map Table						
No.	**Image**	**Type**	**Label**	**No. Of Spots**	**Matched**	**Group**
1	Gel 1	DIGE Min	Cy2	2270	1573	standard
2	Gel 1	DIGE Min	Cy3	2270	1573	infected
3	Gel 1	DIGE Min	Cy5	2270	1573	non-infected
4	Gel 2	DIGE Min	Cy2	2206	2206	standard
5	Gel 2	DIGE Min	Cy3	2206	2206	infected
6	Gel 2	DIGE Min	Cy5	2206	2206	non-infected
7	Gel 3	DIGE Min	Cy2	2313	1405	standard
8	Gel 3	DIGE Min	Cy3	2313	1405	non-infected
9	Gel 3	DIGE Min	Cy5	2313	1405	infected
10	Gel 4	DIGE Min	Cy2	1367	947	standard
11	Gel 4	DIGE Min	Cy3	1367	947	non-infected
12	Gel 4	DIGE Min	Cy5	1367	947	infected
13	Gel 5	DIGE Min	Cy2	2210	1198	standard
14	Gel 5	DIGE Min	Cy3	2210	1198	infected
15	Gel 5	DIGE Min	Cy5	2210	1198	non-infected
16	preparative gel	DIGE Min	post stain	2006	1060	preparative
17	preparative gel	DIGE Min	post stain	2006	1060	preparative

Table 4-3 Spot Map Table from the BVA module. This table profiles the information from all gels in the experiment, including the CyDye label, the number of spots detected on the gel, the number of spots matched to the master, which gel has been assigned master status (M) and which group the gel belongs to; infected, non-infected or standard. The Cy2 labelled standard gel was assigned master status in this 24h DIGE experiment. Spot maps 16 and 17 are the preparative gel.

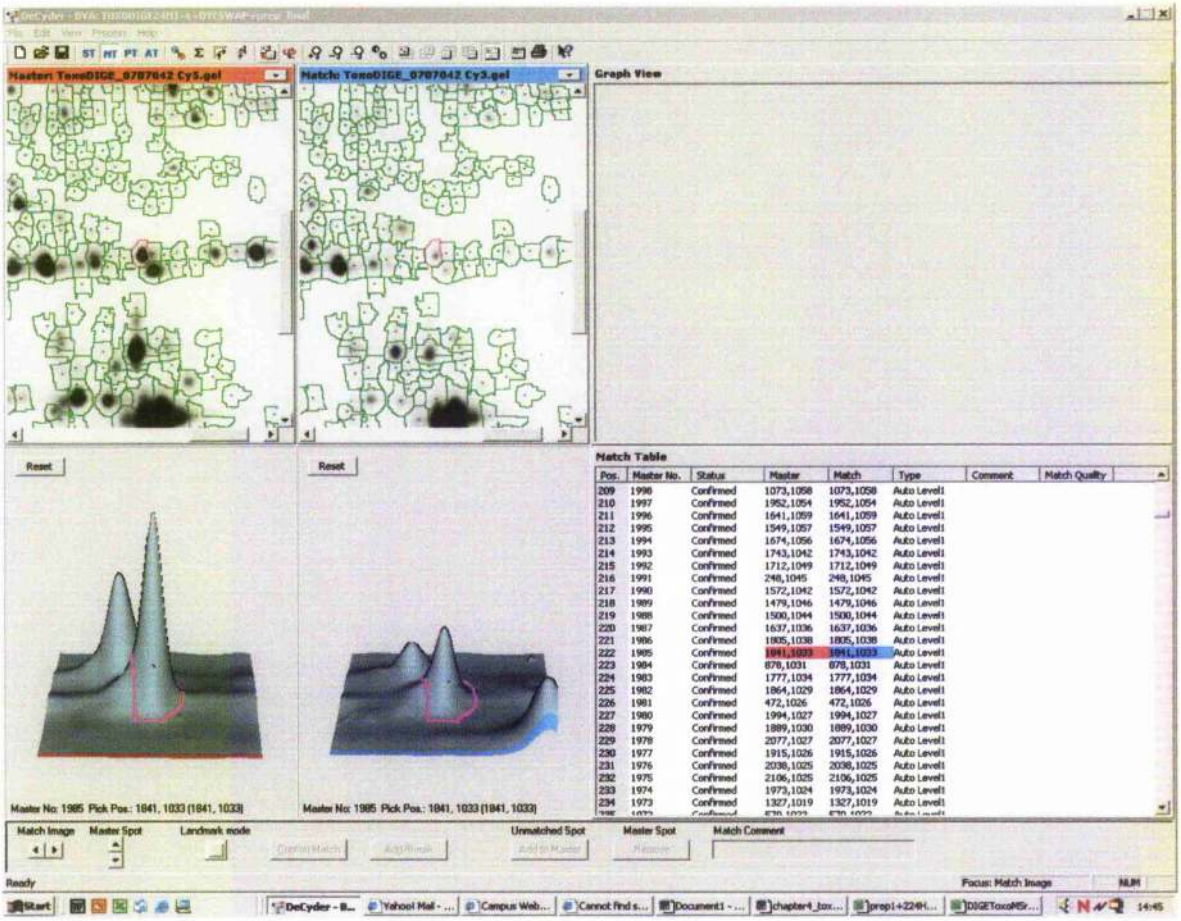


Figure 4-6 Screenshot of the BVA analysis software in the Match Table mode. This mode lists all data associated with inter-gel matching in the 24h infection experiment. This figure shows spot number 1985 matched between two gel images (non-infected Cy5 on the left and infected Cy3 on the right). The Image View can be magnified to show the area of interest of the gel in detail. Bottom left is the 3D View showing the abundance of the matched protein in the two images.

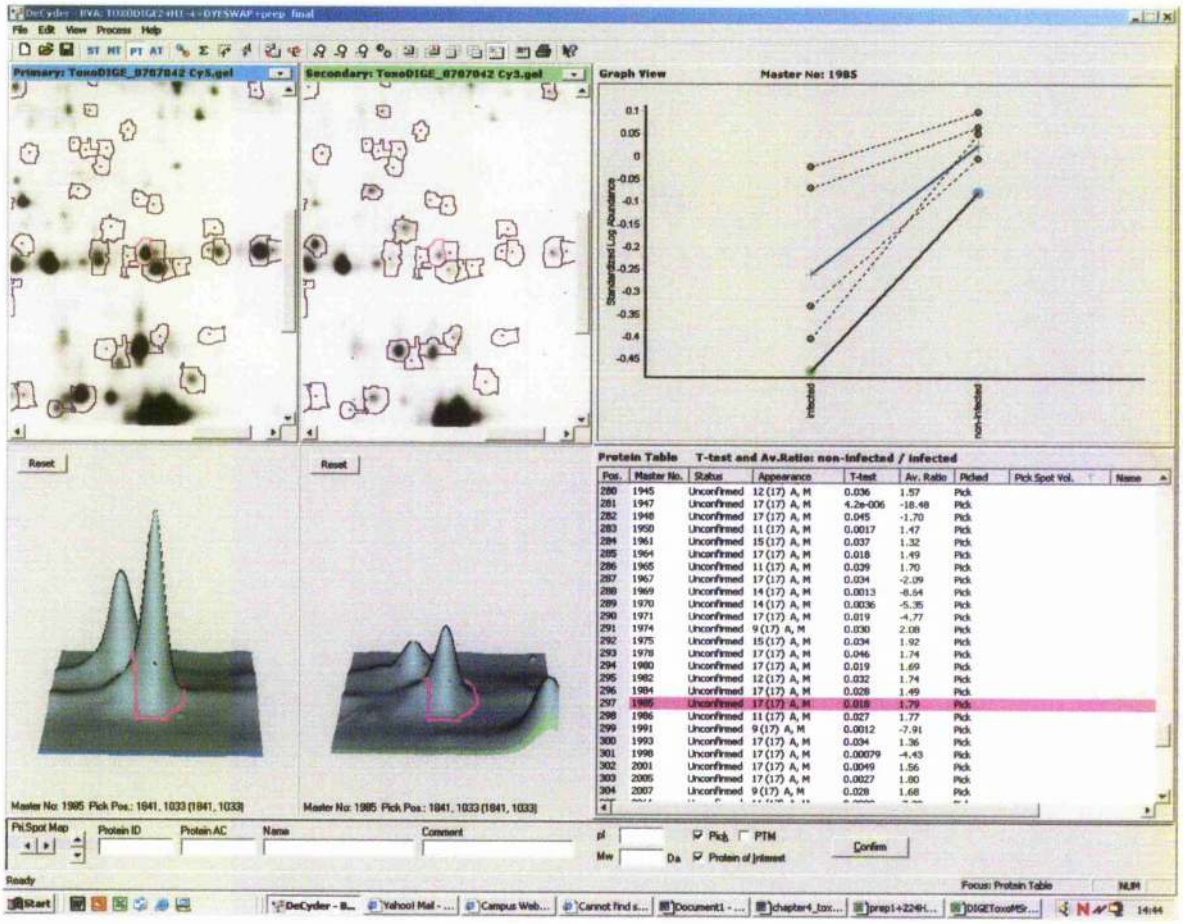


Figure 4-7 Screenshot of the BVA analysis software in the Protein Table mode. The protein match table displays spot specific information including the master number of the spot, how many gel images the spot has been detected in, the *t*-test value and the average ratio of spot volume. This screenshot shows spot number 1985 from the master gel. The graph in the top right section shows that the volume of this spot is consistently lower in *Toxoplasma gondii* infected cells than in non-infected cells (as measured by standardised log abundance).

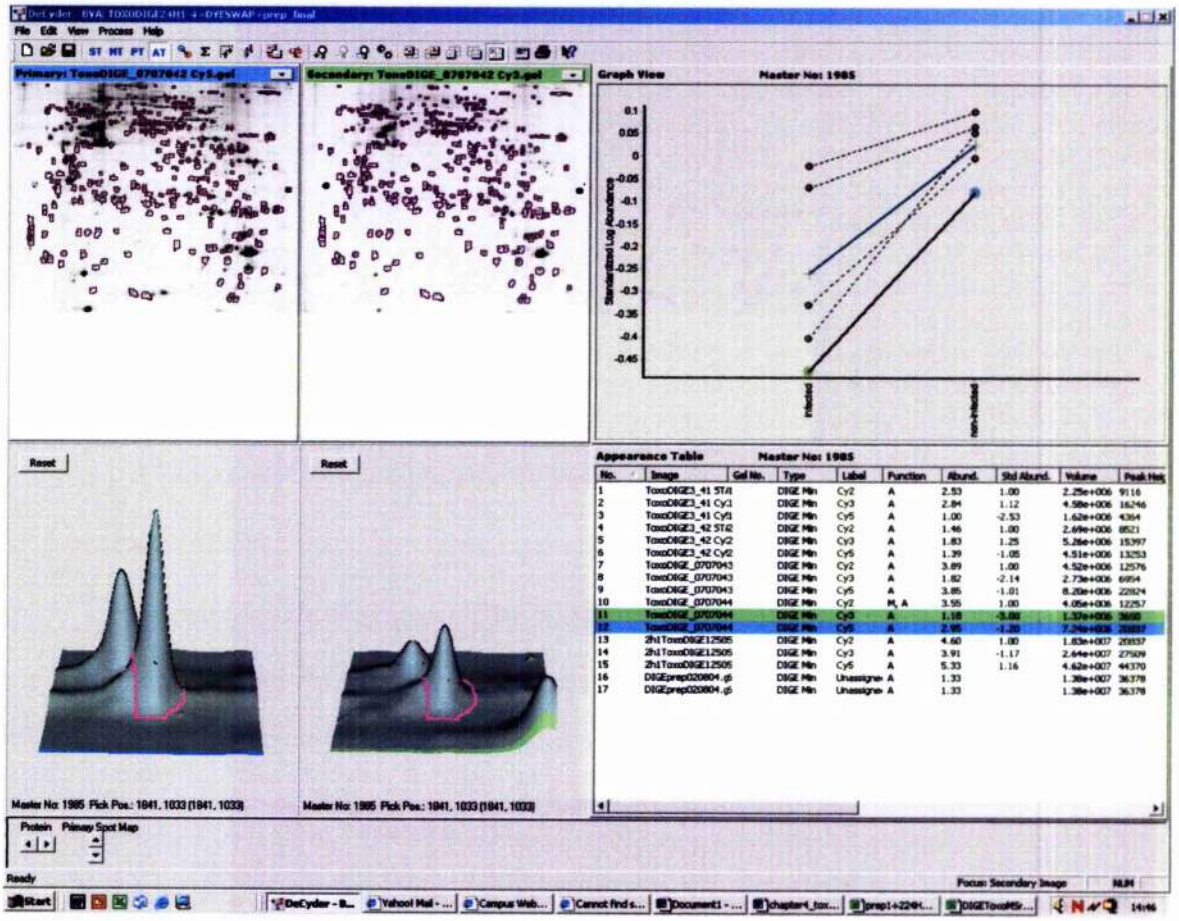


Figure 4-8 Screenshot of the BVA analysis software in the Appearance Table mode. The Appearance Table is used to display and process data associated with a single protein spot identified across the gels. Here we can see master spot number 1985 which is present on all 17 gel images (15 CyDye images and the two SYPRO stained preparative gel images) in the 24h DIGE experiment. The table view contains information about this one spot including the volume and abundance in all images.

The average ratio for each matched spot was calculated. A positive average ratio meant the spot volume had decreased in the infected cell and a negative average ratio meant the spot volume had increased in the infected cell.

After all the gels in the 24h infection experiment were matched and the average ratio of all matched spots calculated it was necessary to analyse the data using a Student's *t*-test to determine whether the volume difference between the infected and non-infected matched spot was significant. The protein statistics tool was used to calculate the average ratio and Student's *t*-test by comparing the two populations (infected and non-infected). A *t*-test value of ≤ 0.05 was deemed significant. Therefore if the *t*-test value was ≤ 0.05 it means that the protein abundance differences are statistically significant and there is only a 1 in 20 chance that the results are random and not due to induced biological variation. In the 24h DIGE experiment, 354 matched spots were present in 9 or more of the 15 gel images and had a *p*-value of 0.05 or less. Of these 354, 149 had a *p*-value of less than or equal to 0.01.

These 354 spots were all marked as a 'protein of interest'. In the infected cells, 141 spots had a decreased volume and 213 had an increase in volume compared to non-infected cells.

In the 12h DIGE experiment, 177 spots of interest were present in 9 or more of the 14 gel images with a p-value of ≤ 0.05. Of these 177 spots, 39 had a decreased volume in *T.gondii* infected cells and 138 had an increased volume.

4.3.4 Spot picking, Mass Spectrometry and Protein Identification

The nature of the minimal labelling method results in populations of labelled and unlabelled species for each protein in a lysate. For each protein spot on a 2-DE gel, the labelled species will be slightly shifted from the unlabelled species due to the addition of a single dye molecule. This effect is more marked for the lower than the higher molecular weight proteins. If a protein is picked using the centre of the spot detected from the CyDye DIGE Fluor fluorescent image (i.e. labelled protein) this may not correspond to the area of highest protein concentration. To circumvent this problem the total protein was visualised using a post-staining method (SYPRO orange staining) and the position of spots for picking based on this new image. This maximised the amount of protein available for MS. However it means that another matching analysis must be carried out; between the master gel and the preparative gel.

Therefore a preparative gel was run at the same time as the CyDye gels. The preparative gel contains a large amount of protein (400µg in this case) representing all the samples in the study. This gel was then used to pick spots for analysis by MS. The preparative gel also contained the Cy2 labelled pooled standard that was run on all gels in the study which allowed for accurate spot matching between the preparative gel and the CyDye comparative gels. All spots present on the preparative gel were detected using the DIA module. This preparative image was then loaded into the BVA module and the 'proteins of interest' (p-value <0.05) from the CyDye labelled master gel matched to the preparative gel. It was possible to match 217 of the 'proteins of interest' to the preparative gel. Of these, 92 had a decreased volume in infected cells and 125 had an increased volume in infected cells. A second preparative gel was run to allow for further matching of the proteins of interest in the BVA experiment which had not been matched with the first preparative gel.

Spots of interest were picked from preparative gels using the Amersham Ettan Spot Handling Workstation 2.1. Protein spots were digested using trypsin, mixed with CHCA matrix and plated onto MALDI targets. All spots were analysed by MALDI and any spots not identified by MALDI were than analysed by MS/MS.

The MS spectra were searched using the Mascot search engine. All MALDI identified proteins with MOWSE scores > 63 ($p>0.05$) and all MS/MS proteins with MOWSE scores >36 ($p>0.05$) were significant hits. Since both parasite and human proteins were present on the gel, all spectra were searched against the NCBI homosapiens database and a locally installed TgEST database. All spots which had significant hit for both a human and a parasite protein were compared. The *Toxoplasma* protein was blast searched against the NCBI database to determine whether it was the same protein as the human protein identified, or if the protein was unique to the parasite. If the human and parasite protein were the same then the result was discarded. This type of result was very rare.

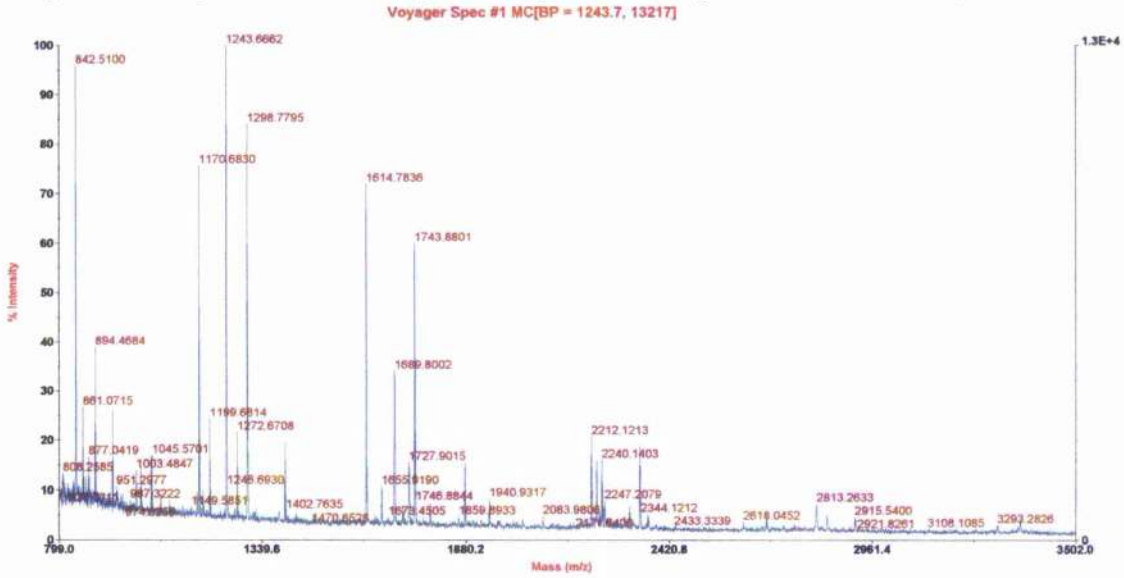
4.3.5 24h DIGE Results

In the 24h infection DIGE experiment protein identifications were gained from 163 spots using MS. Many of these spots contained mixtures of proteins. Therefore the actual number of proteins species identified from the 162 spots was 229. Often several isoforms of the same protein were detected on the same gel resulting in the identification of 96 individual proteins in the 24h DIGE experiment. The expression of the modulated proteins was up regulated, down regulated or modulated. Modulated means that the protein was identified as both up and down regulated and its expression in the infected cell is not clear cut. Of the 96 identified proteins, 39 were down regulated, 43 were up regulated and 14 were modulated due the presence of *T.gondii* in the host cell. Figure 4-9 shows a MALDI trace from peptides digested from a gel plug from a non-infected gel. These peak masses were searched against the NCBI and ToxoDB databases and the protein was identified as tropomyosin 4. An excerpt of the results table from the 24h DIGE experiment can be seen in Table 4-4. The full results table can be seen in Appendix III.

Many of the spots identified using MS/MS turned out to be mixtures of proteins. Figure 4-10 is the MASCOT results page of an MS/MS search. Nine proteins were identified from this one spot on the infected gel. Vimentin has a score of over 750 therefore it is likely that this protein is contributing to the volume change seen between the infected and non-infected cells. However, mixed spots may potentially cause problems in determining which proteins are up or down regulated in volume in the infected cell, as it is not always possible to determine which protein or proteins in the mixed spot are responsible for the change in volume. Ways to overcome this problem include identification of all the proteins present in the mixed spots and then carry out fractionation of the sample or fractionation of the gels (e.g. using narrow pH ranges) to allow for increased spot resolution. This will allow for accurate quantitation of individual protein spots. However this technique is not very practical and as yet, no-one has attempted this approach. The problem of mixed spots is very under looked in the literature.

Figure 4-11 shows information from the BVA module about one of the 24h differentially expressed proteins identified as peroxiredoxin by MS. The primary function of this protein as identified by Harvester and gene ontology is an anti-oxidant protein. It is also involved in cell proliferation. One reason this protein may be down regulated is that the host cell may stop dividing upon infection and it may be that the down regulation of this protein plays a role in the inhibition of cell division.

Figure 4-12 shows the 3D view of two spots of interest on infected and non-infected gels whose volume increased in the host cell due to the presence of the parasite. The 3D images on the left are from the infected cells labelled with Cy3. The 3D images on the right are of the spots from the non-infected cells labelled with Cy5. The histograms on the far right show how the volume of that spot behaved in all five of the 24h DIGE gels. These spots were identified as acetyl coenzyme a acyltransferase 2 and cathepsin B. Figure 4-13 shows two spots whose volume was consistently down regulated in the infected host cell in all five 24h DIGE. These spots were identified as heat shock protein 27 (HSP27) and an RNA-binding protein regulatory subunit. HSP27 is involved in stress response, actin organisation and regulation of translational initiation. HSP27 inhibits translation during heat shock (Cuesta *et al.*, 2000). Perhaps during infection the host cell tries to limit translation but the parasite blocks this so that translation continues uninterrupted for the parasite's gain. How the functions of the modulated proteins fit into the parasite infected cell will be discussed in detail in Chapter VIII.



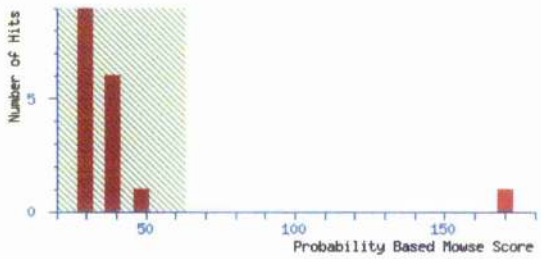
A MALDI spectra.

MATRIX SCIENCE Mascot Search Results

User	: **Morag**
Email	: **9606340n@student.gla.ac.uk**
Search title	: **morag050804DIGEToxo24h1_plate2spot40**
MS data file	: **C:\morag\DIGEToxo24h1\plate2\040804_DIGE-Toxo24h1_plate2__40_0001.dat**
Database	: **NCBInr 20040731 (1934002 sequences; 646347132 residues)**
Taxonomy	: **Homo sapiens (human) (113108 sequences)**
Timestamp	: **5 Aug 2004 at 10:50:29 GMT**
Top Score	: **170 for gi|4507651, tropomyosin 4 [Homo sapiens]**

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 63 are significant (p<0.05).



Concise Protein Summary Report

Format As | Concise Protein Summary | Help

Significance threshold p< 0.05 Max. number of hits 20

Re-Search All | Search Unmatched

1. gi|4507651 **Mass:** 28619 **Score:** 170 **Expect:** 1.1e-12 **Queries matched:** 11
tropomyosin 4 [Homo sapiens]

Figure 4-9 B MS results page searching the peak list from the MALDI spectra seen in A. This protein was identified as tropomysoin 4 and was down regulated after infection with *Toxoplasma gondii*. This protein identification has a score of 170, a score of > 63 is deemed significant.

Gel	Up/down	Spot no	T.g/H	M/D	Spectra	MS	Score	Id	Mass	pI
DIGEprep 290704	down	751	H	M	2 of 2	QStar	260	Caldesmon1 isoform 2	62683	6.18
DIGEprep 290704	down	740	H	M	2 of 2	QStar	253	Caldesmon1 isoform 2	62683	6.18
DIGEprep 290704	up	1893	H	M	2 of 5	Qatar	63	Glyceraldehyde-3-phosphate dehydrogenase	36202	8.26
DIGEprep 290704	up	1947	H	D	3 of 3	Qatar	78	Glyceraldehyde-3-phosphate dehydrogenase	36202	8.26
DIGEprep 290704	down	1985	H	D	1 of 1	MALDI	107	Peroxiredoxin 1, natural killer enhancing factor A	22324	8.27
DIGEprep 290704	down	1986	H	D	2 of 2	Qatar	184	Peroxiredoxin 1, natural killer enhancing factor A	22324	8.27
DIGEprep 290704	down	1984	H	M	1 of 1	MALDI	114	Peroxiredoxin 2 isoform b,	16036	6.12
DIGEprep 290704	down	1950	H	D	3 of 3	Qatar	70	Peroxiredoxin 3 isoform a precursor	28017	7.67
DIGEprep 290704	down	1897	H	D	1 of 1	MALDI	110	Peroxiredoxin 6, anti-oxidant protein 2	25133	6
DIGEprep 290704	down	1910	H	D	1 of 1	Qatar	90	Peroxiredoxin 6, anti-oxidant protein 2	25011	6

p.m	s.c	Appearance	T-test	Av. Ratio	gene name	gi number	Mascot	Harvester	Function	GO number	FUNCAT
9	21	14 (17) A, M	0.019	1.79	CALD1	4826657	http://www	http://harve	actin- and	http://godata	binding protein/mitosis
9	19	14 (17) A, M	0.02	1.79	CALD1	4826657	http://www	http://harve	actin- and	http://godata	binding protein/mitosis
3	7	17 (17) A, M	0.00015	-5.84	GAPD	31645	http://fun-g	http://harve	glycolysis	http://godata	glycolysis
5	18	17 (17) A, M	4.20E-06	-18.48	GAPD	31645	http://fun-g	http://harve	glycolysis	http://godata	glycolysis
9	45	17 (17) A, M	0.018	1.79	PRDX1	4505591	http://fun-g	http://harve	Anti-oxida	http://godata	anti-oxidant/cell proliferation
6	31	11 (17) A, M	0.027	1.77	PRDX1	4505591	http://fun-g	http://harve	Anti-oxida	http://godata	anti-oxidant/cell proliferation
8	44	17 (17) A, M	0.028	1.49	PRDX2	33188452	http://fun-g	http://harve	involved in	http://godata	anti-oxidant/cell proliferation
3	11	11 (17) A, M	0.0017	1.47	PRDX3	5802974	http://fun-g	http://harve	redox regu	http://godata	anti-oxidant/cell proliferation
6	33	17 (17) A, M	0.025	1.59	PRDX6	4758638	http://fun-g	http://harve	involved in	http://godata	anti-oxidant/cell proliferation
3	12	14 (17) A, M	0.022	1.98	PRDX6	3318842	http://fun-g	http://harve	involved in	http://godata	anti-oxidant/cell proliferation

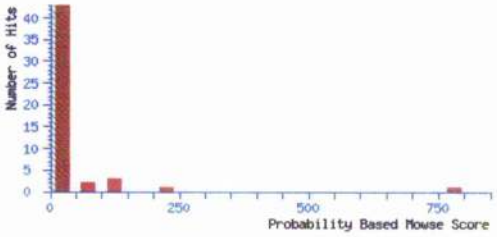
Table 4-4 Exert of 24h DIGE results table. This table contains information about the proteins identified from the 24h DIGE experiment; the experiment name, whether the protein is up or down regulated, the BVA spot number, organism protein is from (Human or *T.gondii*, manual or daemon MS identification, how many proteins were identified per spot, type of MS used, MASCOT score, protein name, mass, pI, peptides matched, sequence coverage, how many gels the spot was seen in, *p*-value from the *t*-test, average ratio, gene name, gI number, MASCOT URL, Harvester URL, function, gene ontology URL, functional category.

Mascot Search Results

User	:	Morag
Email	:	9606348m@student.gla.ac.uk
Search title	:	Morag24hToxnDIGEmchiplate1492spot1296
MS data file	:	C:\morag\MoragDiGEplate1492.wiff
Database	:	NCBInr 20040911 (2888429 sequences; 674368817 residues)
Taxonomy	:	Homo sapiens (human) (108917 sequences)
Timestamp	:	13 Sep 2004 at 11:43:45 GMT
Significant hits:		gi|340219 vimentin
		gi|30582431 CDC37 cell division cycle 37 homolog (S. cerevisiae) [Homo sapiens]
		gi|1710248 protein disulfide isomerase-related protein 5 [Homo sapiens]
		gi|71541 desmin - human
		gi|5453629 dynactin 2; dynactin complex 50 kD subunit; 50 kD dynein-associated polypeptide [Homo sapiens]
		gi|1346343 Keratin, type II cytoskeletal 1 (Cytokeratin 1) (K1) (CK 1) (67 kDa cytokeratin) (Hair alpha protei
		gi|12654715 TXNDC5 protein [Homo sapiens]
		gi|7717364 homolog to cAMP response element binding and beta transducin family proteins [Homo sapiens]
		gi|4826870 nucleobindin 2 [Homo sapiens]
		gi|284165

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 36 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As: Peptide Summary — Help

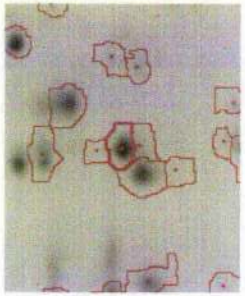
Significance threshold p< 0.05 — Max. number of hits 20

Standard scoring / MudPIT scoring — Ions score cut-off 0 — Show sub-sets

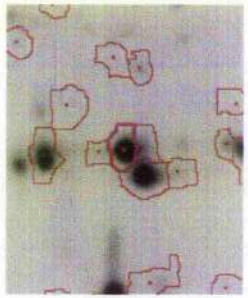
Show pop-ups / Suppress pop-ups — Sort unassigned: Decreasing Score — Require bold red

Select All | Select None | Search Selected | Error tolerant | Archive Report

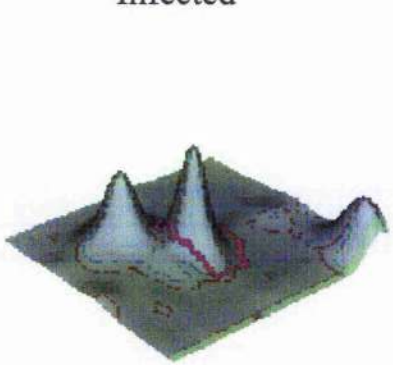
Figure 4-10 MASCOT results page of a mixed spot. Nine different proteins were identified using tandem MS from one spot on the DIGE 2-DE gel.

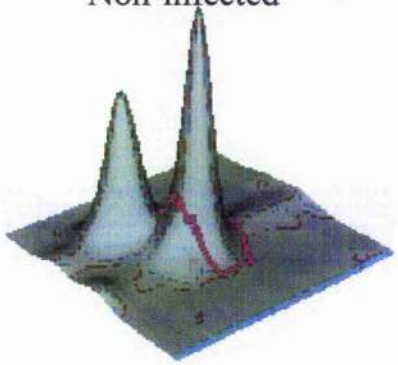


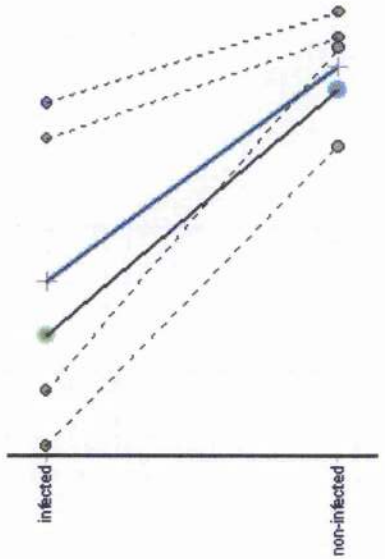
Infected



Non-infected



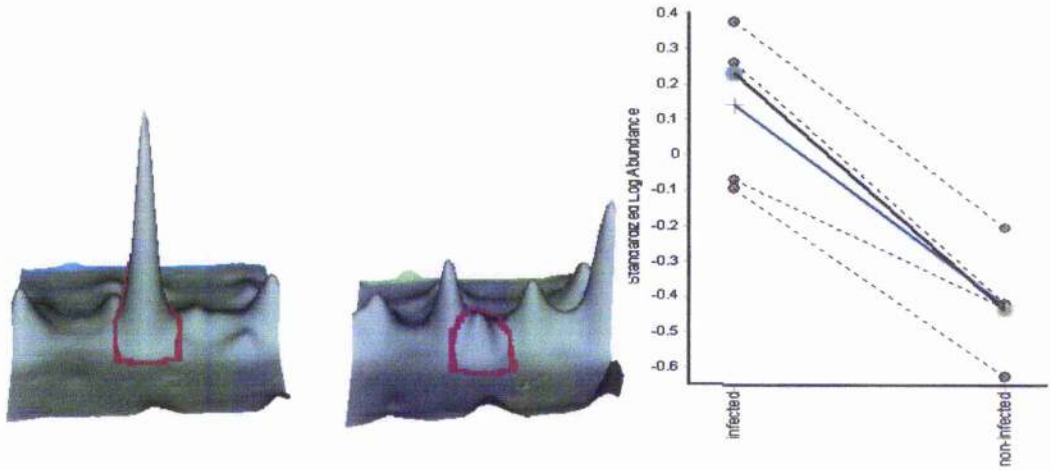




Protein Table	T-test and Av.Ratio: non-infected / infected		
Master No.	Appearance	T-test	Av. Ratio
1985	17 (17) A, M	0.018	1.79

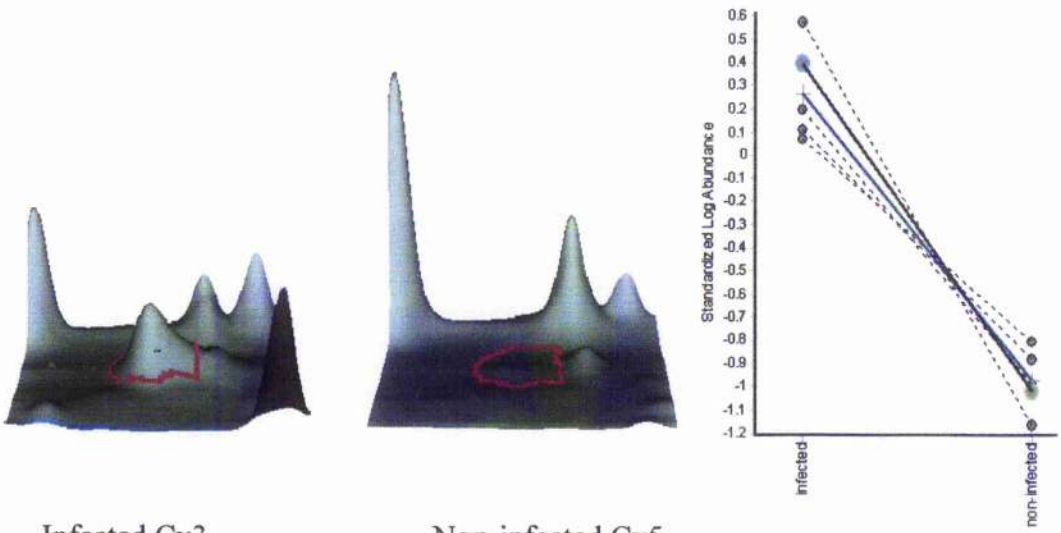
Master No	MS	Score	Protein	Mass	pI	P.M	Seq.cov%	Gene nam	gI number	Mascot	Harveseter	Function	GO no.	Functional categories	
1985	MALDI	107	Peroxiredo	22324	8.27	9	45	PRDX1	4505591	http://fun-g	http://harve	Anti-oxidar	http://goda	11	anti-oxidant
														3	cell proliferation

Figure 4-11 Information collated from BVA about the differentially expressed protein peroxiredoxin. The two gel images show the matched spot from one set of gels and below this are the 3D images of the spots. The spot volume is clearly down regulated in infected cells. The graph in the upper right corner shows how the volume of the spot behaves in all 5 gel sets. As can be seen the volume behaves similarly in all experiments, down regulated in infected cells. The table under the graph shows the information from the BVA module, with the spot number, how many gels the spot was seen in (every gel in the experiment in this case) and the average volume change between infected and non-infected cells. The table that runs along the bottom shows information about the identified spot; the spot number, the type of MS used, the protein name, mass, pI, peptides matched and sequence coverage, the gene name and gene number, the MASCOT and Harvester URLs, the protein function and the gene ontology URL.



Infected Cy3 Non-infected Cy5

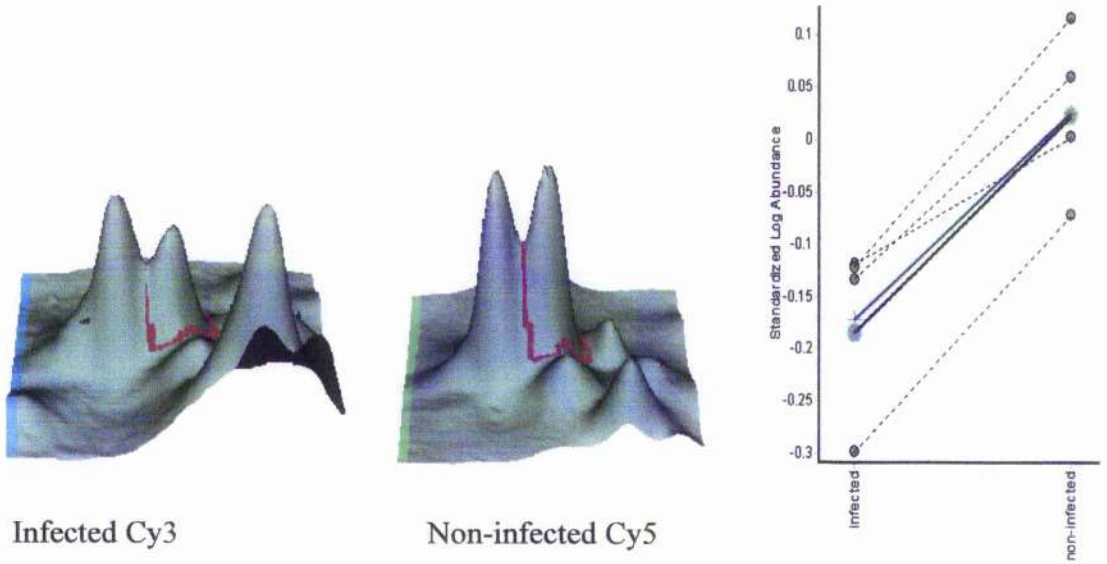
Spot number 1540. Acetyl coenzyme a acyltransferase 2. Up regulated in infected cells. -3.83 average ratio. *p*-value = 0.0012



Infected Cy3 Non-infected Cy5

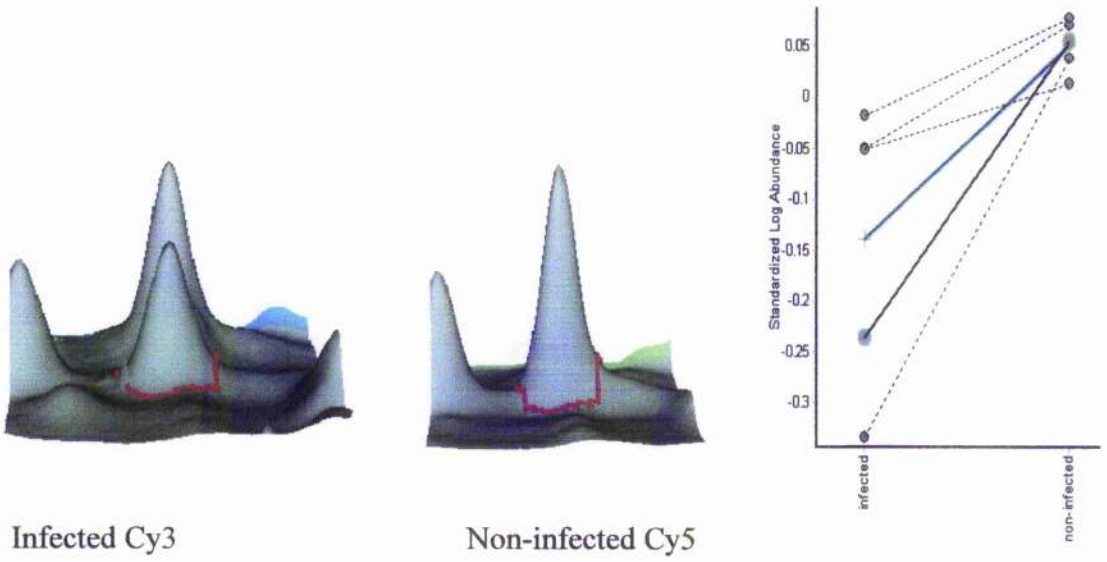
Spot number 1947. Cathepsin B. Up regulated in infected cells. -18.48 average ratio *p*-value = 0.0000042

Figure 4-12 3D view of two spots of interest on infected and non-infected gels whose volume increased in the host cell due to the presence of the parasite. The 3D images on the left are from the infected cells labelled with Cy3. The 3D images on the right are of the spots from the non-infected cells labelled with Cy5. The histograms on the far right show how the volume of that spot behaved in all 5 of the 24h DIGE gels. These spots were identified as acetyl coenzyme a acyltransferase 2 and cathepsin B.



Infected Cy3 Non-infected Cy5

Spot number 1902. Heat shock protein 27. Down regulated in infected cells
1.58 average ratio. *p*-value = 0.0025



Infected Cy3 Non-infected Cy5

Spot number 1964. RNA-binding protein regulatory subunit
Down regulated in infected cells 1.49 average ratio. *p*-value = 0.018

Figure 4-13 3D view of two spots of interest on infected and non-infected gels whose volume decreased in the host cell due to the presence of the parasite. These spots were identified as heat shock protein 27 (HSP27) and an RNA-binding protein regulatory subunit.

4.3.6 12h DIGE Results

In the 12h DIGE experiment, 177 spots were present in 9 or more of the 14 gel images with a p-value of ≤0.05. Of these 177 spots, 39 had a decreased volume in *T.gondii* infected cells and 138 had an increased volume. Of the 177 spots of interest it was possible to match 79 spots to the preparative gel. Using a combination of MALDI and MS/MS it was possible to identify 31 spots from the 12h infection DIGE experiment. This constituted 9 human proteins and 10 *Toxoplasma* proteins. Of the 9 identified human proteins, 5 were up regulated and 4 down regulated in the parasite infected cell. An excerpt of the results table from the 12h DIGE experiment can be seen in Table 4-5. The full results table can be seen in Appendix IV.

A comparison of the differentially expressed spots across time can be seen in Table 4-6. Of the five proteins which were modulated at both 12h and 24h, four were consistently up regulated. The only protein not to behave in the same manner between 12h and 24hpi was actin beta, the expression of this protein was mixed therefore correlation over time was difficult to determine.

To compare time points post infection with technique Table 4-6 was compared with Table 3-7 from Chapter III which compared the differentially expressed proteins across time as identified by conventional gel-to-gel electrophoresis. Actin beta, Aldolase A and BiP protein were all consistent in terms of expression modulation in the infected cell. Vimentin was shown as down regulated/modulated in the gel-to-gel experiments and in the DIGE experiments vimentin was identified as up regulated in the infected cell. As the DIGE experiments involve at least four gels per experiment then these results are likely to be more significant than the mixed vimentin response seen in Chapter III.

Up/Down	Spot no.	Appearance	T-test	Average Ratio	Gel id	T.g or H	M/D	MS	Score	Protein name	Mass	pI	p.m	s.c
DOWN	1646	14 (14)	0.023	-1.37	12hDIGE	H	M	MALDI	449	ACTB	40536	5.55	11	27
DOWN	1647	14 (14)	0.0018	-1.33	12hDIGE	H	M	MALDI	247	Actin gamma	41977		8	
UP	774	14 (14)	0.012	1.52	12hDIGE	H	M	Qatar	183	BiP	72185	5.03	7	16
DOWN	1116	14 (14)	0.035	-1.62	12hDIGE	H	M	MALDI	79	Lamin A/C isoform2	65153	6.4	7	14
UP	2043	11 (14)	0.028	5.4	12hDIGE	H	M	MALDI	66	PIAS-Nyprotein	45671	9.52	4	9
UP	989	14 (14)	9.90E-05	3	12hDIGE	H	M	MALDI	74	Protein kinase NYD-Sp9	34362	4.9	5	17
DOWN	246	14 (14)	0.021	-1.19	12hDIGE	H	M	Qatar	107	Protein serine/threonine kinase	43424	6.28	4	23
UP	1412	14 (14)	0.00075	2.41	12hDIGE	H	M	MALDI	160	Vimentin	53738	5.06	15	7
UP	1632	14 (14)	0.0064	1.57	12hDIGE	H	M	MALDI	155	Aldolase A	39706	8.34	5	17

Table 4-5 Excerpt of 12h DIGE results table. This table contains information about the proteins identified from the 12h DIGE experiment; the experiment name, whether the protein is up or down regulated, the BVA spot number, organism protein is from (Human or *T.gondii*, manual or daemon MS identification, how many proteins were identified per spot, type of MS used, MASCOT score, protein name, mass, pI, peptides matched, sequence coverage, how many gels the spot was seen in, *p*-value from the *t*-test, average ratio, gene name, gI number, MASCOT URL, Harvester URL, function, gene ontology URL, functional category.

Protein name	Gene name	24h DIGE	12h DIGE	Correlates?
Lamin A/C, isoform 2	**LMNA**	DOWN	DOWN	YES
Aldolase A	**ALDOA**	UP	UP	YES
BiP protein	**HSPA5**	UP	UP	YES
Vimentin	**VIM**	UP	UP	YES
Actin	**ACTB**	MODULATED	DOWN	MAYBE?

Table 4-6 Comparison of modulated proteins at 12hpi and 24hpi. Of the six proteins which were differentially expressed in both the 12h and 24h infection experiments, five of them were up regulated across time. The only protein which did not have a definite correlation between 12h and 24h was actin beta.

4.4 Discussion

Previous work (Chapter III) showed that many changes were stimulated in the proteome of *Toxoplasma gondii* infected cells. There are several limitations in the gel-to-gel technique used in Chapter III, including gel and IPG strip variation, user dependant analysis and poor analysis software, often making gel-to-gel comparisons inaccurate. The high degree of gel-to-gel variation in spot patterns often results in difficulties in distinguishing biological from experimental variation (Karp *et al.*, 2004). Thus, it is often necessary to run numerous 2-DE gels to analyse one sample. To overcome these issues, Unlu *et al* developed an approach involving the multiplexing of samples, called difference gel electrophoresis (DIGE) which has since been commercialised by Amersham Biosciences.

DIGE provides several advantages over the gel-to-gel technique. The combination of CyDye DIGE Fluor minimal dyes and DeCyder Differential Analysis Software exploits the multiplexing capability of the 2-DE DIGE methodology. Inclusion of a pooled internal standard eliminates system variation, allowing highly accurate measurement of protein abundance changes. The use of biological replicates in the experimental design ensures a true measurement of induced biological differences above the background of inherent biological variation (Lilley and Friedman, 2004). The DIGE system is capable of routinely detecting and quantifying differences as small as 10% between samples with greater than 95% statistical confidence. DIGE leads to highly accurate qualitative and quantitative results because gel-to-gel variations are eliminated. Inclusion of an internal standard on every gel allows easy matching of gels. This helps to avoid the complications of inter-gel comparison but also speeds up the analyses and reduces the number of gels that need to be run for a reliable comparison of protein patterns (Czupalla *et al.*, 2005).

Minimal labelling causes slight differences in molecular weight between labelled and non-labelled protein. This molecular weight shift does not pose a problem for DIGE where both sets of proteins are labelled to the same extent. The effect of the dyes on pI is more difficult to estimate because the ionisable positive charges of the primary amino groups are replaced by the non-ionisable quaternary amino group of the dye. However, at the range of pH in an IEF gel, this change should not have a noticeable effect (Unlu *et al*, 1997). Disadvantages of the DIGE technology include mixed spots where multiple proteins may be present in one spot making if difficult to determine which protein is differentially expressed. This is also a problem with conventional gel-to-gel electrophoresis and was discussed in Chapter III.

The goal of the work in this chapter was to employ DIGE technology to investigate the protein expression of infected host cells and to compare this technique with the more conventional gel-to-gel approach. The aim was to identify protein spots in the 2-DE DIGE gels that change in volume in a way that correlates with infection status or progression.

Toxoplasma gondii infected and non-infected HFF cell extracts at 12hpi and 24hpi were analysed using DIGE. Infected, non-infected and pooled standard CyDye labelled extracts were mixed together and separated by 2-DE on the one gel. Analysis of the gel images using the DIGE DeCyder software identified several hundred spots whose volume changed significantly and reproducibly in the pathogen infected cell. A statistically significant change in spot volume was defined as a protein spot which had a Student's *t*-test value of $p<0.05$. The Student's *t*-test null hypothesis was that there was no change in the protein abundance between infected and non-infected cells. The spots of interest were excised from the gels and identified by MS.

In the 24h infection experiment 96 differentially expressed human proteins were identified with statistical variance of the infected versus non-infected spot volume ratio within the 95th confidence interval (Student's *t*-test; $p<0.05$). An ontology analysis of these proteins revealed that there is a great deal of expression modulation in the infected host cell metabolism (cathepsins and heterogeneous ribonucleoproteins), glycolysis (aldolase, enolase, phosphoglycerate kinase, pyruvate kinase and glyceraldehyde phosphate dehydrogenase) and structural proteins (actin beta, laminins, tropomyosins and vimentin) among others. The significance of the functions of the modulated proteins in the *Toxoplasma gondii* infected cell will be discussed in detail in Chapter VIII.

The modulated proteins identified in the 24h DIGE experiment were compared with those modulated in the 24h gel-to-gel experiments discussed in Chapter III. Combining the results from both sets of experiments showed 147 individual proteins changed in expression in the host cell 24hpi. Thirty four of these proteins were common to both sets of experiments. Of the 34 proteins whose expression was modulated in both sets of experiments, 17% did not have the same expression response, e.g. the protein was down regulated in gel-to-gel and up regulated in the DIGE experiment. Therefore, considering the proteins common to both techniques, the DIGE data backed up the gel-to-gel data. Over two-thirds of the proteins whose expression changed 24hpi were unique to either the gel-to-gel 2-DE or the DIGE. As DIGE provides statistically significant results and is a great deal more sensitive at detecting minor changes than conventional 2-DE, these results highlight the advantages of DIGE over conventional 2-DE, but also shows that the

techniques can be complementary. DIGE is a lot less time consuming than the manual gel-to-gel approach and also profiles more proteins. The fluorescent staining used in DIGE is also more linear and sensitive than the colloidal Coomassie staining used for the gel-to gel technique. The DeCyder software allows for univariate statistical testing (e.g. Student's t-test and ANOVA). It has recently been proposed that univariate tests may increase the likelihood of false positive results (Karp *et al.*, 2004). Multivariate tests (e.g. partial least squares discriminant analysis) can be used to complement the univariate approach in identification of differentially expressed protein spots. Multivariate tests are useful for recognising pathway changes and other correlated events (Karp *et al.*, 2004). Further statistical analysis of the data presented here may have provided complementary results concerning the host response to infection.

Combining the results from all of the time course infection experiments using both the gel-to-gel and DIGE techniques showed modulation of 157 proteins in *Toxoplasma gondii* infected cells. These results provide a dataset which is comparable to the gene expression data from the study by Blader *et al.* (Blader *et al.*, 2001). This will allow for a direct gene to protein comparison.

Chapter V Investigation into the gene to protein correlation in HFF cells during infection with *Toxoplasma gondii*

5.1 Introduction

This chapter aims to investigate the gene to protein correlation in *Toxoplasma gondii* infected cells. Host cell gene expression reflects RNA transcription and degradation rates, but the extent to which this correlates with the corresponding protein levels which are influenced by translational and post-translational mechanisms is not known. It has been suggested that there is little correlation between mRNA and protein abundance in some systems (Le Roch *et al*, 2004).

Several comparative studies have been carried out which showed a wide range in the correlation between gene and protein expression (Le Roch *et al*., 2004; Gygi *et al*., 1999; Cox *et al*., 2005). Differing levels of correlation can be explained by post-translational processing or splicing events which can influence protein expression. There are also difficulties in comparing gene and protein data as no single proteomic technique can provide truly complete proteome data whereas microarrays can more easily profile the entire genome transcriptional response. One of the comparative studies was a genome-scale epitope tagging study of protein abundance in yeast (Cox *et al*., 2005). The authors concluded that many essential proteins and transcription factors are present at levels that are not readily predicted by mRNA levels. Gygi *et al* (1999) tested gene to protein correlation in yeast using ^{35}S methionine labelling for protein quantitation and serial analysis of gene expression (SAGE) analysis for mRNA transcript quantitation and showed a modest positive correlation between mRNA and protein abundance. A third group used ICAT labelling and the ratio of mRNA using spotted cDNA arrays to compare mRNA and protein abundance of the parasite *Plasmodium falciparum* (Le Roch *et al*., 2004). This study concluded that there was a moderately high positive relationship between mRNA and protein abundance. Therefore the gene to protein correlation remains unclear.

To investigate gene to protein correlation in the context of the dynamic event of host cell invasion by a pathogen, and therefore the role of post-transcriptional controls in the regulation of protein expression during infection with *Toxoplasma gondii*, we compared mRNA transcript data (Blader *et al*, 2001) with protein abundance changes measured in

this study (Chapters III and IV). A moderately low positive relationship between mRNA and protein abundance was observed.

Several limitations exist in comparing genomic and proteomic datasets. Microarrays can only detect transcripts that have a representative probe on the array. To calculate the degree of correlation, corresponding genes and proteins must be matched. The sequence identifiers for the individual gene or protein may be from the same database (e.g. SwissProt, NCBI, ENSEMBL) but more often it is necessary to cross-reference between different databases and link gene and protein identifiers manually. Often there will be gaps in the data sets where there is no gene expression data available for a protein or vice versa, e.g. the gene was not present on the array. The most common problem arises from the fact that proteomic techniques do not profile the entire proteome. The techniques most commonly used may fail to detect secreted or membrane proteins that are present in low abundance (Le Roch *et al*, 2004). Approximately 3000 spots can be resolved on a 2-DE gel (Jenkins and Pennington, 2001) and microarrays can contain tens of thousands of genes. Comparison of the overlapping genes and proteins excluding all data points that do not cross-map can be performed. Alternatively entire data sets can be compared placing blank values for missed matches. As microarray studies are typically of far greater scope at present than most proteomic experiments, this second comparison may not be helpful due to the large amount of unmatched data (Cox *et al.*, 2005). An advantage that proteomics has over gene expression studies is the ability to profile post-translational modifications (PTM). A single gene may be represented by several spots on a gel (different protein isoforms) and these are not detectable by microarray studies. Thus to some extent microarrays and 2-DE represent different types of data.

Although comparing protein lists can be informative some estimate of relative quantity across samples is generally required for meaningful comparisons with microarray data. Protein levels can be measured using quantitative proteomics e.g. isotope coded affinity tags (ICAT) or as in this study by difference gel electrophoresis (DIGE).

Quantitative changes in protein expression in *Toxoplasma gondii* infected HFFs were determined using DIGE (Chapter IV). These proteins were identified using MASCOT software to search the MS data against the complete set of human proteins in the NCBI database. A previous study was carried out under identical experimental conditions in which the expression of approximately 22,000 genes were assessed by microarray assay (Blader *et al.*, 2001) thus presenting an ideal opportunity to perform a comparison with proteins modulated in expression under the same conditions.

To confirm further some of the changes seen in gene and protein expression in the pathogen infected cells Western blotting (WB) was used. This technique allows for a further comparison of the gene and protein response (Mukherjee *et al.*, 2003; Zhu *et al.*, 1998; Alfonso *et al.*, 2005; Jiang *et al.*, 2005). The proteins chosen for WB were selected based on the following criteria; their expression was modulated in the microarray and proteomic studies, the function of the protein was of biological interest in the pathogen infected cell and antibody specific for that protein was available. Another consideration was whether the protein could be inhibited or over expressed in the laboratory. For example, glyceraldehye-3-phosphate dehydrogenase (GAPDH) is up regulated in gene and protein expression in *Toxoplasma gondii* infected cells. Inhibitors of this protein are available which could potentially be added to the infected cells to investigate whether inhibition of this protein resulted in a decrease in invasion or infection.

Three proteins of interest were chosen for further study. These were caldesmon, cathepsin B and GAPDH. Table 5-1 shows the changes in expression in HFFs of these genes and proteins after infection with *Toxoplasma gondii* as determined by the MA study of Blader and 2-DE analysis (this study). Caldesmon 1 appears not to have been present on the MA, this will be discussed in detail later in the chapter.

Protein name	Gene name	Microarray	24h gel to gel	24h DIGE
Caldesmon	CALD1	no match to array	down	down
Cathepsin B	CTSB	down	down	down
Glyceraldehyde 3 phosphate dehydrogenase	GAPDH	up	up	up

Table 5-1 Proteins chosen for further confirmation of expression by Western Blotting. Table shows the changes in expression in the gene (MA column) and protein (24h gel-to-gel and 24h DIGE).

5.2 Materials and Methods

5.2.1 Comparison of microarray and proteomic data

In the proteome investigation, a set of proteins were identified that were changed in expression between non-infected and *Toxoplasma gondii* infected host cells (Chapters III and IV). The proteins were identified using the MASCOT software to search the MS data against the complete set of human proteins (NCBI). A comparison of proteins modulated in volume in this study to the genes that were found to be differentially expressed in the study by Blader *et al* (2001) was performed. One of the difficulties with this type of comparison is the incompatibility of accession numbers returned by MASCOT for human proteins with the accession numbers identifying clones or mRNA sequences on the microarrays. It is also not always possible to detect that a protein record corresponds with a particular clone record by matching the gene or protein name because there are many examples in the databases in which inconsistent names have been assigned to genes and proteins. Therefore the following processes were used to ascertain which proteins identified in this study matched clones on the microarray in the Blader study.

5.2.1.1 Retrieving compatible identifiers for MA and proteomic data

The first stage involved obtaining the GenBank accession numbers for the corresponding nucleotide sequence for each of the proteins identified in this study. Many of the GenBank protein records have a field called "DB_SOURCE" that specifies a database identifier for the source of the protein sequence. The source of a protein sequence can be a GenBank nucleotide record, a SwissProt record or a protein 3-D structure from the Protein Database of 3-D Structures of Biological Macromolecules (PDB). The BioJava (http://www.biojava.org/) API (application programming interface) was used to retrieve the value in the "DB_SOURCE" field for each protein record. Approximately 60% of the protein records returned an identifier for a GenBank nucleotide record.

Using the harvester search engine, each protein was assigned its HUGO (Human Genome Organisation) gene name. This was done manually by searching the EMBL (European Molecular Biology Laboratory) Harvester database (www.embl.harvester.de). The HUGO website (http://www.gene.ucl.ac.uk/nomenclature/) provides a table that maps the gene name to a GenBank nucleotide record. The nucleotide records were retrieved for the proteins and approximately 65 % of accession numbers were obtained by this method. Not

all accession numbers could be retrieved by this method because the HUGO gene list is not exhaustive, i.e. there are gene symbols that have not been mapped to an accession number.

Once the nucleotide accession numbers had been obtained for as many protein records as possible, a process was performed to find the microarray clones that corresponded to the same gene. The DoTS (Database of Transcribed Sequences) resource (www.allgenes.org) contains clusters of accession numbers that correspond to the same underlying gene. The database has been created by performing sequence alignment over several different human and mouse databases to create so called "DoTS Genes". Each DoTS gene has been mapped to mRNA, cDNA and EST sequences residing in public sequence databases. A copy of DoTS was generated locally and for each protein identified in the study, DoTS was queried using the nucleotide accession number, to retrieve the corresponding DoTS gene. For each of the clones on the microarray, the accession numbers were retrieved and also queried against DoTS. If a protein and a clone matched the same DoTS gene, it could be concluded that they correspond to the same gene. This process is summarised in Figure 5-1.

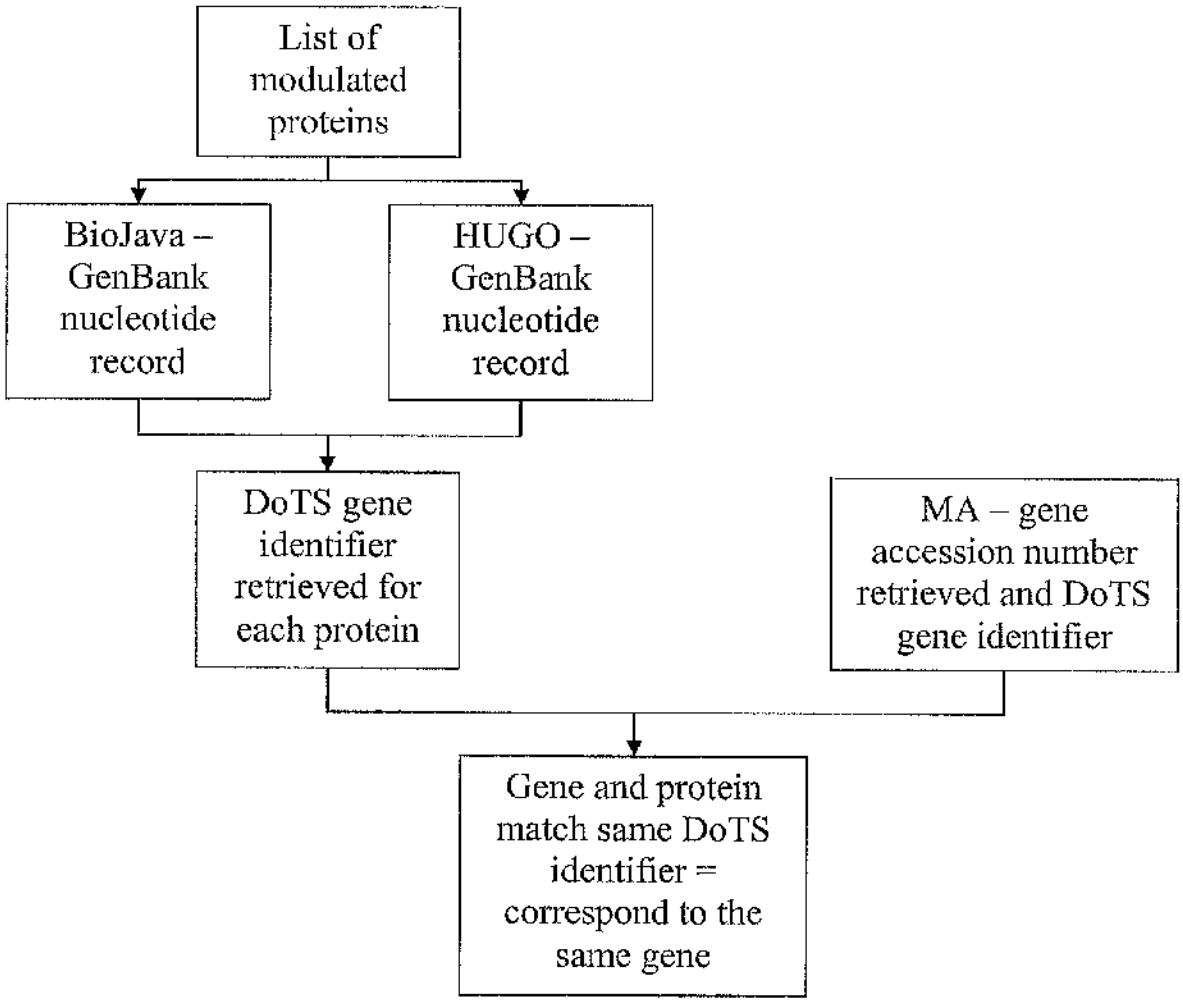


Figure 5-1 Schematic of the process used to match MA clones and differentially expressed proteins.

5.2.2 DeCyder

The DeCyder BVA module calculates the average ratio for each matched spot. The program converts the calculated normalised log volume ratio (R, equation 1) to an expression ratio E, by equations 2 and 3, to give a value greater than one or less than minus one that is reported to the user as the measured ratio change.

Equation 1: $R = \log10(V2/V1)$

V1 is the normalised volume of the spot in the primary image and V2 is the volume of the spot in the secondary image

Equation 2: $E = (V2/V1)$ for $R>0$, i.e. $V2/V1>1$

Equation 3: $E = -(V1/V2)$ for $R<0$, i.e. $V1/V2 >1$, (Karp *et al.*, 2004)

To get the DIGE data into the same format as the MA data; formula 1 was used.

Formula 1: =IF(V1>0,V1,1/V1) where V1 is the DIGE value.

5.2.3 Statistical Analysis

To examine if the observed changes in protein abundance correlated with changes in mRNA abundance as reported by Blader *et al* (2001) several statistical tests were carried out. The Spearman rank correlation has been commonly used for comparisons between mRNA and protein data sets, as it provides a statistical method to compare the two data sets while taking into account the different methods by which mRNA and protein abundance are measured (Le Roch *et al.*, 2004). To calculate the Spearman rank correlation between transcriptome and proteome data sets, each data set was sorted by mRNA or protein volume change, and each value in each data set assigned a rank. The Spearman rank correlation coefficients were calculated using the following equation:

$$Sr = 1 - (6\sum D^2/N(N^2-1))$$

where D is the difference between the ranks of the corresponding mRNA and protein expression values for a particular gene and N is the number of mRNA and protein value pairs (Le Roch *et al.*, 2004).

The 23 genes and proteins which matched between the MA and DIGE data sets were assigned a rank of 1-23 based on the MA value or DIGE adjusted value (using formula 1). The ranks were then entered into the statistical package offered by http://faculty.vassar.edu/lowry/webtext.html and the r-value calculated.

A Mann Whitney U test was also carried out to test the null hypothesis that changes in mRNA and protein abundance were similar and therefore correlated with each other. The combined MA and proteomics abundance data were ranked in order (1-46, based on the MA value and DIGE adjusted value using formula 1) and the Mann Whitney U statistical package from http://faculty.vassar.edu/lowry/webtext.html used to calculate the U-value.

5.2.4 Western blotting

HFFs were infected or mock-infected with *Toxoplasma gondii* tachyzoites and left for 24h. The cells were harvested as previously described in Chapter II. All 1-DE and 2-DE sample preparation was as previously described in Chapter II.

Equal amounts (determined using the Bradford protein assay) of infected HFF protein extract, uninfected HFF protein extract and *T.gondii* protein extract were loaded onto 1-DE gels. The proteins in the 1-DE western blots were separated on a 12% resolving gel at 120V in 1x SDS electrophoresis buffer until the dye front had run to the bottom of the gel. The standard used in the 1-DE western blots was the SeeBlue® Plus2 Pre-Stained standard (Invitrogen) which consists of 10 pre-stained protein markers. For the 2-DE western blots, proteins were separated in the 1st dimension on pH4-7, 7cm immobiline dry strips (following the IEF protocol described in Chapter II), equilibrated and then separated in the second dimension on a 12% resolving gel at 120V in 1X SDS electrophoresis buffer until the dye front had reached the bottom of the gel.

Proteins were transferred from the gel to Hybond membrane (Amersham Biosciences) using a wet transfer system. The transfer was carried out at 100V for 45min in 10% transfer buffer (10x transfer buffer: 24.22g Tris, 112.5g glycine in 1L ddH_2O), 20% methanol and 70% ddH_2O. The membrane was then stained with Ponceau S solution (Ponceau S concentrate-2% (Sigma, Cat P7767) 1:100 with ddH_2O) to check transfer efficiency and to confirm that equal amounts of starting material had been loaded onto the 1-DE gel. The blot was blocked in 5% milk solution in 1x TBS (10x TBS: 20mM Tris, 137mM NaCl in 1L ddH_2O adjusted pH to 7.6 using HCl) + 0.1% TWEEN-20 for 2h at 37

°C. Antibody was applied in the appropriate dilution in 1% milk solution in 1x TBS + 0.1% TWEEN-20 and left overnight, shaking, at 4°C.

Monoclonal antibodies were chosen for this study as they are more specific than polyclonal (there are several isoforms of caldesmon and cathepsin present on the 2-DE gels therefore the antibody needed to be specific to cathepsin B and caldesmon 1) and generally give a lower background.

Mouse monoclonal anti-caldesmon (SIGMA, mouse IgG isotype, specific to chicken, turkey, rabbit, guinea pig and human caldesmon) was used at a dilution of 1:400 and 1:800 (as recommended by the antibody data sheet) for Western blotting. Mouse monoclonal anti-cathepsin B (CALBIOCHEM, mouse IgG isotype, specific to human cathepsin B) was used at a dilution of 250ng/ml and 500ng/ml (as recommended by the antibody data sheet) for Western blotting. Mouse monoclonal anti-GAPDH (CALBIOCHEM, mouse IgG isotype, specific to canine, chicken, fish, frog, human, mouse, porcine, rabbit and rat GAPDH) was used at a concentration of 2μg/ml (as recommended by the antibody data sheet) for Western blotting.

Blots were then washed four times in 1% milk solution in 1x TBS for 15min each at room temperature. The secondary antibody was rabbit anti-mouse IgG, (Pierce, catalogue # 31450) horseradish peroxidise conjugated, used at a concentration of 1 in 10,000 in 1% milk solution in 10 x TBS for 2h at room temperature. The blot was then washed three times in 1% milk solution in 1x TBS for 15min each at room temperature.

The blot was briefly rinsed in 10x TBS, and the bound antibody visualised using Supersignal West Pico Chemiluminescent Substrate from Pierce (according to the manufacturer's instructions) and developed on film (Hyperfilm ECL from Amersham Biosciences). Films were developed and fixed using Kodak GBX developer/replenisher and fixer/replenisher (Sigma) respectively. Western blotting was repeated several times for each antibody to confirm the patterns observed.

5.3 Results

5.3.1 Microarray and proteomics comparison

In the MA study of Blader *et al.* (2001), approximately 22,000 genes were assayed and approximately 1200 genes (about 5%) were modulated in expression, 600 of these significantly (>2 fold). Data from the 24h gel-to-gel and 24h DIGE experiments were compared to the 24h MA time point. One hundred and forty seven proteins were identified as being modulated in expression from the 24h infection proteomic studies (Chapters III and IV). Of these, 96 proteins matched to a gene present on the MA. The matching procedure (which compared all the proteins modulated in the proteomic experiments with the MA data) highlighted 48 proteins that corresponded with microarray genes that also changed in expression. Each of these matches was manually verified by checking that the protein name and clone name were similar, to ensure that an incorrect match had not been made. 48 proteins were matched to clones that were assayed on the array but did not show a significant change in gene expression. The rest of the differentially expressed proteins (51) in this study were not matched to any microarray clone.

The distribution of proteins and genes identified as modulated between the different techniques is represented as a Venn diagram in Figure 5-2.

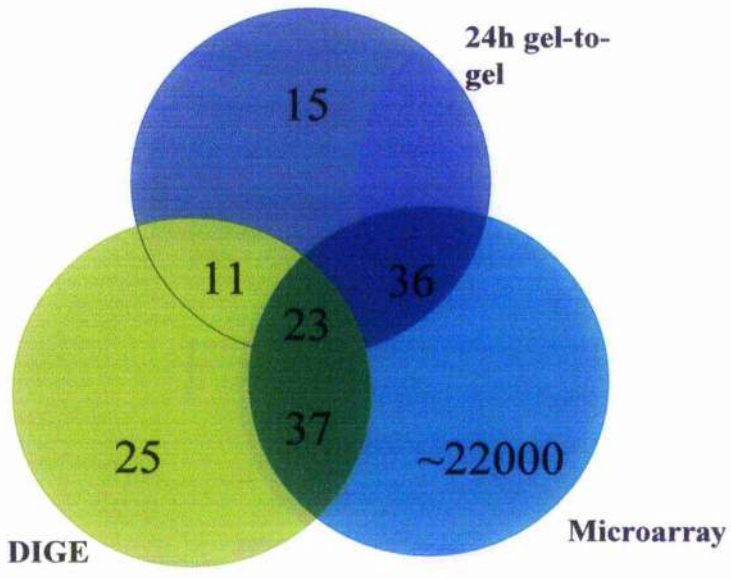


Figure 5-2 Venn diagram showing the distribution of modulated proteins identified using gel-to-gel electrophoresis and DIGE and all the genes arrayed. Sections which overlap indicate genes/proteins which match between techniques.

24h - 85 - unique pros across all expts

DIGE - 96

As was discussed in chapter IV, 23% of differentially expressed proteins 24hpi were common to both the gel-to-gel and DIGE experiments. This highlights the fact that the techniques are complementary. The data from the DIGE experiment is more statistically significant due to the inclusion of an internal standard and the sophisticated software used to analyse the gels. The fluorescent dyes used in DIGE are also more sensitive and linear than the colloidal Coomassie used in the gel-to-gel experiments.

As a preliminary comparison, the 96 modulated proteins which matched to genes on the array were compared. Of these, 35% agreed in the direction of change, i.e. both gene and protein expression were up regulated due to the presence of the parasite, or both gene and protein expression were down 24hpi. Fifteen percent of the matched genes and proteins differed in terms of expression direction. Fifty percent matched to clones on the array that had a volume change of less than 2 fold and therefore were considered not to be a significant difference.

There were 51 proteins differentially expressed 24hpi for which it was not possible to obtain a nucleotide accession number by any of the automated processes. For these proteins it was not possible to say whether they matched a gene product on the microarray, although it is highly unlikely that they matched a gene that exhibited a change in volume because these proteins were manually checked for possible matches, or a DoTS gene record was obtained but no microarray clone matched the same record. For these clones, it is likely that a matching clone was not present on the array, although it is also possible that there was a matching sequence that has been incorrectly mapped by DoTS.

In order to provide a more quantitative comparative approach, a comparison of the 24h DIGE results only and the 24h MA results was carried out. In the microarray experiment 567 known genes differed in volume 2fold or greater at 24hpi. The 24h DIGE showed 354 spots whose volume differed 2fold or greater. Of these 354, 184 spots were identified by MS. In several cases the same protein was identified at several different positions on the gel. Therefore 96 individual proteins species were detected as modulated in expression in the 24h DIGE experiment. Of these 96, 21% correlated in expression (e.g. down regulated in both gene and protein expression), 10% did not correlate, 29% were seen in both the MA and proteomic experiments, although did not change significantly in the MA study and 40% could not be matched between gene and protein. This data can be seen in Figure 5-3 and Table 5-2.

The MA experiment profiled approximately 22,000 known genes and uncharacterised expressed sequence tags. The DIGE experiment profiled approximately 2500 proteins some of which were isoforms encoded by a single gene. For example, Zyxin was identified at four different locations on the DIGE gel. Therefore the proteome analysis probably corresponds to less than 10% of the genes that were analysed on the microarray. However, the proteome analysis will contain information on post-translational modifications that are not tractable by MA.

Protein name	Gene name	24h DIGE	MA	Correlates?
6-phosphogluconolactonase	PGLS	down	down	Yes
Actin	ACTB	down	down	Yes
Adenylate kinase 2 isoform a	AK2	up	up	Yes
Aldehyde dehydrogenase 1A3	ALDH1A3	up	up	Yes
Aldose reductase	AKR1B1	up	up	Yes
BiP protein	HSPA5	up	up	Yes
Calponin isoform a	CNN2	down	down	Yes
Cathepsin B	CTSB	down	down	Yes
Elongation factor 1 alpha	EEF1A2	up	up	Yes
Enolase 1	ENO1	up	up	Yes
Glyceraldehyde 3P dehydrogenase	GAPD	up	up	Yes
Heat shock protein 27	HSPB1	down	down	Yes
Hnrp A2/B1 isoform A2	HNRPA2B1	up	up	Yes
HS70kd pro 42kd ATPase	HSPA8	up	up	Yes
Lamin A/C, isoform 2, 70kDa lamin	LMNA	down	down	Yes
MTHFD2 protein	MTHFD2	up	up	Yes
PDI related protein 5	PDIR	up	up	Yes
Phosphoglycerate kinase 1	PGK1	up	up	Yes
Porin 31HM	VDAC1	up	up	Yes
Tubulin alpha	TUBA1	up	up	Yes
Aldolase B	ALDOB	up		unchanged
ATP binding cassette transporter A13	ABCA13	up		unchanged
ATP synthase beta subunit	ATP5D	up		unchanged
B23 nucleophosmin (280 AA)	NPM1	up		unchanged
Capping protein (actin filament)	CAPZA1	down		unchanged
Carboxyl terminal LIM domain protein	PDLIM1	modulated		unchanged
Cathepsin D preprotein	CTSD	up		unchanged
CDC37 cell division cycle 37 homolog	CDC37	up		unchanged
Chloride intracellular channel 4	CLIC4	down		unchanged
Coiled-coil-helix-coiled-coil-helix 3	CHCHD3	up		unchanged
Collagen type 7	COL7A1	down		unchanged
Coproporphyrinogen oxidase	CPOX	up		unchanged
Dynactin 2	DCTN2	up		unchanged
ERprotein29 precursor	C12orf8	down		unchanged
Lrp protein	LRP1	down		unchanged
Methylumbelliferyl-acetate deacetylase	CHIT1	down		unchanged
Myristoylated alanine-rich C-kinase	MARCKS	up		unchanged
Nuclear chloride ion channel protein	CLIC1	down		unchanged
P47	NSFL1C	up		unchanged
Peroxiredoxin 6	PRDX6	modulated		unchanged
Pyridoxine 5'-phosphate oxidase	PNPO	down		unchanged
THAP domain containing 11	THAP11	up		unchanged
Thioredoxin domain 5 isoform 2	TXNDC5	up		unchanged
Transgelin	TAGLN	down		unchanged
Tubulin epsilon 5	TUBE1	down		unchanged
Ubiquilin 2	UBQLN2	up		unchanged
Visceral adipose-specific SERPIN	SERPINA1	modulated		unchanged
Zyxin	ZYX	down		unchanged
Acetyl coenzyme a acyltransferase 2	ACAT1	modulated	up	No
Cofilin 1	CFL1	down	up	No
Microtubule associated protein 1B	MAP1B	down	up	No

Peroxiredoxin 1	**PRDX1**	down	up	No
Peroxiredoxin 2	**PRDX2**	down	up	No
Ran-binding protein 1	**RANBP1**	down	up	No
Triose phosphate isomerase	**TPI1**	down	up	No
Tyrosine 3-monooxygenase/tryptophan 5	**YWHAH**	down	up	No
Vimentin	**VIM**	modulated	up	No
Clathrin, light polypeptide A isoform a	**CLTA**	down	up	No
Aldolase A	**ALDOA**	up		no match
Annexin 5	**ANXA5**	modulated		no match
Caldesmon	**CALD1**	down		no match
carbonyl reductase 1	**CBR1**	down		no match
Cardiotrophin like cytokine	**CLC**	modulated		no match
Desmin	**DES**	up		no match
dnak-type molecular chaperone HSPA6	**HSPA6**	up		no match
Dystrophin	**DMD**	up		no match
Galactose-specific lectin	**LGALS3**	up		no match
Heme binding protein 1	**HEBP1**	down		no match
Hnrp C-like protein	**HNRPC**	down		no match
High density lipoprotein binding protein	**HDLBP**	down		no match
Human Diff6,H5,CDC10 homologue	**NEDD5**	down		no match
Inter alpha (globulin) inhibitor H5	**ITIH5**	down		no match
Lamin B1	**LMNB1**	down		no match
Lasp-1 protein	**LASP1**	modulated		no match
Myosin alpha heavy chain	**MYH6**	up		no match
N-acetyl-beta-glucosaminidase	**HEXB**	down		no match
Neuropolypeptide h3	**PBP**	down		no match
Peroxiredoxin 3 isoform a precursor	**PRDX3**	down		no match
Progesterone receptor membrane	**PGRMC1**	up		no match
Protein-L-isoaspartate methyltransferase	**PCMT1**	up		no match
Pyrophosphatase (inorganic)	**PP**	down		no match
Pyruvate kinase M2 isozyme	**PKM2**	down		no match
RNA-binding protein regulatory subunit	**PARK7**	down		no match
Serum albumin precursor	**ALB**	up		no match
snf7dc2	**snf7dc2**	down		no match
Thyroid hormone binding protein	**ETHE1**	down		no match
Transgelin 2	**TAGLN2**	down		no match
Tropomyosin 1	**TPM1**	down		no match
Tropomyosin 2	**TPM2**	modulated		no match
Tropomyosin 3	**TPM3**	modulated		no match
Tropomyosin 4	**TPM4**	down		no match
Tumor rejection antigen (gp96) 1	**TRA1**	down		no match
Ubiquitin A-52 residue ribosomal protein	**UBA52**	up		no match
Ubiquitin carboxyl terminal esterase L1	**UCHL1**	modulated		no match
UMP-CMP kinase	**UMP-CMPK**	down		no match
UV excision repair protein RAD23 B	**RAD23B**	up		no match
Vacuolar protein sorting 33B	**VPs33B**	up		no match

Table 5-2 Table profiling the mRNA and protein comparison. All proteins which had a significant volume difference in *Toxoplasma gondii* infected cells in the 24h DIGE experiment were compared to the modulated genes present on the array. The table contains information on the gene name and whether the gene or protein was up regulated or down regulated. In the correlation column no match indicates no match for the protein was found on the array, unchanged on array indicates that a match for the protein was found on the array but the gene did not differ in volume significantly, No indicates that the gene and protein expression differs, Yes indicates that both the gene and protein behave similarly in expression.

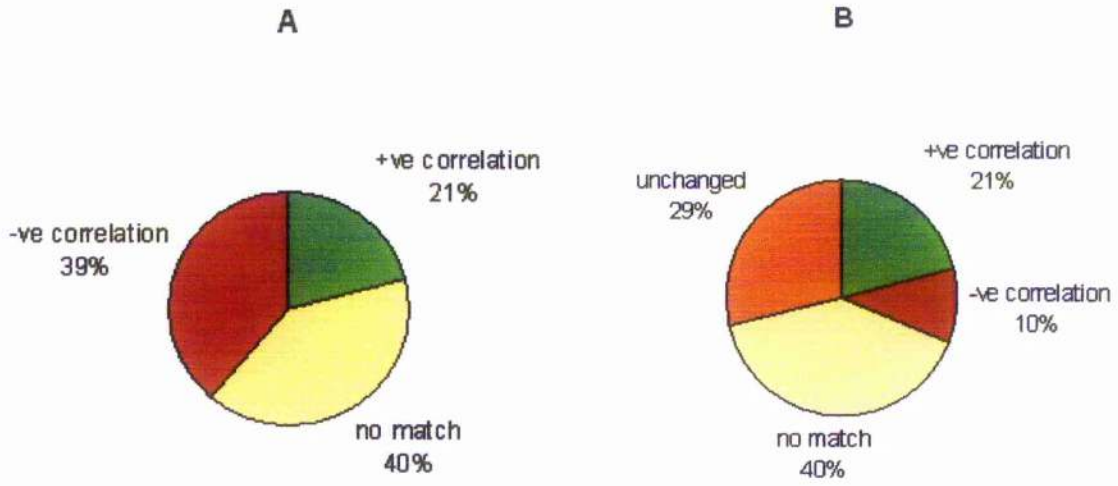


Figure 5-3 Pie charts representing the correlation of gene and protein expression. A. As can be seen in pie chart A 21% of modulated genes and proteins behaved similarly, e.g. both gene and protein were up regulated in the infected cell. 40% of the modulated proteins did not match to a gene on the array and 39% of the modulated proteins differed in expression response to the corresponding gene on the array. B. The genes and proteins which did not match in expression response were further categorised showing that 10% of the matched genes and proteins had converse expression e.g. the gene was up regulated and the protein down regulated in the pathogen infected cell. 29% of the modulated proteins matched to genes on the array which did not change in expression in the infected cell.

To investigate whether the volume changes in the gene and DIGE protein expression were correlated, two statistical tests were carried out. Only the data from the 24h DIGE experiment was included in this comparison as it provided accurate volume changes in protein expression. The volume ratios measuring the differences in gene or spot volume in the infected cell were used to compare the degree of change between gene and protein. The correlation between genes and proteins with similar expression was analysed. The 23 proteins chosen for the statistical analysis all had similar results (e.g. consistently up or down regulated in volume) across the DIGE experiment. Proteins which had a mixed volume response in the DIGE experiment (e.g. not clear overall whether the protein was up or down regulated due to infection) were discarded. The 23 pairs of data were then compared. For the microarray data, if the reported value lay between 0 and 0.05, volume was down regulated in the infected cell. If the value was >1, the volume was up regulated in the infected cell. mRNA abundance was represented by the expression levels as calculated by the Scanalyze program (Blader *et al.*, 2001). The proteomic data was converted into the same format as the MA data using the methods discussed in section 5.2.2.

Once the MA and proteomic values were in the same format they were ranked in order of highest to lowest value. Two statistical tests were performed. The null hypothesis was that there was no difference between the groups, therefore the gene and protein expression was correlated. The data here is unpaired and non-parametric (based on ranked data). Parametric tests require that the data from sample groups be normally distributed with equal variances, it is not possible to conclude that this is the case here, therefore non-parametric tests were carried out.

Table 5-3 shows the relationship between the 23 pairs of mRNA and proteins. Each experimental value was assigned a rank of 1-23 for both mRNA and protein and the ranks used to calculate the Spearman's rank correlation. For the Mann Whitney U test, the mRNA and protein values were ranked together e.g. 1-46.

Spearman's correlation was used to quantify association between two variables. Spearman's Rank correlation calculates an r-value. The closer r is to –1 or 1 then the stronger the correlation. A perfect positive correlation would be an r-value of 1, a perfect negative correlation would be an r-value of –1. The calculated r-value from our data was 0.3989 showing a weak positive correlation. The statistical package offered by http://faculty.vassar.edu/lowry/webtext.html was used to calculate the r-value.

The Mann-Whitney U test compares two unpaired groups. If the calculated U-value is greater than the tabled U-value then the null hypothesis is rejected. The calculated U-value was 320 and the table dU-value was 175 therefore the null hypothesis was rejected and it was concluded that there was no relationship between the volume differences of the gene and protein expression. The statistical package offered by http://faculty.vassar.edu/lowry/webtext.html was used to calculate the U value.

In conclusion there appeared to be a low level of positive correlation between gene and protein expression in terms of up or down regulation in the pathogen-infected cell. Very little correlation was found between actual volumes.

	DIGE		Microarray		
Protein name	**24h value**	**rank**	**MA 24h value**	**rank**	**mRNA and protein correlate**
Glyceraldehyde-3-phosphate dehydrogenase	12.16	1	3.7	2	Yes
Tubulin alpha	4.84	2	2.7	10	Yes
Elongation factor 1 alpha	4.76	3	3.175	8	Yes
Porin 31HM	3.4	4	2.4	13	Yes
PDI related protein 5	3.31	5	2.26	18	Yes
Heterogeneous nuclear ribonucleoprotein A2/B1	2.79	6	3.385	6	Yes
HS70kDa protein 8 isoform1	2.77	7	2.4	11.5	Yes
Aldose Reductase	2.73	8	4.74	1	Yes
BiP protein	2.39	9	2.4	11.5	Yes
Aldehyde dehydrogenase 1A3	2.15	10	3.53	4	Yes
MTHFD2 protein	1.78	11	3.35	7	Yes
ENO1 protein	1.42	12	3.076	9	Yes
Lamin A/C isoform 1 precursor	0.94	13	0.32	23	Yes
MAP1B protein	0.91	14	2.395	15	No
Peroxiredoxin 2 isoform b	0.67	15	2.1	19	No
Cathepsin B	0.66	16	0.47	21	Yes
6-phosphogluconolactonase	0.65	17	0.49	20	Yes
Triosephosphate Isomerase chain B	0.62	18	3.63	3	No
Calponin 2 isoform a	0.61	19	0.34	22	Yes
Ran-binding protein 1	0.59	20.5	3.52	5	No
Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase	0.59	20.5	2.4	14	No
Peroxiredoxin 1, natural killer enhancing factor A	0.56	22	2.298	16	No
Cofilin1	0.47	23	2.25	17	No

Table 5-3 Volume comparison of gene and protein expression differences. The correlation between matching mRNA and proteins was investigated. The expression volumes from the DIGE experiment were converted into the same format as the microarray values (DIGE 24h value column). All volumes were assigned a rank of between 1-23 for both gene and protein. These ranks were then used to calculate Spearman's rank correlation.

5.3.2 Confirmatory studies by Western blotting

Protein extracts from *Toxoplasma gondii* infected HFFs and non infected HFFs were subjected to 1-DE and 2-DE and blotted onto Hybond membranes. The membranes were incubated with specific antibodies to three of the proteins of interest from the proteomic experiments. A major bottleneck in confirmatory studies is the lack of antibodies for many of the identified proteins (Alfonso, 2005) therefore this had to be taken into account when choosing which proteins to investigate further. Figure 5-4, Figure 5-5 and Figure 5-6 show the results obtained by Western blotting with caldesmon, cathepsin and GAPDH in 1-DE. It is desirable to run a protein extract on the gel that is constitutively expressed in all the samples of interest for use as a control. It was not possible to run internal protein standards on the Western blots as the traditional house-keeping genes i.e. GAPDH, tubulin beta and actin beta (Ferguson *et al.*, 2005; Ikemoto *et al.*, 2003), all varied in expression in the parasite infected cell. Therefore no internal standard was included to monitor constitutive protein expression.

Three proteins of interest were chosen for further study; caldesmon, cathepsin B and glyceraldehye-3-phosphate dehydrogenase (GAPDH). These proteins were selected on the basis of changes in expression at the gene or protein level and that the function of the protein was of biological interest in the infected cell.

5.3.3 1-DE Western blotting

5.3.3.1 Caldesmon

Caldesmon (CALD) is an actin and myosin binding protein that plays an essential role during cellular mitosis and receptor capping (http://harvester.embl.de/). The protein occurs in both a high molecular weight (120-150kDa) and a low molecular weight (71-80kDa) form, depending on the tissue in which it is located (SIGMA product data sheet). Protein expression was down regulated in *Toxoplasma gondii* infected host cells. In the 24h DIGE experiment caldesmon was down regulated 1.79 fold in infected cells. However, a difference in mRNA expression was not seen in the MA experiments. The human caldesmon protein when blasted against the ToxoDB database has 49% protein homology with *Toxoplasma* protein TgTwinScan_4914 (www.toxodb.org). 1-DE gels were run with extracts from *Toxoplasma gondii* infected and non infected cells and pure parasite extract (Tg1) was also run as a control to ensure that there was no cross reactivity with *Toxoplasma* protein. Proteins were transferred onto Hybond membranes and incubated

with caldesmon antibody. The blot in Figure 5-4 shows that the *Toxoplasma* protein extract had a higher signal for the parasite protein than the HFF extracts. This antibody was supposed to be specific to human caldesmon but was clearly not of good specificity giving multiple bands in the infected and non-infected lanes. The sizes of the bands at 120kDa and 71kDa correspond to the appropriate sizes of human caldesmon forms. These bands should be stronger in the non infected lanes but it is not possible to confirm this from this blot. The experiment was repeated several times, always with the same result.

5.3.3.2 Cathepsin B

Cathepsin B (CTSB) is an acid lysosomal proteinase which is expressed in all cells and is active in intracellular protein breakdown. The protein is classed into the functional category of protein metabolism and has been shown to be involved in apoptosis as expression can enhance the mitochondrial death pathway and phosphatidylserine exposure which are involved in the final step of apoptosis (http://harvester.embl.de/). CTSB is synthesised as an inactive 43kDa pro-enzyme which, by removal of a 623 amino acid propeptide, is activated to the single-chain form of 31kDa or the two chain form of 25 and 5kDa subunits. The human cathepsin B protein when blasted against the ToxoDB database showed 61% homology with *Toxoplasma* protein TgTwinScan_4336 (www.toxodb.org). An inhibitor to cathepsin B is available called Ac-LVK-CHO. This compound is a lysinal analogue of leupeptin and a potent inhibitor of cathepsin B (McConnell *et al.*, 1993).

Cathepsin B expression was down regulated in both gene and protein expression in the pathogen infected cell. Figure 5-5 shows a Western blot with cathepsin B reacting with HFF proteins separated by 1-DE. As can be seen from the blot, there was very little signal in the non-infected lanes compared to the infected. This experiment was repeated several times with the same results. This result contradicts the proteomics DIGE result where CTSB was down regulated by 1.44 fold.

5.3.3.3 GAPDH

GAPDH is one of the key enzymes involved in glycolysis and catalyses glyceraldehyde-3 phosphate to 1,3-bisphosphoglycerate. Besides functioning as a glycolytic enzyme in the cytoplasm, mammalian GAPDH is also involved in a great number of intracellular processes such as membrane fusion, microtubule bundling, phosphotransferase activity, nuclear RNA export, DNA replication and DNA repair. While GAPDH is constitutively expressed in almost all tissues at high levels, there are some physiological factors such as

hypoxia and diabetes that increase GAPDH expression in certain cell types (CALBIOCHEM data sheet). GAPDH is comprised of four 36kDa subunits with a total molecular weight of 146kDa. GAPDH expression was up regulated in *Toxoplasma* infected cells in both the microarray and proteomic studies.

Figure 5-6 shows the Western blot of GAPDH which shows bands of the correct size of 36kDa for GAPDH protein. The protein was seen as up regulated in infected cells using proteomics. The signal is stronger in the infected lanes than in the non-infected lanes. Image master 2DElite software was used to calculate the spot volume of the bands and the data can be seen in Table 5-4. The volume of the GAPDH band from infected HFFs is indeed greater than the volume of the GAPDH band from non infected HFFs.

Band	Spot volume
Inf1	486281
Non1	301465
Inf2	488087
Non2	224034

Table 5-4 Volumes of the bands corresponding to GAPDH by western blotting from infected and non infected HFFs.

GAPDH enzyme activity can be inhibited by iodoacetate (Ikemoto *et al.*, 2003; Chang *et al.*, 2000) therefore iodoacetate could possibly be used in an inhibition study.

5.3.4 2-DE Western blotting

No protein signal was detected on the 2-DE western blots for any of the three proteins of interest. Several different primary and secondary antibody concentrations were tested. Due to time constraints it was not possible to continue with the 2-DE Westerns. Further optimisation of the technique may have yielded results.

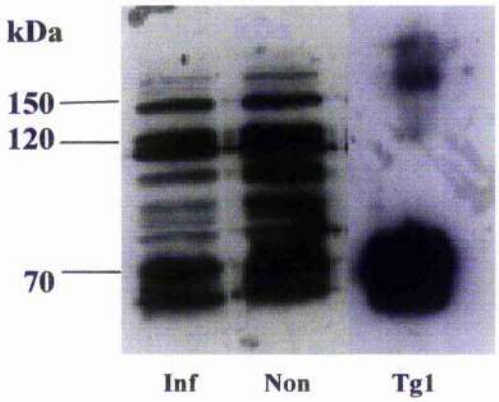


Figure 5-4 Western blot analysis to determine levels caldesmon protein expression. Extracts from *Toxoplasma gondii* infected and non infected cells were separated on 1-DE gels. Pure parasite extract (Tg1) was run as a control to check for cross hybridisation. Proteins were transferred onto Hybond membranes and incubated with caldesmon antibody. This blot shows that the caldesmon antibody has a higher affinity for the *Toxoplasma* protein than the HFF extracts and that the antibody was very non specific.

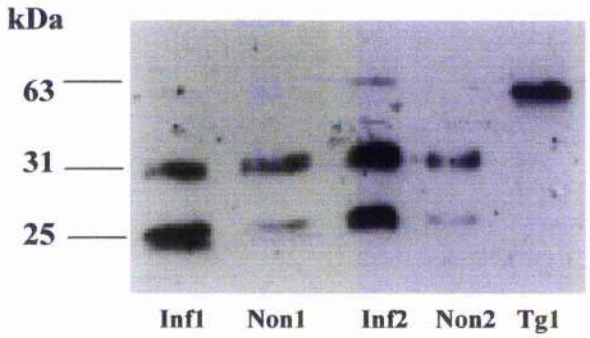


Figure 5-5 Western blot analysis to determine protein expression of cathepsin B. Extracts from *Toxoplasma gondii* infected and non infected cells were separated on 1-DE gels. Pure parasite extract (Tg1) was also run as a control to check for cross hybridisation. Proteins were transferred onto Hybond membranes and incubated with cathepsin B antibody. This blot shows bands of the correct sizes of cathepsin B, 31kDa and 25kDa, however the signal should be stronger in the infected lanes to correspond with the MA and proteomic results. This could be an example of a gene and protein that do not correlate in expression in the pathogen infected cell.

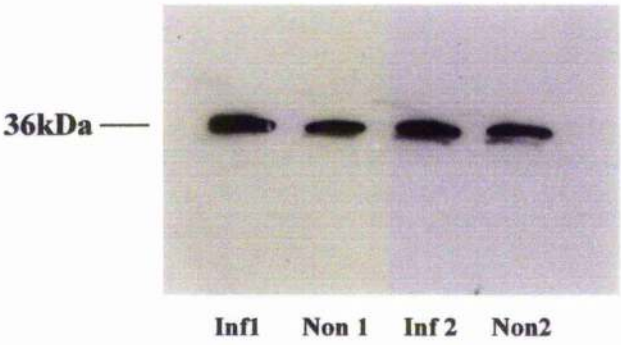


Figure 5-6 Western blot analysis to determine level of GAPDH protein expression. Extracts from *Toxoplasma gondii* infected and non infected cells were separated on 1-DE gels. Pure parasite extract was also run as a control but no signal was detected so the area has been cropped from the image. Proteins were transferred onto Hybond membranes and incubated with GAPDH antibody. The blot shows a band of the correct size of 36kDa for GAPDH protein. This protein was identified as up regulated in infected cells by proteomics. The signal is indeed stronger in the infected lanes in the Western blot therefore confirming the MA and proteomic results.

5.4 Discussion

This chapter investigated the correlation between mRNA and protein expression in *Toxoplasma gondii* infected HFF cells. The proteomic studies presented in this thesis mirrored the gene expression studies of Blader *et al* (2001). Comparing all of the differentially expressed proteins from the proteomic experiments with the matched differentially expressed genes on the microarray gave a relatively low level of correlation (34%). A more accurate method of comparison was to compare the actual gene and protein volume differences between each matching mRNA and protein pair and this showed that there was a weak, albeit significant, level of correlation. Western blotting was further used to confirm changes seen in the MA and proteomic studies. The results in this chapter highlight the fact that studies which look at only one parameter e.g. gene expression or protein expression, do not provide the whole picture. To look at the global response these techniques must be used in concert as they provide complementary results.

Comparison of mRNA and protein abundance levels can help to discern the role of post-transcriptional regulation and therefore high-throughput transcript and protein expression profiling techniques are powerful tools to study this (Le Roch *et al*, 2004). The correlation between mRNA and protein expression is still largely unknown (Alfonso *et al*., 2005). Several studies have compared gene and protein expression and most have reported a weak correlation between mRNA and protein levels (Cox *et al*., 2005; Gygi *et al*., 1999). The level of correlation may however depend on the organism. The aim of the work in this chapter was to determine the level of gene to protein correlation in cells during infection with *Toxoplasma gondii* and therefore gain a better understanding of the functional modifications in infected host cells and what contribution post-translational modifications make to the protein expression profile.

It is very difficult to gain a true understanding of the gene to protein correlation when comparing microarrays with the type of proteomic experiments used here. Microarrays can profile tens of thousands of genes. There are currently no protein profiling technologies available that can match this kind of detail (Greenbaum *et al*., 2003) and 2-DE can resolve around 3000 spots on a gel (although many of these will be isoforms of the same protein (Jenkins and Pennington, 2001). Microarrays provide a wealth of data about the gene abundance; however it is now accepted that gene array and proteomic technologies are complementary, and the protein expression ought to also be studied as the protein concentration and protein-protein interactions are most likely the true causative forces in

the cell (Greenbaum *et al.*, 2001). It has been noted that in some cases 40-fold changes in mRNA concentration do not lead to any change in protein concentration (Predic *et al.*, 2002).

In this study, the first comparison of differentially expressed proteins and genes gave a correlation of 34%. One major problem in this analysis was that there was no data for genes which were present on the array but perhaps had a poor signal. Therefore it is not possible to conclude that because a match was not found on the array for a specific protein, that the gene expression was not modulated, as it may not have been present on the array or through experimental error no signal was detected. Over half of the matched proteins corresponded to genes which did not change in expression at any of the time-points investigated in the study by Blader *et al* (2001). The second comparison investigated the correlation between the actual volume changes of matched mRNA and proteins. Using Spearman's rank correlation and the Mann Whitney U test, a weak level of correlation was observed. In general, it was possible to identify about 50% of the differentially expressed spots in the 24h DIGE experiment. Therefore there may be many more proteins which were modulated in expression due to the presence of the parasite that it was not possible to compare to the genes present on the microarray.

No strong correlation was seen between the MA and proteomic data. This was expected as there are at least four reasons to explain the poor correlation between mRNA and protein; (1) proteomic bias, (2) mRNAs differ in their rates of translation into protein, (3) proteins differ in their in vivo half lives and (4) there is a significant amount of error and noise in both protein and mRNA experiments that limit the ability to get a clear understanding of the gene to protein correlation (Greenbaum *et al.*, 2001).

Protein can be controlled by the cell at the level of transcription or translation. It is assumed that proteins which show a large degree of variation in their expression are controlled at the transcriptional level. The cell has already put significant energy into dictating the final level of protein through tightly controlling the mRNA expression, and it is therefore assumed that there would be minimal control at the protein level. In contrast, genes that show minimal variation in their mRNA expression are more likely to have little or no correlation with the final protein level; the cell is controlling these genes at the translational and/or post-translational level, therefore the mRNA levels are somewhat independent of the final protein concentration (Greenbaum *et al.*, 2003).

The general lack of correlation between mRNA and protein abundance may also be explained by the fact that proteins have different half-lives as the result of varied protein synthesis and degradation rates. Understanding what role degradation and synthesis play in protein expression is essential to fully interpret protein abundance data from proteomic experiments (Pratt *et al.*, 2002). Protein levels may change because of a change in the rate of protein synthesis, mRNA or protein turnover. Protein turnover can vary significantly as the cell can control the rates of degradation or synthesis for a given protein, and there can be significant differences even within proteins that have similar functions. Measurement of these rates has been attempted recently. In general, change in protein concentration over time will be equal to the rate of translation minus the rate of degradation. By analogy to concepts in chemical kinetics this equation has been proposed:

$$dP(i,t)/dt = SE(i,t) - DP(i,t)$$

P is protein abundance i at time t, E is the mRNA expression level of protein P, S is a general rate of protein synthesis per mRNA and D is a general rate of protein degradation per protein. This equation is of course highly simplified and in a more general context it would be expected that synthesis and degradation rates will differ for each gene/protein and be dependant on the regulatory effects of other genes/proteins over time. In addition, the equation does not take into account the stochastic nature of gene and protein expression (Greenbaum *et al.*, 2001; Pratt *et al.*, 2002; Greenbaum *et al.*, 2003).

Proteomic analysis generally deals with steady state protein levels. Pratt *et al* (2002) used stable isotope labelled amino acids to determine the rate of breakdown of individual proteins by inspection of mass shifts in tryptic fragments. They determined that the average rate of protein degradation was 2.2% per hour, although some proteins had rates up to 10% per hour. They concluded from this study that protein turnover is a significant missing dimension in proteomic experiments and should therefore be considered when assessing protein abundance data and comparing it to the relative abundance of similar mRNA species (Pratt *et al.*, 2002).

Several studies have shown that for a substantial number of genes the absolute amount of protein in the cell is not strongly correlated with the amount of mRNA. These conclusions are based on simultaneous measurement of mRNA and protein at just a single time point. The study by Lian *et al* (2002) investigated the simultaneous changes in mRNA and protein relationship over several time points. This resulted in a much stronger correlation,

consistent with the hypothesis that a substantial proportion of protein change is a consequence of changed mRNA levels, rather than PTM (Lian *et al.*, 2002).

To confirm further the changes seen in the MA and proteomic experiments, three proteins were chosen for expression analysis using Western blotting (WB). The results show three examples of WB outcome. (1) The up regulation of GAPDH (as shown by the gene and protein expression studies) in *Toxoplasma gondii* infected cells was confirmed by WB. (2) The antibody used to detect caldesmon was extremely non-specific to the human protein (although a specific monoclonal antibody had been selected) therefore it was not possible to conclude the change in protein expression from the WB of this protein. (3) The cathepsin WB results differed from what was expected. The MA and proteomic experiments showed a down regulation in cathepsin B gene and protein expression, the western blot results contradicted this showing a stronger signal in the infected lanes indicating an up regulation of protein expression. It could be that different isoforms of the protein were changing in expression in parasite infected cells, this would not be detected using Western blotting. As many of the spots analysed by MS (particularly in DIGE) were mixed proteins, this could account for the CALD and CTSB results seen in the Western Blotting. Mixed spots highlight how important it is to do further confirmatory studies. Western Blotting is a much easier technique to carry out than fractionation of sample and gels.

There are several reasons why the WB may not have shown the expected results for CALD and CTSB; the concentration of protein or antibody, and specificity, or the development time may not have been optimal. This could have been resolved with further optimisation of the technique. Isoforms which appear at different positions on 2-DE gels are often grouped together on 1-DE westerns and therefore it may be difficult to differentiate between isoforms of interest. WB can often struggle with sensitivity over small fold differences. DIGE and MS can overcome these issues. DIGE can identify small fold differences and using MS, human and *Toxoplasma gondii* proteins can be differentiated between. To conclude, for the Western blot data, it is clear that the host cell response may be more complex than a protein being up or down regulated due to infection.

Several 2-DE westerns were carried out and no signal was detected for any of the three proteins of interest. Several different primary and secondary antibody concentrations were tried. Due to time limits it was not possible to continue with the optimisation of this protocol.

If time had allowed it would have been extremely interesting to over express or inhibit the expression of these three proteins of interest in the pathogen infected cell to determine whether the expression was essential for the survival of the parasite or the survival of the host cell and if not essential, to determine whether the change in protein expression was due to a bystander effect (Roos *et al.*, 1994). Growth studies of the parasite with the inhibited/over expressed protein cell line could have been measured using tritiated hypoxanthine, a naturally occurring purine derivative that is selectively taken up and incorporated by the growing parasite and not the host cell. Therefore specifically parasite growth can be measured. Another possibility would have been to use two strains of *Toxoplasma gondii*; virulent and avirulent, to determine whether the proteins of interest differ in expression in infected cells according to which parasite strain infects.

In conclusion, in this chapter we compared the steady state proteome with the host cell gene expression. The infected cell is a dynamic system and therefore the steady state proteome provides merely a snapshot of the host cell protein response at a given time point. This also contributes to the difficulty of comparing the proteomic data to the MA study. Further studies using [^{35}S] methionine labelling to detect the newly synthesised proteins in the infected host cell would give a more comprehensive insight into which proteins are being synthesised due to the presence of the parasite. As the amount of incorporated radiolabel is measured, new protein synthesis is detected independent of the total amount of a given protein that is present (Predic *et al.*, 2001). Comparison of the newly synthesised proteins with the differentially expressed genes may show a different correlation result than was seen when comparing the steady state proteome.

Chapter VI The proteome of *Leishmania* sp infected cells

6.1 Introduction

The aim of the work in this chapter was to investigate the uniqueness of the host response to *Toxoplasma gondii*. This was done by profiling the host cell protein response to a second intracellular protozoan pathogen, *Leishmania major*, and comparing this response to the *Toxoplasma* infected cell protein profile.

An extremely important biological question is whether cells have a common response to intracellular pathogens (parasites, viruses or bacteria) or whether the cellular response is unique to individual invading pathogens. Similarities in response could prove useful for drug targets or vaccine design. Some work has been done using microarrays to investigate the host transcriptional response to individual pathogens (Blader *et al.*, 2001; Eckmann *et al.*, 2000; Zhu *et al.*, 1998). Blader *et al* (2001) compared the host cell gene expression of *Toxoplasma gondii* infected cells with *Trypanosoma cruzi* infected cells using microarrays. They found only a 4% correlation in gene expression between the two pathogen infected cell sets, and very few changes in host cell gene expression early in infection. This result was surprising to the investigators as a general feature of host-pathogen interactions (as seen in existing microarray analyses) is the quite rapid induction of host cell gene expression (de Avalos *et al.*, 2002). These results indicate that the host response is tailored to individual pathogens at the level of gene transcription. However, many of the early responses to infection may take place at the protein level.

6.1.1 Leishmania sp life cycle

Leishmania sp are intracellular pathogens of the phylum Kinetoplastida. Infection with *Leishmania* results in a spectrum of human diseases, called the Leishmaniases, with two million cases in 88 countries annually (Ivens *et al.*, 2005). *Leishmania* parasites exhibit two forms in their life cycle, (Figure 6-1) amastigotes which reside in macrophages of the mammalian host and promastigotes which are found in the gut of the vector, the sandfly (Diptera, Psychodidae). Whilst feeding the infected fly vector injects metacyclic promastigotes into the human. In contrast to the active invasion of cells by *Toxoplasma gondii*, *Leishmania* promastigotes induce phagocytosis by macrophages. Once inside the cell the promastigotes transform into amastigotes and commence repeated binary division.

Within the cell the parasite resides within a parasitophorous vacuole (PV) which is surrounded by host membrane. The PV fuses with lysosomes and the amastigotes can survive within the acidic environment of the phagolysosome, despite the presence of powerful lysosomal enzymes. When infected cells eventually rupture, amastigotes are released and are taken up by new macrophages.

6.1.2 Leishmania-host cell interaction

The aim of this chapter was to compare the host cell response to invasion by *Leishmania sp* and *Toxoplasma gondii* using proteomic techniques. Two approaches were used to investigate the host cell protein expression response to infection; conventional gel to gel 2-DE and DIGE. HFFs were infected with *L.major* promastigotes at several infection ratios and time points. Infected cells were examined and photographed to confirm the presence of intracellular parasites. *L.major* infected and non-infected HFFs were compared using 2-DE gels and then by DIGE. The proteome of *Leishmania* infected cells was then compared with the *Toxoplasma* infected HFF protein profile.

Fibroblasts are not a typical host cell type for *Leishmania sp,* however, a study by Bogdan *et al* (2000) proposed that fibroblasts may serve as host cells in latent leishmaniasis. Chang *et al* (1978) also documented that human skin fibroblasts were able to be infected by *L.major* (Chang, 1978). Macrophages are generally considered to be the most important host cell for *Leishmania* parasites. Dendritic cells, epithelial cells, eosinophils, and neutrophils have also been shown to harbour *Leishmania* both in vivo and in vitro. Using confocal laser microscopy, Bogdan *et al* (2000) found that mice which had healed a cutaneous *L.major* infection continued to have parasites persisting in the fibroblasts in the lymph nodes. In vitro, both promastigotes and amastigotes of *L.major* infected primary skin or lymph node fibroblasts. The findings suggested that fibroblasts may provide a safe environment for the parasites during the chronic phase of infection (Bogdan *et al.*, 2000). A proposed microbial survival strategy is the ability of a pathogen to enter and reside within cells that do not elicit an immune response.

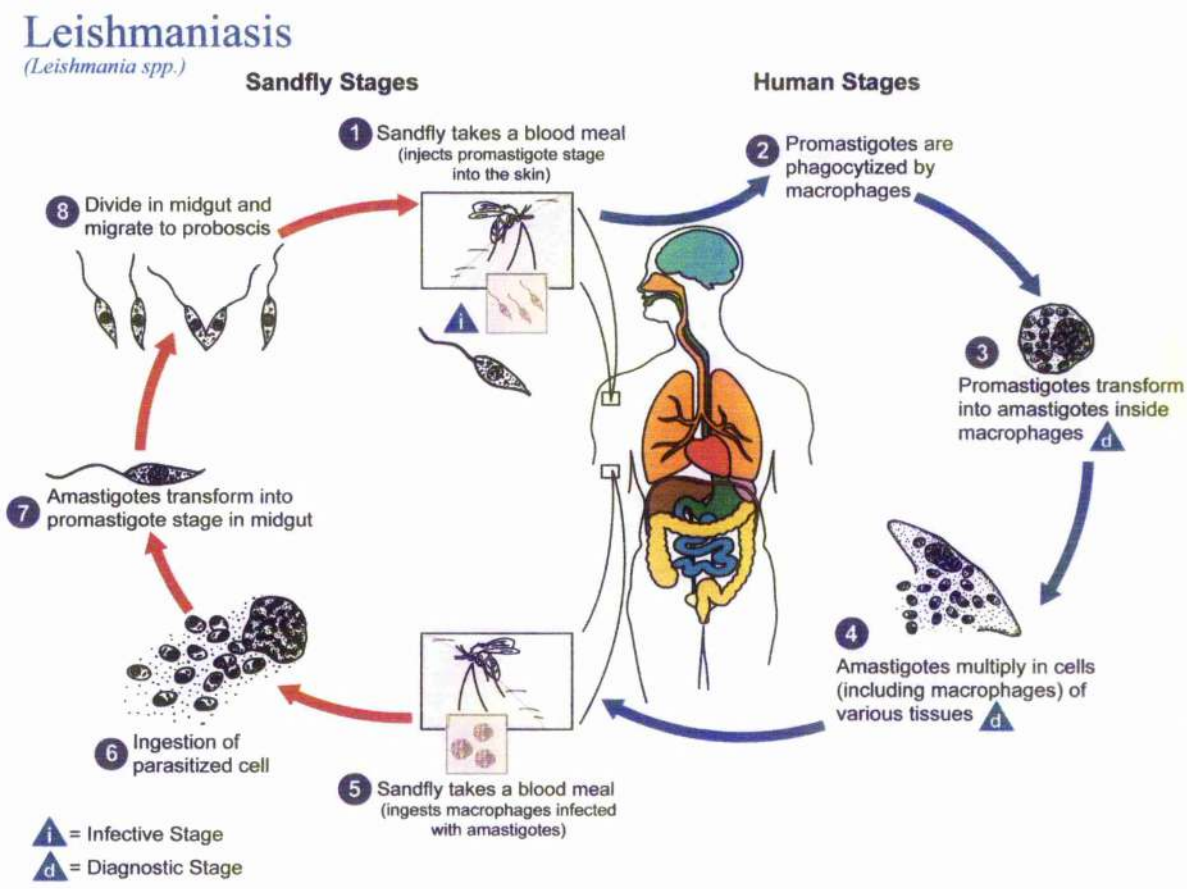
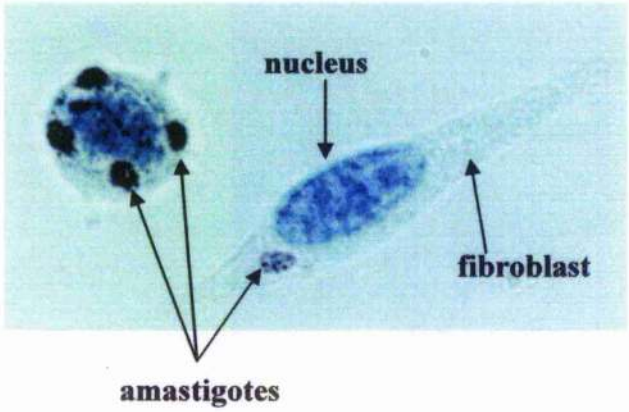


Figure 6-1 Life cycle of *Leishmania sp.* (Image taken from Hailu, 2005 (Hailu *et al.*, 2005)) Metacyclic promastigotes are injected into the human as the sandfly feeds. The promastigotes are actively phagocytosed by macrophages and reside within a parasitophorous vacuole within the cell, where they differentiate into amastigotes and replicate until the cell lyses. The released amastigotes are then phagocytosed by new macrophages or taken up in the blood meal by a sandfly. In the fly, the amastigotes differentiate into promastigotes within the gut.

A.



B.

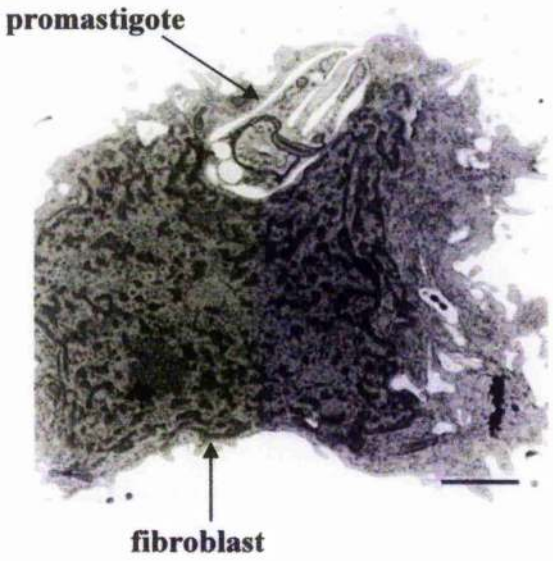


Figure 6-2 Images of *L.major* infected fibroblasts taken from the paper by Bogdan, C. *et al*, 2000. A. Skin fibroblasts 48 h after infection with amastigotes. Intracellular amastigotes are visible. B. Reticular lymph node fibroblasts 24 h after infection with promastigotes. An intracellular promastigote is visible. Bar, 2.1 μM (Bogdan, C. *et al*, 2000)

Leishmania was chosen as the intracellular pathogen for this study for several reasons. Firstly, *Leishmania* is an intracellular parasite that is known to survive within fibroblasts (Bogdan *et al*, 2000) and using HFFs allows for easy comparison between the HFF protein response to *Leishmania* and *Toxoplasma*. Secondly, *L.major* and *L.mexicana* were available to us in this study and are relatively easy to culture. Disadvantages of using *Leishmania* as the intracellular pathogen include the fact that *Leishmania* and *Toxoplasma* have quite different lifestyles. *T.gondii* actively invades cells whereas *Leishmania* is taken up by phagocytosis. *Toxoplasma* resides within a parasitophorous vacuole made entirely of parasite proteins. *Leishmania* also resides within a parasitophorous vacuole, which in this case is formed as a result of targeted fusion of host cell lysosomes with the plasma membrane. It has also been documented that 48h post infection of fibroblasts with *L.major* that only up to a maximum of three amastigote parasites were seen in a cell. This is indicative that the parasites are not replicating within the fibroblasts at the same rate as they would do if in a more typical host cell, e.g. a macrophage where *Leishmania* parasites double every 12-24h (R.Burchmore, personal communication). *Toxoplasma* replicates by endodyogeny in fewer than 8h (Black and Boothroyd, 2000). We would therefore expect that the protein profiles of host cells infected with the different pathogens would be quite different. However, there are several elements of the host response to the two pathogens that may be similar. For example, an induction of heat shock proteins or remodelling of host cell constituents due to the formation of the parasitophorous vacuole. *L.donovani* and *T.gondii* have both been shown to modulate host cell apoptosis (Bogdan and Rollinghoff, 1999).

In macrophages, *L.donovani* amastigotes have been shown to interfere with host cell gene expression (Buates and Matlashewski, 2001). Using arrays, 588 macrophage transcripts were analysed from parasite infected and non-infected cells and a down regulation of 40% of the arrayed genes was seen in infected cells. House-keeping gene expression remained consistent and a small number of genes were up regulated due to infection. The investigators propose that a similar set of results would have been seen if the entire genome had been analysed (Buates and Matlashewski, 2001). Table 6-1 compares the differentially expressed genes in *L.donovani* and *T.gondii* infected cells from the MA studies by Buates and Matlashewski (2001) and Blader *et al* (2001) respectively. As can be seen from the table, only 16 differentially expressed genes were common to both studies and of these only 4 behaved similarly. This indicates that the host cell has quite a different cellular response to individual pathogens at the gene level.

Protein/Gene name	Gene response	Fold Regulation	*T.gondii* MA study	MA studies correlate?
Bone morphogenetic protein 7	DOWN	-14.9	down 24h	YES
Cytochrome P450 1A1	UP	2.3	down 24h	NO
G1/S-specific cyclin D1	DOWN	-2.2	up 24h	NO
GTPase-activating protein	DOWN	-2.8	down 24h	YES
IL-1	DOWN	-2.2	up 24h	NO
IL-6R	DOWN	-2.4	up 24h	NO
IL-7 receptor	DOWN	-4.5	up 24h	NO
Insulin-like growth factor receptor II	DOWN	-4.1	down 24h	YES
Mad-related protein 2	DOWN	-3.7	up 24h	NO
Membrane-type matrix metalloproteinase 1	DOWN	-2.2	up 24h	NO
Phosphotyrosine phosphatase	DOWN	-2.1	up 24h	NO
Ribosomal protein S6 kinase II 1	DOWN	-3.5	up 24h	NO
RXR- *cis*-11-retinoic acid receptor	DOWN	-2.1	down 24h	YES
Translin	DOWN	-2.2	up 24h	NO
Ubiquitin-conjugating enzyme E2	DOWN	-2.8	up 24h	NO
Vascular endothelial growth factor precursor	DOWN	-2.1	up 24h	NO

Table 6-1 Comparison of differentially expressed genes in *L.donovani* and *T.gondii* infected cells as determined by microarray analysis (Buates and Matlashewski, 2001; Blader *et al.*, 2001).

Investigating the proteome of *Trypanosoma cruzi* infected HFFs would perhaps have been a more ideal choice to compare to *T.gondii* infected cells as this comparison was also done at the gene expression level by Blader *et al*. However as a category 3 organism, it was not possible to culture *T.cruzi* in our laboratories. *T.cruzi* and *Leishmania* are quite closely related as both are protozoan parasites from the family Trypanosomatidae therefore comparison of the *Leishmania* infected host protein profile to the *T.cruzi* infected host transcriptional profile will be of interest in investigating any common host response.

6.2 Materials and Methods

6.2.1 Infection of HFFs by Leishmania major

To determine whether HFFs could be host cells for *Leishmania major* promastigotes, HFFs were incubated with different ratios of *L.major*. HFFs (p<20) were grown in NunTek chamber slides and infected with different ratios of *Leishmania major* promastigotes (0:1, 2:1, 5:1 or 10:1). The HFFs plus parasites were left in DMEM at 32°C, 5% CO_2 for 4h. The medium was then removed from all wells and the cells washed three times in fresh DMEM to remove any extracellular parasites. Cells were then left in fresh DMEM for 48h at 32°C, 5% CO_2. After 48h, the medium was removed and the cells were Geimsa stained as described in Chapter II. Using a Zeiss Axiovert 200M light microscope, images of the non-infected HFFs and *Leishmania major* infected HFFs were captured at 200x and 400x magnification. The number of parasites per cell was counted using a light microscope at 1000x magnification under oil immersion.

6.2.2 Conventional gel-to-gel 2-DE

Leishmania infected and non-infected cells were harvested and the proteins prepared for 2-DE using the method discussed in Chapters II and III. An infection ratio of 7 parasites: 1 HFF cell was used. *Leishmania* infected HFFs were compared to non-infected HFFs using 2-DE (7cm and 24cm pH 3-10NL IPG strips). The infected and non-infected images were compared by eye and also using the 2DElite software. Spots which were differentially expressed were chosen for MS analysis.

6.2.3 DIGE comparison of Leishmania major infected and non-infected cells

Confluent HFFs (passage number 18) were infected with *Leishmania major* promastigotes at an infection ratio of 2 parasites: 1 cell. The cells and parasites were left at 32°C, 5%CO_2 for 24h and then washed with fresh DMEM to remove any extracellular parasites. Cells were then left for 3 days at 32°C, 5%CO_2 and washed with fresh DMEM every 24h to remove any remaining extracellular parasites.

Cells were harvested as described in Chapter II. The samples were prepared for 2-DE using the DIGE lysis method detailed in Chapter IV. The Amersham 2D Clean-up kit was used

to remove any contaminants from the sample. The protein concentration of the samples was determined using the BCA assay as 3.9µg/µl and the pH of the samples confirmed as between 8 and 9. 50µg of the infected sample was labelled with Cy3 and 50µg of the non-infected sample labelled with Cy5. The labelling reaction took 30min (on ice in the dark) at which point it was stopped by adding 1µl 10mM lysine and left on ice for 10min. Although only one DIGE gel was run, a pooled internal standard of the two samples of interest was included and labelled with Cy2.

The preparative and DIGE samples were run on the same gel. The three CyDye labelled samples were mixed ready for IEF and 184µg of both infected and non-infected sample (the preparative sample) were mixed together and added to the CyDye labelled samples. This collective sample was then mixed with an appropriate volume of rehydration buffer including DTT and IPGs pH4-7. Separation by first and second dimension was carried out as detailed in Chapter IV.

To detect the total protein on the gel, the gel was stained with SYPRO Orange. The gel was fixed overnight in 40% MeOH, 10% acetic acid, 50% ddH_2O. The gel was then washed in 0.5% SDS for 1h, washed briefly in 7.5% acetic acid and then stained for 1h in SYPRO Orange total protein stain (1 in 10000 dilution) in 7.5 % acetic acid. The gel was washed briefly in 7.5% acetic acid and scanned using the Typhoon scanner at the Cy3 setting: 580 BP 30 emission filter, 532nm green laser.

The CyDye images were analysed using the DIA software and all of the labelled spots present on the gel detected. The volume of each spot was calculated and any differences of two fold or greater between the matched spots on the infected and non-infected images was marked as differentially expressed. The spot map of these differentially expressed proteins was then exported to the BVA module. All of the spots on the preparative scan were detected using DIA and the resulting spot map exported to BVA where it was matched to the CyDye images. The differentially expressed spots were then excised from the gel using the Ettan Spot Handling Workstation (Amersham Biosciences) and the proteins digested with trypsin and analysed using a combination of MALDI and MS/MS.

6.2.4 DIGE comparison of Leishmania mexicana infected and non-infected cells.

Confluent HFFs (passage number 18) were infected with *Leishmania mexicana* promastigotes at an infection ratio of 10 parasites: 1 cell. The cells and parasites were left

at 32°C, 5%CO_2 for 4h and then washed with fresh DMEM to remove any extracellular parasites. Cells were then left for 48h at 32°C, 5%CO_2 and washed with fresh DMEM every 24h to remove any remaining extracellular parasites. Cells were harvested as described in Chapter II. The samples were prepared for 2-DE using the DIGE lysis method detailed in Chapter IV. The Amersham 2D Clean-up kit was used to remove any contaminants from the sample. Proteins from infected and non-infected cells were labelled with Cy5 and Cy3 respectively. A pooled sample of infected and non-infected proteins was labelled with Cy2.

6.3 Results

6.3.1 HFF infection with L.major

HFF cells were grown in chamber slides and infected with different ratios of *L.major* metacyclic promastigotes. After 48h, the cells were stained and the number of parasites per cell was counted. Table 6-2 shows the number of intracellular amastigotes and extracellular promastigotes counted in HFFs which were infected at a ratio of 10 parasites: 1 HFF. To determine the ratio of infection in these cells, the number of amastigotes and host cells were counted in ten different positions per well under the microscope. The percentage of infection was 39% and 33% in wells 1 and 2 respectively (i.e. the percentage of HFFs that had an intracellular amastigote).

	Number of HFFs	Number of metacyclic promastigotes	Number of amastigotes	% infection
well 1	206	102	86	39%
well 2	150	279	49	33%

Table 6-2 Number of amastigotes and promastigotes seen in HFF cells. Numbers of parasites and cells were counted under oil immersion at 1000x magnification from 10 different places per well. In well 1 an average of 39% of HFFs had an intracellular amastigote. In well 2, 33% of HFFs had an intracellular amastigote on average.

When HFFs were incubated with *L.major* promastigotes at a parasite:cell ratio of 10:1 for 4h, parasites removed and cells left at 32°C, 5% CO_2 for 48h, the average infection rate was 6.45 (±6.34) (mean±standard deviation). HFFs are typically grown at 37°C, 5% CO_2. To introduce *Leishmania* parasites to this temperature during the infection experiment would heat shock the parasites therefore the confluent HFFs were transferred to a 32°C, 5% CO_2 incubator once the parasites had been added.

Figure 6-3 shows four images (A-D) of non-infected and infected cells. Panels A and B show non-infected HFFs at 200x and 400x magnification respectively. Panels C and D show HFFs infected with *L.major* metacyclic promastigotes (10:1 infection ratio) at 200x and 400x magnification respectively. Round bodies which look similar to the amastigotes seen in the images from the study by Bogdan *et al.* (2000) (Figure 6-2A) can be seen in panel D and are highlighted as amastigotes. Many extracellular promastigotes can be seen in the images.

Figure 6-4 shows two images of infected cells at 400x magnification with the HFFs, amastigotes and promastigotes highlighted.

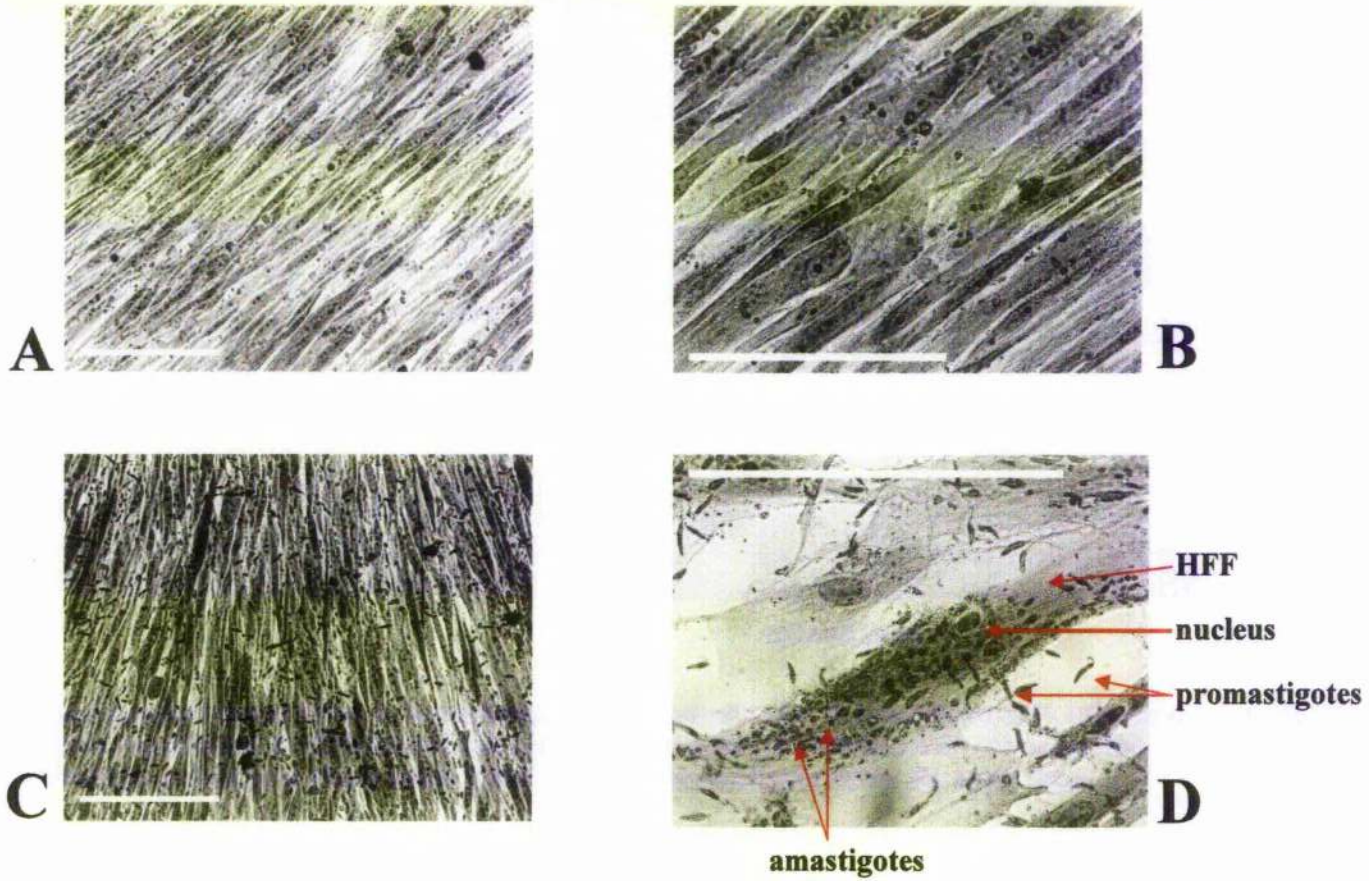


Figure 6-3 Images of non-infected and *L.major* infected HFFs. A. Non-infected HFFs 200x magnification. B. Non-infected HFFs 400x magnification. C. HFFs infected with *L.major* promastigotes (10:1 infection ratio) 200x magnification. D. HFFs infected with *L.major* promastigotes (10:1 infection ratio) 400x magnification. In A and B (non-infected HFFs) cells run parallel to one another. Panel C; extracellular promastigotes clearly visible. Panel D; intracellular amastigotes visible. Bar equals 50μm.

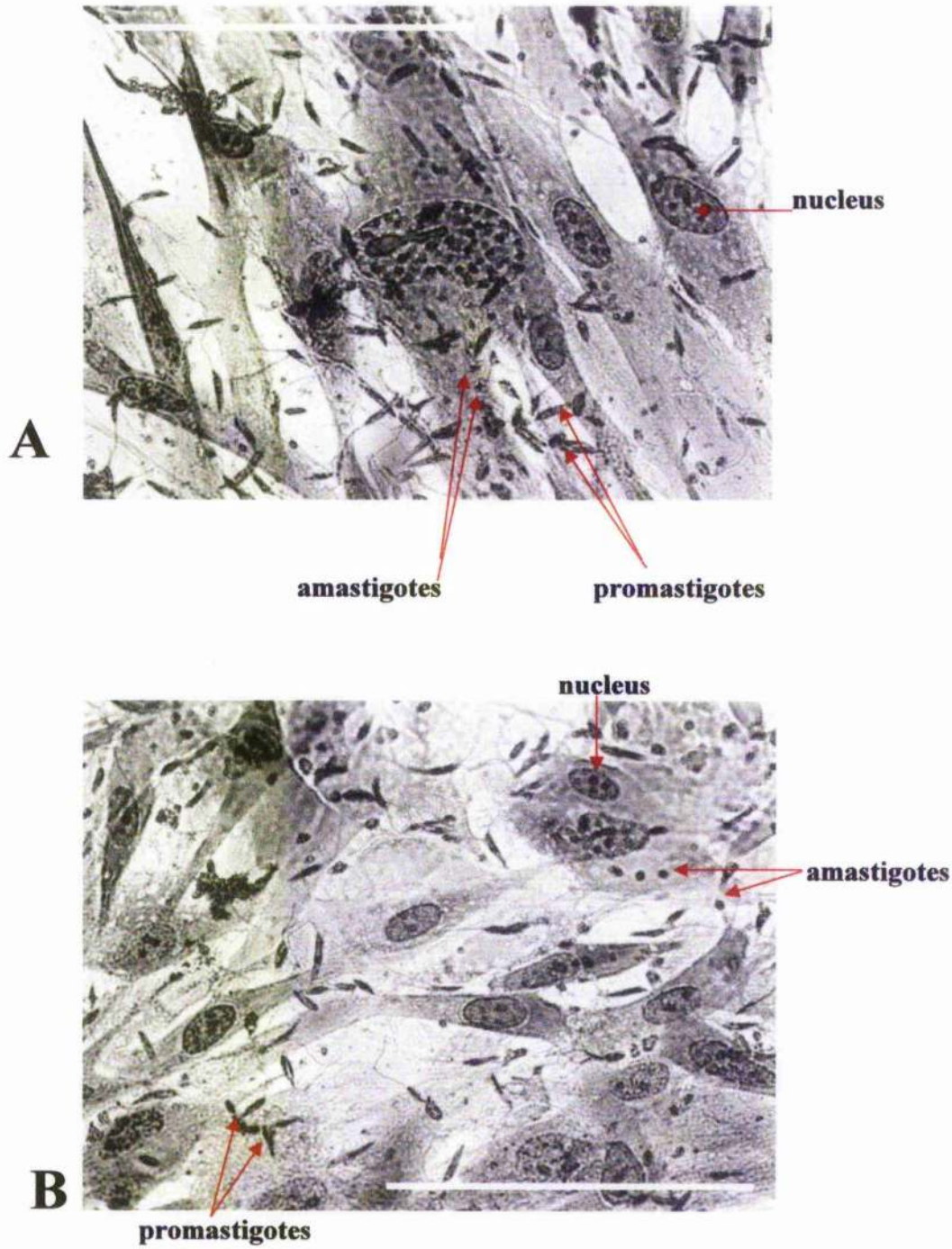
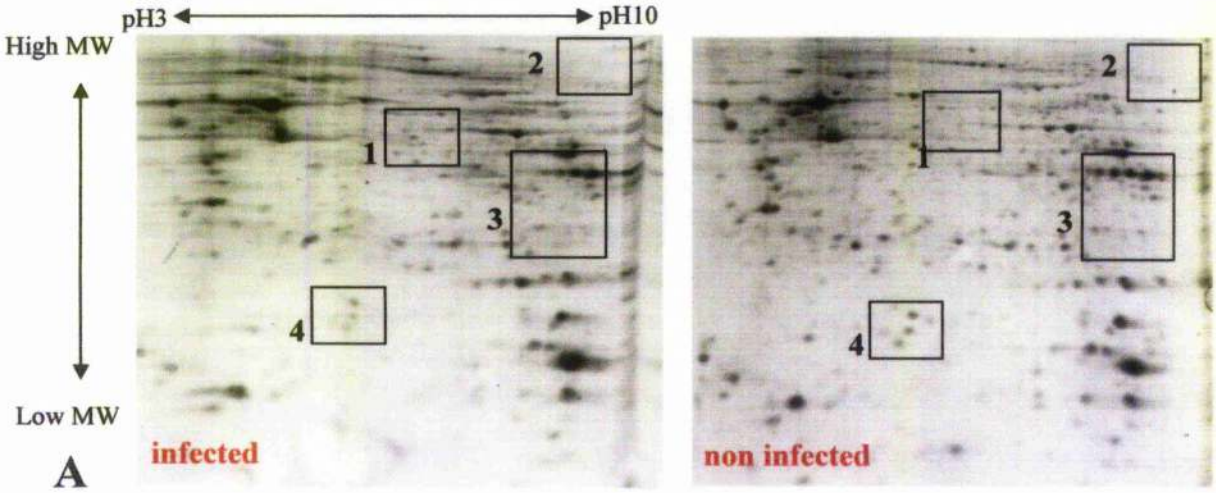


Figure 6-4 HFFs infected with metacyclic promastigotes of *L.major*. 400x magnification. HFF nuclei, amastigotes and promastigotes are highlighted. Many extracellular promastigotes can be seen. Intracellular amastigotes are present within the cells. Bar equals 50μm.

6.3.2 Analysis of L.major infected HFFs by 7cm 2-DE

Leishmania infected HFFs were compared with non-infected HFFs using 2-DE. HFFs were infected with *L.major* metacyclic promastigotes at an infection ratio of 7 parasites: 1 cell. After 4h the cells were washed with fresh DMEM to remove any extracellular parasites and the cells were then left for a further 24h at 32°C, 5% CO_2.

Differentially expressed proteins in *Leishmania* infected cells were first separated using 7cm IPG strips. Forty µg of protein extract from infected or non-infected HFFs was separated on 7cm pH3-10NL IPG strips in the first dimension and then on 12% resolving gels in the second dimension, following the protocols discussed in Chapter II. All gels were stained with colloidal Coomassie. Figure 6-5 shows the colloidal Coomassie stained *L.major* infected HFFs (A) and non-infected cells (B). The infected and non-infected images were compared by eye and also using the 2DElite software. Spots that were differentially expressed between gels C and D can be seen in Figure 6-6. These gels were used for qualitative purposes to show that there were indeed differences in protein expression between *L.major* infected and non-infected cells. It was not possible to identify these differentially expressed proteins as perhaps there was too little protein present per spot to be detected by MS.



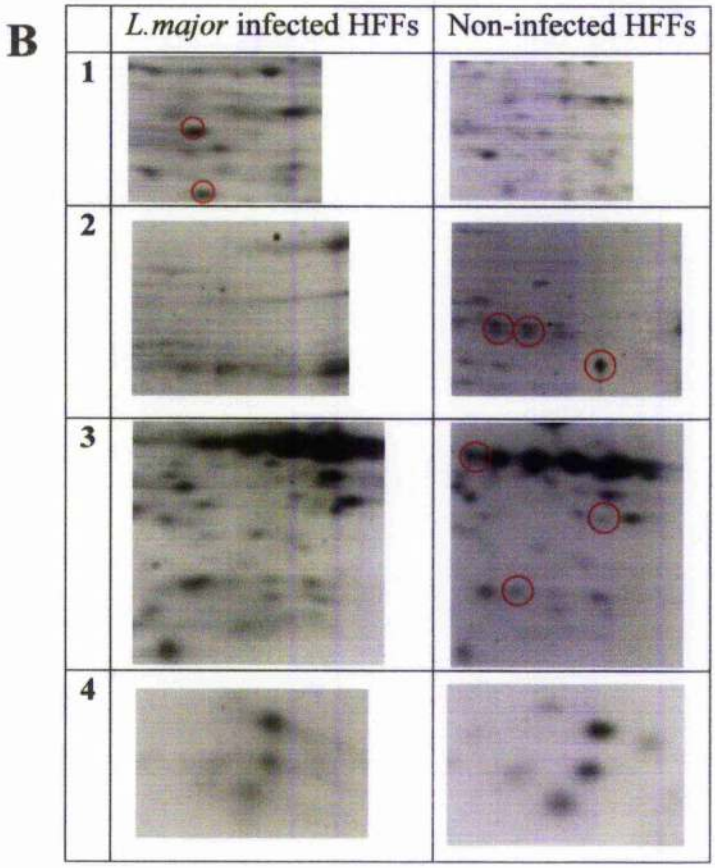


Figure 6-5 A. 7cm mini gels profiling proteins from *L.major* infected and non-infected HFFs. Gels show good resolution of spots and differences can be seen between the infected and non-infected extracts. Gels were stained with colloidal Coomassie. **B** Four sections of the gel magnified to show the differences between the infected and non infected images. Red circles indicate some of the clear differences between the infected and non infected gels.

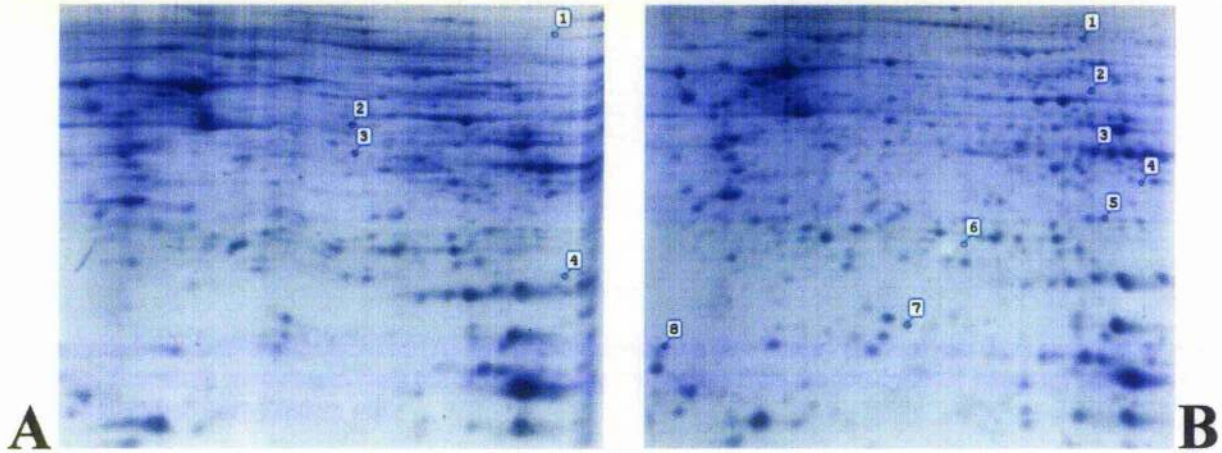


Figure 6-6 Spots chosen for MS analysis from *Leishmania* infected (A) and non-infected cells (B). These 7cm mini gels were analysed by eye to identify any differences between infected and non-infected cells. Four spots were chosen from the infected image and 8 spots from the non-infected image for MS analysis. Spots were excised from the gel manually, digested with trypsin and analysed using Tandem MS.

6.3.3 Analysis of L.major infected HFFs by 24cm 2-DE

Leishmania infected HFFs were then compared with non infected HFFs using 24cm IPG strips. Samples were separated on 24cm pH3-10NL IPG strips following the protocols discussed in Chapter II. An infection ratio of 7 parasites: 1 HFF cell was used. All gels were stained with colloidal Coomassie. The infected and non-infected images were compared by eye and also using the 2DElite software. Spots which were differentially expressed were chosen for MS analysis.

Figure 6-7 and Figure 6-8 show gel images of proteins from *L. major* infected and non-infected HFFs respectively. Differentially expressed spots chosen for MS analysis are shown in Figure 6-9 and Figure 6-10 (infected and non-infected respectively). The 2DElite software was used to determine the volume of the spots that matched across the two gels. Figure 6-11 shows the composite gel image of the infected gel, with the matching spots from the non-infected gel overlaid. Spots shown in green indicate spot volumes that increased in infected cells. Spots shown in red indicate spot volumes that decreased in infected cells. From the *L. major* infected cells, 71 protein spots from the gel were chosen for MS analysis as they were either not present in the same position on the non-infected gel or the volume difference between the matched spots was 30% or greater difference. From the non-infected gel, 55 spots were chosen for MS analysis.

Of the 126 spots picked and analysed by MALDI, it was only possible to identify actin beta twice, once from the infected and once from non-infected gel. This was due to an error with the Ettan Spot handling Work Station. Therefore no conclusions about the protein expression in *L.major* infected cells can be drawn from this experiment.

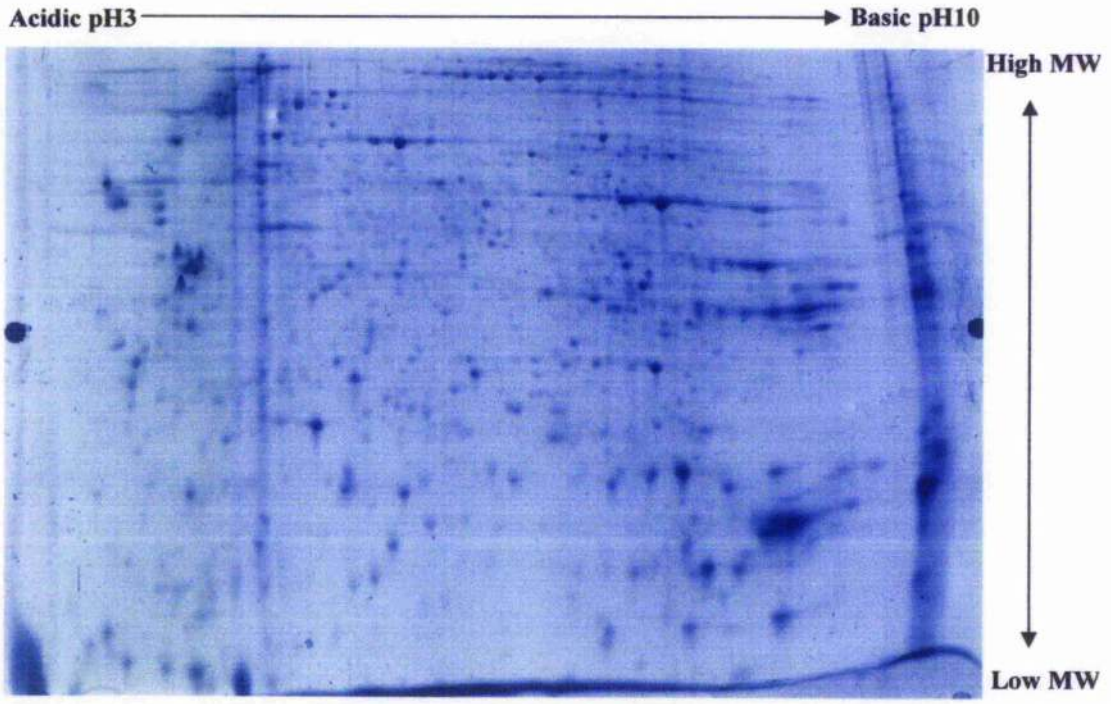


Figure 6-7 Proteome map of *Leishmania major* infected HFFs. Colloidal Coomassie stained gel of whole cell extract from infected HFFs after 28h infection. 251µg of protein was separated by 2-DE on pH3-10NL. Good separation of proteins from acidic to basic pH and high to low molecular weight.

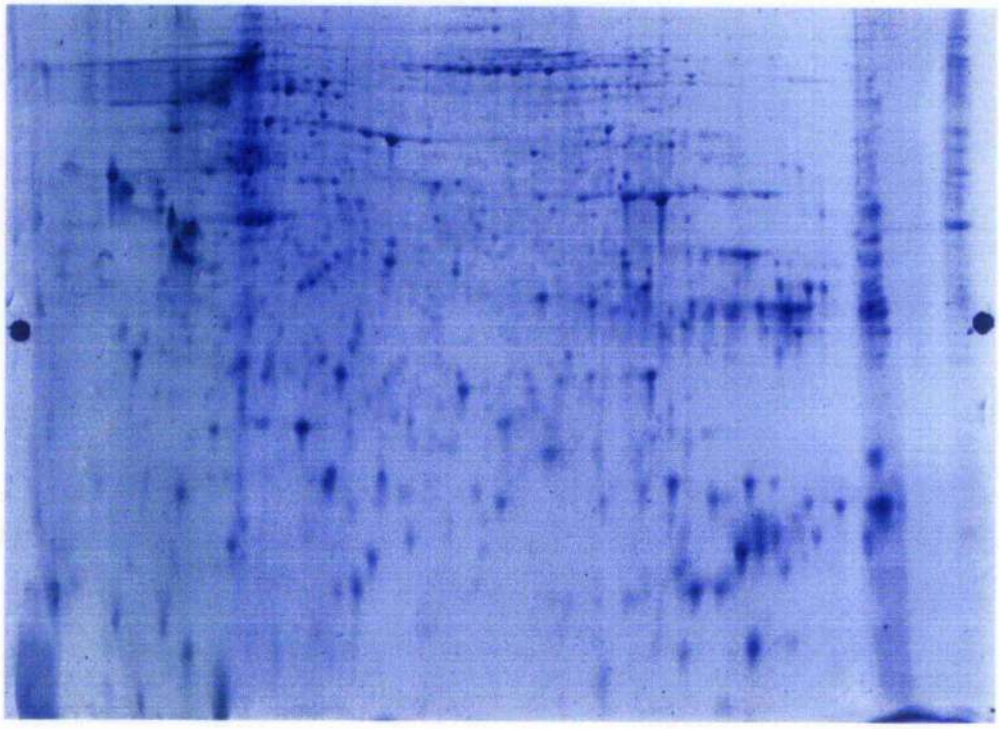


Figure 6-8 Proteome map of non infected HFFs. Colloidal Coomassie stained gel of whole cell extract from non-infected HFFs. 251µg of protein was separated by 2-DE on pH3-10NL. Good separation of proteins from acidic to basic pH and high to low molecular weight.

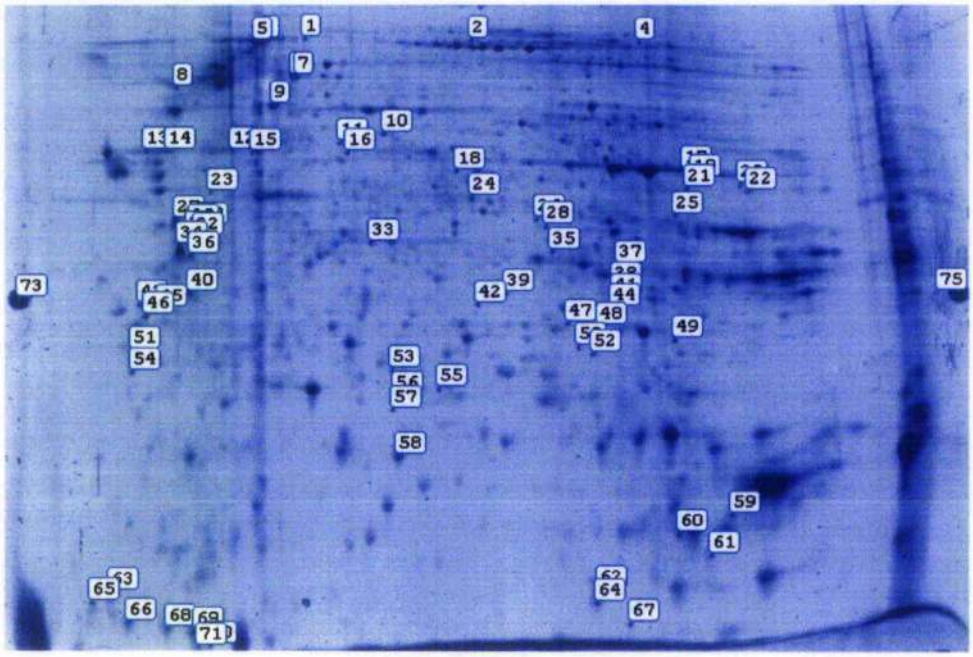


Figure 6-9 2-DE gel of *L.major* infected HFFs. All spots on the gel were detected using the 2DElite software. The numbers refer to protein spots present on this gel that are unique to *L.major* infected cells (e.g. the spots do not appear in the same position on the non-infected gel) or the protein spots are up regulated in volume 30% or greater compared to the matched spot on the non-infected gel. Dust and marks on the gel detected as spots by the software were removed manually. The volume of each spot was calculated relative to background and normalisation.

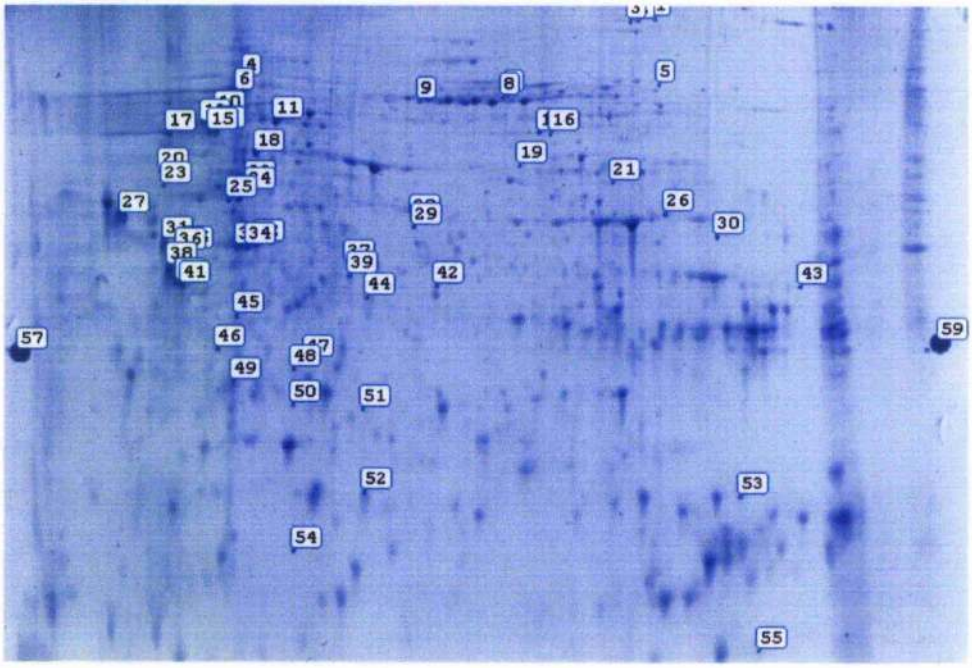


Figure 6-10 2-DE gel of non-infected HFFs. All spots on the gel have been detected using the 2DElite software. The numbers refer to protein spots present on this gel that are unique to non-infected cells (e.g. the spots do not appear in the same position on the infected gel), or the protein spots are down regulated in volume 30% or greater compared to the matched spot on the infected gel. Dust and marks on the gel detected as spots by the software were removed manually. The volume of each spot was calculated relative to background and normalisation.

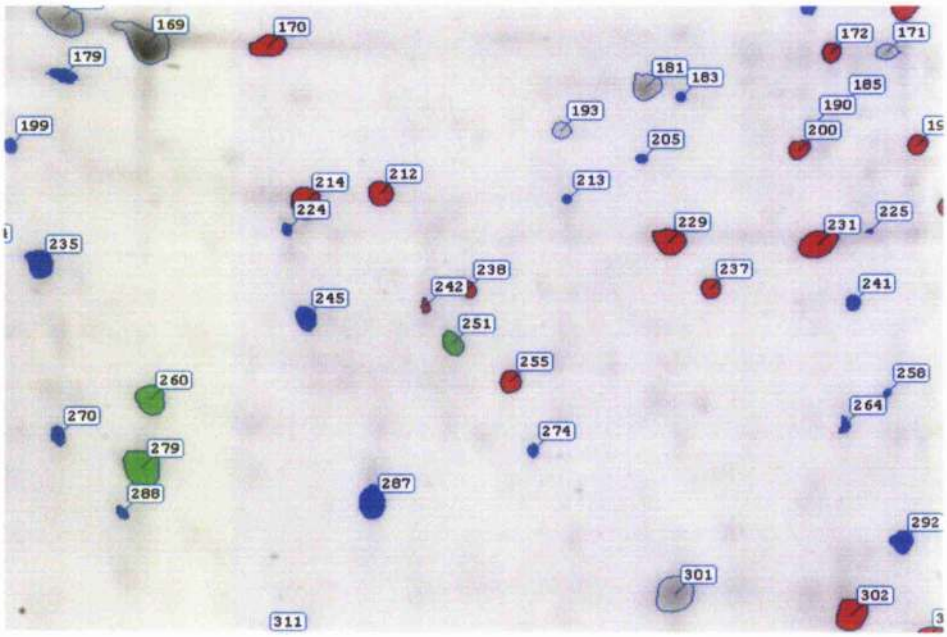


Figure 6-11 Image of the *L.major* infected gel with the matching spots from the non-infected gel overlaid on top. Green indicates spot volume is increased in infected cells, red indicates spot volume is decreased in infected cells. Gels were analysed using the 2DElite software.

6.3.4 Analysis of L.major infected HFFs using DIGE

A second 2-DE technique was then used to analyse the *L.major* infected proteome; DIGE. Figure 6-12 shows the CyDye gel images of the infected, non infected and standard from the experiment. Although the spot resolution on these gels was not optimal, the software was still able to detect several thousand spots. The *Leishmania* infected HFFs were compared with the non-infected HFFs in the DIA module to examine which protein spots increased or decreased in volume during infection. Spots which changed in volume 2-fold or greater were deemed significant. Using this parameter, 34 spots were down regulated and 19 spots up regulated in the parasite infected HFFs. A 2-fold or greater volume ratio was used in the *L.major* DIGE experiment since there was only one gel and hence no way to account for experimental variation. The DIA spot map profiling the 53 spots of interest was exported into the BVA module and 26 of the spots of interest were matched to the SYPRO stained preparative gel.

Figure 6-13 shows a screenshot from the DIA analysis module. The non-infected gel image is upper left with the infected gel image to the right. The 3D view shows a spot from the non-infected image, matched to a spot from the infected image on its right. This spot has a volume ratio difference of -2.64, indicating that the expression of this spot was down regulated in infected cells.

Figure 6-14 shows a screenshot of the BVA module looking at the same spot as in the DIA module. Here the image on the left is again from the non-infected gel image, the image on the right is the matching spot on the preparative gel. Of the 26 spots of interest matched to the preparative gel, 13 were down regulated in infected cells and 13 were up regulated. Of these 26 it was possible to identify 8 of the spots, which represented 9 human protein isoforms using MS/MS. All MS/MS data was searched against the human NCBI database and the locally installed *L.major* PEP database. The MASCOT® results pages for two of the identified proteins can be seen in Figure 6-15 and Figure 6-16. Figure 6-15 shows the MASCOT® results page of beta actin which is down regulated in the infected cell. Figure 6-16 is a MASCOT® results page of a *Leishmania major* identified protein. The results can be seen in Table 6-3. One *L.major* was identified. Down regulated proteins were identified as HSP27, tubulinβ, actinβ, BiP and eukaryotic translation initiation factor 5A. Globinβ was identified from two different spots on the gel, once as up regulated and once as down regulated in the infected cell. This protein is a component of haemoglobin therefore we would not expect it to be present in HFFs.

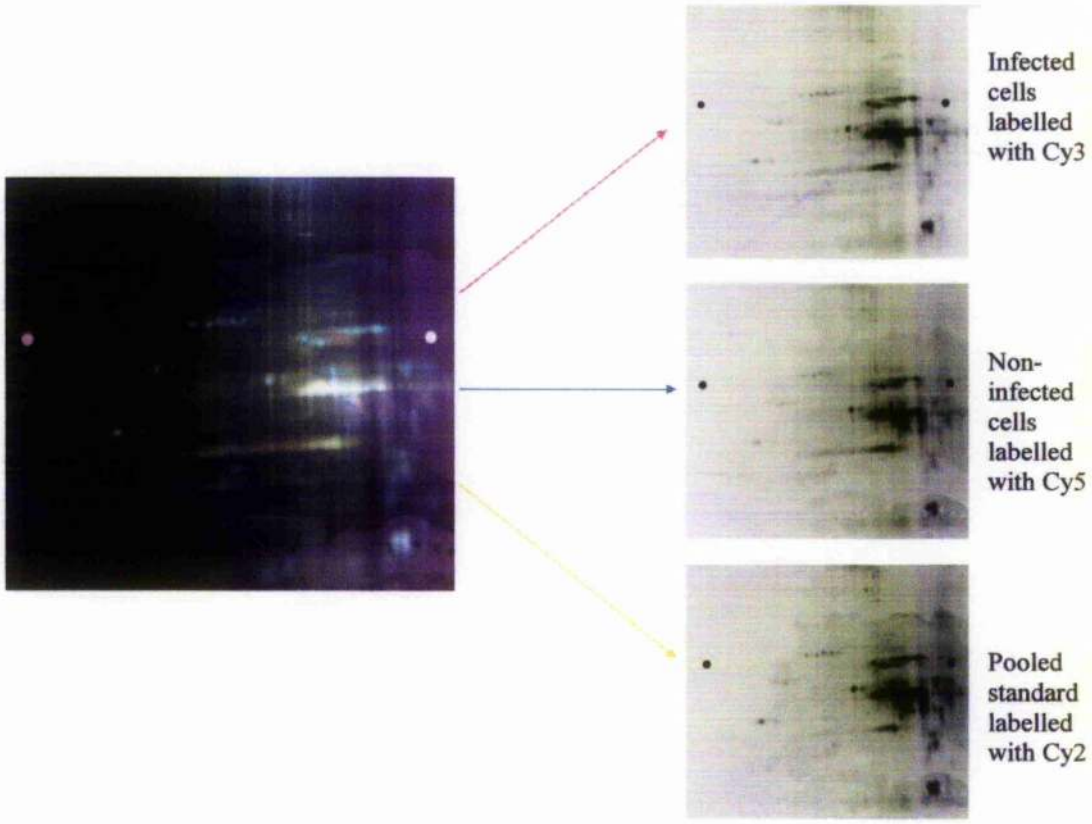


Figure 6-12 Differential protein expression analysis of *Leishmania major* infected and non-infected HFFs by DIGE. Protein extracts (50μg) from *Leishmania major* infected and non-infected HFFs were differentially labelled with Cy3 and Cy5 respectively. A mixed internal standard representing proteins from these two samples in this 24h infection experiment was labelled with Cy2. Proteins were separated on pH4-7 IPG strips in the 1st dimension. The colour image on the left is a composite picture of the 3 CyDye labelled samples overlaid. The three black and white images on the right are the Cy3, Cy5 and Cy2 labelled images separately. Several thousand spots were resolved on each image and with quite good resolution.

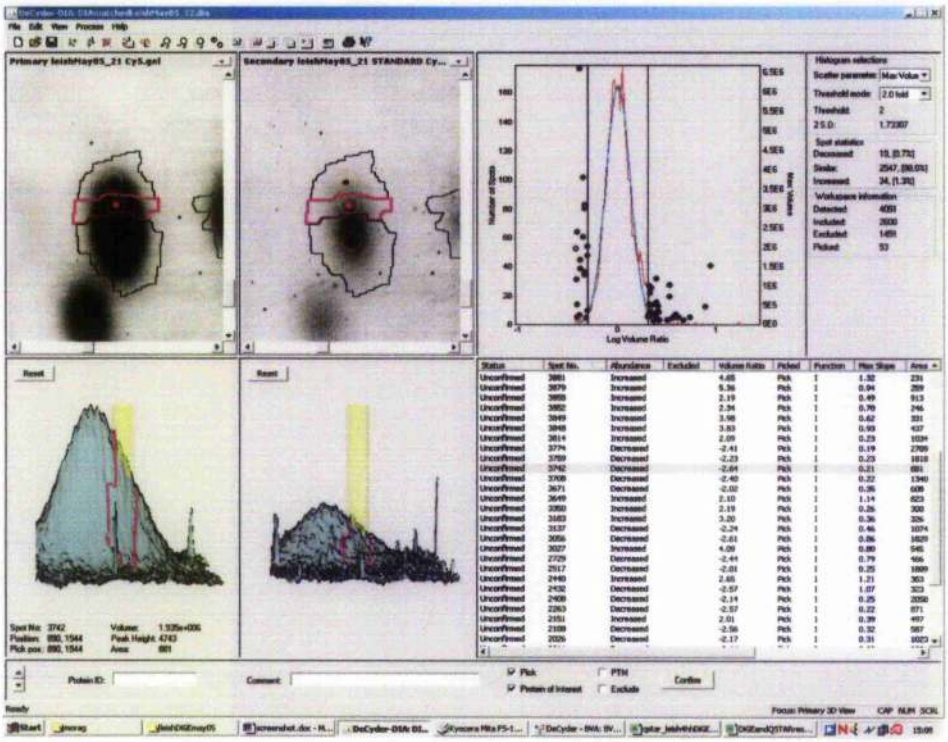


Figure 6-13 Screenshot of the DIA module comparing infected and non-infected gel images. A spot of interest has been highlighted and is shown on the gel and in 3D. The yellow cylinder indicates the protein has been marked 'of interest' and has been assigned pick status.

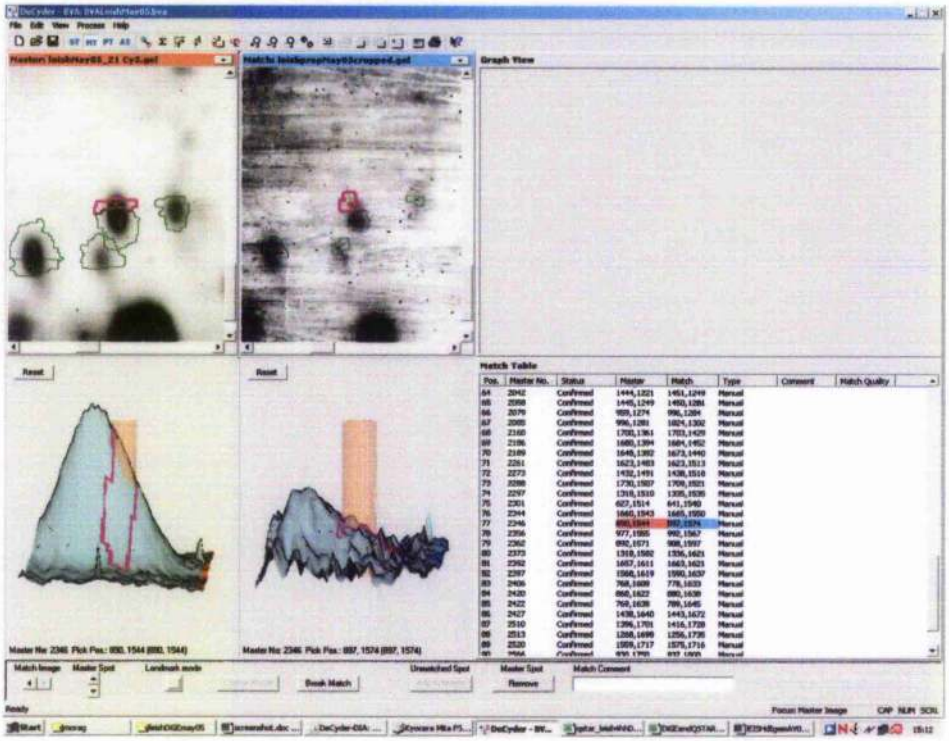


Figure 6-14 Screenshot of the BVA module comparing the non-infected image with the preparative gel. The spot from the Cy5 labelled image is matched to the corresponding spot on the preparative gel from which it will be picked and subjected to MS analysis.

Figure 6-17 is a 3D view of four differentially expressed proteins of interest from *L.major* infected and non-infected HFFs. The top images are from the Cy5 labelled non-infected cells and the lower images are from the Cy3 labelled infected cells. The proteins were identified using MS/MS and the gene names are indicated. ACTB = actin beta, HBB = beta globin, EIF5A = eukaryotic initiation factor 5A isoform I variant CD. These four proteins were all identified as down regulated in *L.major* infected cells.

1	2	3	4	5	6	7	8	9	10	11	12	13
UP	801	Ms/MS	*L.major*	41	LmjF14.0790	445070	6.18	1%	8	possible kine	http://fun-g	2.1
DOWN	2729	MS/MS	H	64	Heat shock 27kDa	22826	5.98	13%	2	gi|54696638	http://fun-g	-2
DOWN	3137	MS/MS	H	58	Beta globin	4604	4.7	30%	1	gi|687685	http://fun-g	-2
UP	3183	MS/MS	H	46	Beta globin	4604	4.7	30%	1	gi|687685	http://fun-g	3.2
DOWN	3671	MS/MS	H 1 of 2	292	Tubulin beta	50240	4.75	14%	9	gi|338695	http://fun-g	-2
DOWN	3671	MS/MS	H 2 of 2	173	BiP	72185	5.03	10%	7	gi|1143492	http://fun-g	-2
DOWN	3742	MS/MS	H	162	ACTB protein	40536	5.55	13%	5	gi|15277503	http://fun-g	-3
DOWN	3759	MS/MS	H 1 of 2	212	eukaryotic initiation factor 5A	17049	5.08	46%	6	gi|33383460	http://fun-g	-2
DOWN	3759	MS/MS	H 2 of 2	151	ACTB protein	40536	5.55	13%	5	gi|15277503	http://fun-g	-2
DOWN	3774	MS/MS	H	199	ACTB protein	40536	5.55	17%	5	gi|15277503	http://fun-g	-2

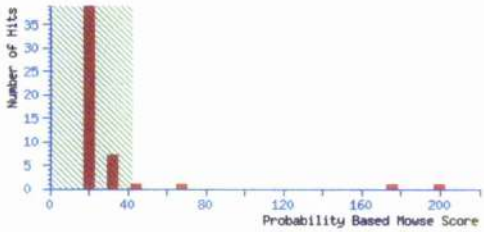
Table 6-3 Proteins identified as differentially expressed in *L.major* infected cells. Column 1: protein is up or down regulated, 2: DIA module number, 3: type of MS analysis used, 4: human or *Leishmania* protein identified, 5: MASCOT score, 6: protein name, 7: molecular weight, 8: pI, 9: sequence coverage, 10: number peptides matched, 11: gene id, 12: MASCOT URL, 13: volume ratio from DIA.

Mascot Search Results

```
User              : morag_250405_ncbi_PP
Email             : 9606340n@student.gla.ac.uk
Search title      : Sample19
MS data file      : C:\morag\Morag18-05-05.wiff
Database          : NCBInr 20050507 (2452561 sequences; 830617002 residues)
Taxonomy          : Homo sapiens (human) (106343 sequences)
Timestamp         : 19 May 2005 at 16:59:40 GMT
Significant hits: gi|28336     mutant beta-actin (beta'-actin) [Homo sapiens]
                  gi|9506371   actin, alpha 1, skeletal muscle [Rattus norvegicus]
                  gi|12408252  FKSG30 [Homo sapiens]
                  gi|181402    epidermal cytokeratin 2 [Homo sapiens]
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 42 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As: Peptide Summary — Help

Significance threshold p< 0.05 — Max. number of hits: AUTO

Standard scoring / MudPIT scoring — Ions score cut-off: 0 — Show sub-sets

Show pop-ups / Suppress pop-ups — Sort unassigned: Decreasing Score — Require bold red

Select All | Select None | Search Selected | Error tolerant | Archive Report

1. gi|28336 **Mass:** 42128 **Score:** 199 **Queries matched:** 5
mutant beta-actin (beta'-actin) [Homo sapiens]

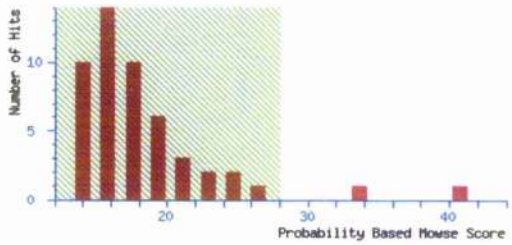
Figure 6-15 MASCOT® results page identifying mutant actin beta as down regulated in *L.major* infected cells.

Mascot Search Results

```
User            : moragLeishPEP
Email           : 9606340n@student.gla.ac.uk
Search title    : Sample02
MS data file    : C:\morag\Morag18-05-05.wiff
Database        : LmajorPEP 20040903 (8216 sequences; 5220500 residues)
Taxonomy        : Homo sapiens (human) (8216 sequences)
Timestamp       : 20 May 2005 at 08:18:28 GMT
Warning         : Taxonomy 'Homo sapiens (human)' ignored. No taxonomy indexes for this database
Significant hits: LmjF14.0790  |||hypothetical protein, unknown function|Leishmania major|chr 14|||Manual
                  LmjF16.1220  |||hypothetical protein, conserved|Leishmania major|chr 16|||Manual
```

Probability Based Mowse Score

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 28 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

Format As: Peptide Summary — Help

Significance threshold p< 0.05 — Max. number of hits AUTO

Standard scoring / MudPIT scoring — Ions score cut-off 0 — Show sub-sets

Show pop-ups / Suppress pop-ups — Sort unassigned: Decreasing Score — Require bold red

Select All | Select None | Search Selected | Error tolerant | Archive Report

1. LmjF14.0790 **Mass:** 445070 **Score:** 41 **Queries matched:** 8
   |||hypothetical protein, unknown function|Leishmania major|chr 14|||Manual
   Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
20	421.79	841.56	842.42	-0.86	0	(10)	4.9	2	GDVGVGSPR
21	421.79	841.57	842.42	-0.86	0	(8)	8.4	3	GDVGVGSPR
☑ 22	421.79	841.57	842.42	-0.86	0	(17)	0.95	1	GDVGVGSPR
☑ 23	421.79	841.57	842.42	-0.86	0	17	0.94	1	GDVGVGSPR
26	421.80	841.58	842.42	-0.85	0	(10)	4.9	3	GDVGVGSPR
121	522.32	1563.95	1563.72	0.23	1	0	31	10	AMRLTSDAAQGAGCR
☑ 144	747.10	2238.28	2239.07	-0.79	1	(15)	0.82	1	TSGHANSMVASNSRLCQLVSK + Oxidation (M)
☑ 145	560.58	2238.30	2239.07	-0.77	1	26	0.073	1	TSGHANSMVASNSRLCQLVSK + Oxidation (M)

Figure 6-16 MASCOT® results page identifying a *Leishmania* protein as present on the *L.major* infected gel.

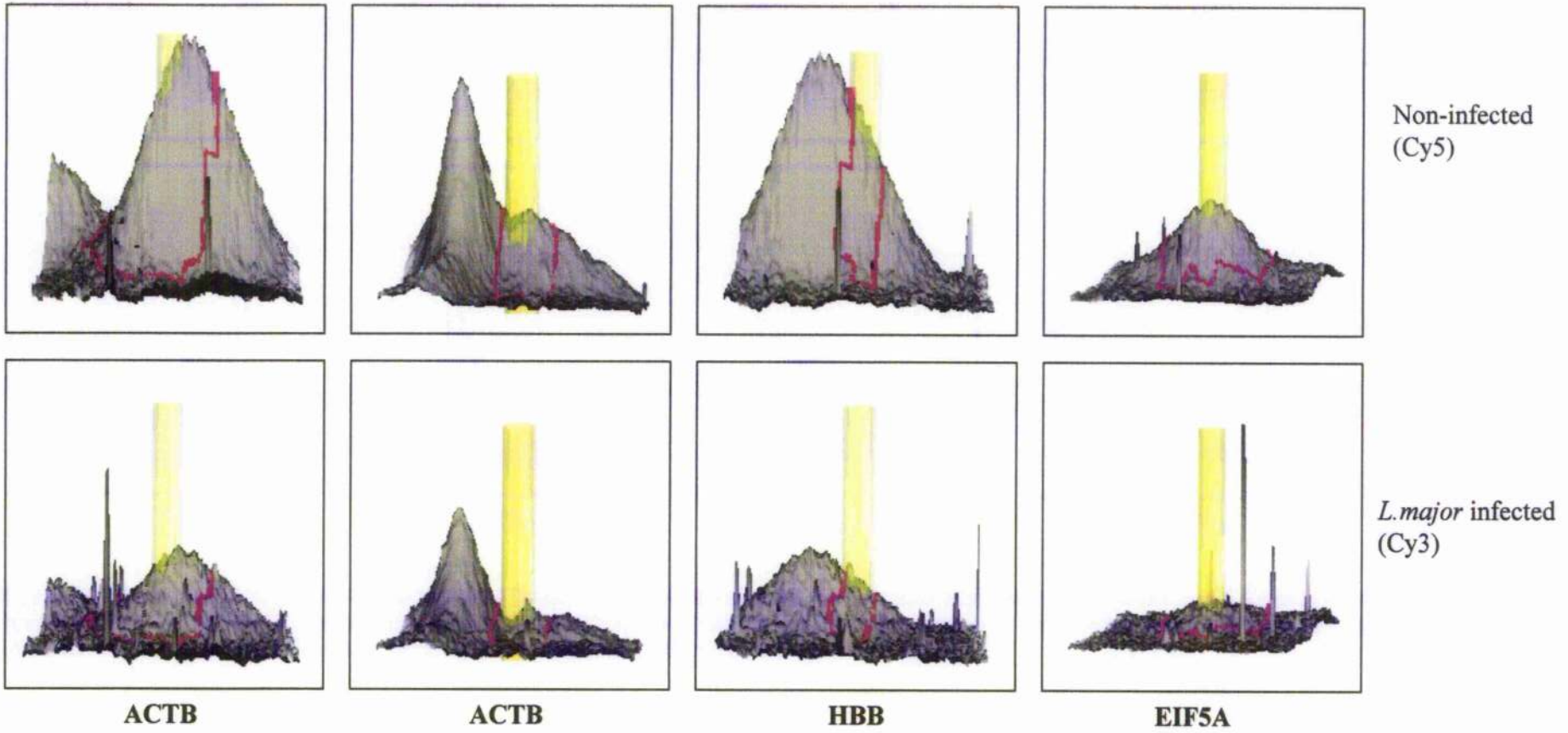


Figure 6-17 Example of four proteins down regulated following *L.major* infection. 3D view of differentially expressed proteins from *L.major* infected and non-infected HFFs. The top images are from the Cy5 labelled non-infected cells and the lower images are from the Cy3 labelled infected cells. The proteins were identified using MS/MS and the gene names are indicated. ACTB = actin beta, HBB = beta globin, EIF5A = eukaryotic initiation factor 5A isoform I variant CD.

6.3.5 Analysis of L.mexicana infected HFFs using DIGE

The *Leishmania mexicana* infected HFF cell proteome was compared with an HFF non infected proteome using DIGE. The premise behind this experiment was that it would be possible to compare the proteome of HFFs when infected by two different strains of the same parasite and look for similarities/differences in the host response. The HFFs were infected with *L.mexicana* promastigotes for 48h after which time the infected and non infected cells were harvested and the sample prepared for DIGE. The DIGE gel images can be seen in Figure 6-18. Analysis of the corresponding infected and non infected images using the DIA software highlighted many differences between the samples of interest, in one experiment, 81 spots were down regulated and 61 up regulated in infected cells by 2-fold or greater. However, it was not possible to identify any of these regulated spots due to difficulties in protein transfer to the second dimension on the spot gel, and also technical difficulties with the Amersham Work Station. Due to time limitations it was not possible to repeat this experiment. Thirty four spots were down regulated and 19 up regulated by 2-fold or more in the *L.major* experiments as compared to the down regulation of 81 spots and up regulation of 61 spots by 2-fold or greater in *L.mexicana* infected cells. Perhaps these differences relate to differences between the parasite strains. Further work needs to be done before any firm conclusions can be drawn from this observation.

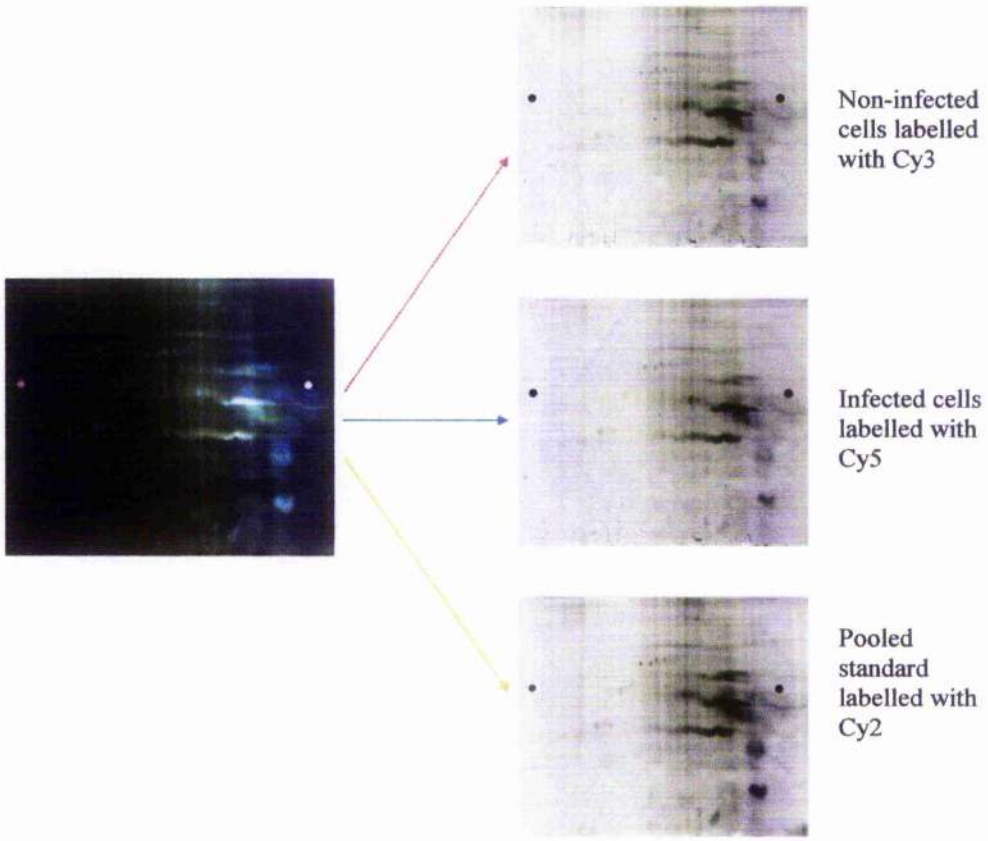


Figure 6-18 Differential protein expression analysis of *Leishmania mexicana* infected and non-infected HFFs by DIGE. Protein extracts (50µg) from *Leishmania mexicana* infected and non-infected HFFs were differentially labelled with Cy5 and Cy3 respectively. A mixed internal standard representing proteins from these two samples in this 24h infection experiment was labelled with Cy2. Proteins were separated on pH4-7 IPG strips in the 1st dimension. The colour image on the left is a composite picture of the 3 CyDye labelled samples overlaid. The three black and white images on the right are the Cy5, Cy3 and Cy2 labelled images separately. These gels do not show good spot resolution; there may have been a problem with the focusing or transfer of proteins to the second dimension.

6.3.6 *Comparison of protein expression changes in Leishmania and Toxoplasma infected cells*

Modulated proteins common to both the *Toxoplasma* and *Leishmania* DIGE studies were compared. As relatively few proteins were identified in the *Leishmania* study it was only possible to match four proteins. Two of the proteins behaved similarly (consistent down regulation of ACTB and HSP27) and two behaved differently (BiP and tubulin beta). Due to the very small data set it is not possible to infer any commonalties in the host cell response when infected with these different parasites.

Protein name	Expression in *L.major* infected cells	Expression in *Toxoplasma* infected cells	Correlates?
ACTB	down	down	yes
BiP	down	up	no
HSP27	down	down	yes
Tubulin beta	down	up	no

Table 6-4 Comparison of the protein changes seen in *L.major* infected HFFs with *Toxoplasma gondii* infected HFFs. Due to the small comparative sample size, it is difficult to draw any conclusions about whether the protein expression is similar or not in the two sets of pathogen infected cells.

6.4 Discussion

To investigate whether the host cell has a common response to intracellular pathogens, the protein expression of *Toxoplasma gondii* infected cells was compared to the protein expression of *Leishmania major* and *Leishmania mexicana* infected cells. To allow for close comparison of *Toxoplasma gondii* and *Leishmania sp* infected cells, the host cell HFF was used. The infection ratios and time points were kept as similar to the microarray and proteomic experiments to allow easy and valid comparisons of data between *Leishmania* and *Toxoplasma* infected cells. Our results show that HFFs can be a host cell for *L. major* and *L. mexicana* parasites. Analysis of *Leishmania* infected cells by 2-DE and DIGE showed that there are modulations in the protein expression in the cell due to the presence of the parasite. Our results indicate that in comparison to *Toxoplasma gondii* infected cells, *Leishmania major* or *Leishmania mexicana* infected cells showed much fewer protein expression differences. In the conventional gel to gel experiment, 71 spots were up regulated and 55 spots were down regulated in *L.major* infected cells. This is in comparison to 165 spots up regulated and 144 spots down regulated in the *Toxoplasma gondii* infected cells in 24h experiment replicate 11 (Chapter III). In the *Leishmania* DIGE experiments, 19 spots were up regulated and 34 spots down regulated two fold or more, this is compared to 384 spots down regulated in volume and 54 increased in volume on just one of the *Toxoplasma* DIGE gels. From this it might be possible to conclude tentatively that there is a difference in response of HFFs to these parasites, at least with respect to the numbers of proteins modulated. However, disappointingly due to technical problems with MS identification, too few proteins were identified to draw firm conclusions as to differences in the nature of the response to the two parasites.

HSP27 was identified as down regulated in the *L.major* infected cell. This contradicts the results of a study by Kantengwa *et al* (1995) which showed that macrophages infected with *L.major* had no effect on host stress proteins synthesis which may relate to the inability of *L.major* to activate the cells respiratory burst (Kantengwa *et al.*, 1995).

The differences in the protein expression of the pathogen infected cells most likely reflects the different lifestyles of the two parasites, including the different mechanisms of entry, survival strategies and the different rates of replication. *Leishmania major* promastigotes bind to specific receptors on the macrophage surface, inducing phagocytosis and the parasite is then internalised by the phagocyte into a phagosome. Macrophage lysosomes fuse with the phagosome to produce a phagolysosome within which *Leishmania* double

every 12-24h (R. Burchmore, personal communication). In contrast, *Toxoplasma gondii* binds to specific receptors on the host cell surface and actively invades the host cell. Within the cell the parasite resides within a parasitophorous vacuole (PV) composed entirely of parasite proteins. *Toxoplasma* replicates by binary fission with a replication time of 6-9h (Coppens and Joiner, 2001).

A comparison was carried out to see if there was any relationship between the gene expression of *T.cruzi*-infected cells (Blader *et al.*, 2001) and the protein expression of *Leishmania* infected cells. In both cases much fewer cellular changes were seen in the *Leishmania* infected cells than observed in *T.cruzi* infected cells. None of the genes modulated in the *T.cruzi*-infected cells were the same as the proteins modulated in *Leishmania* infected cells. An explanation for this could again be attributed to the parasites lifestyles. *Trypanosoma cruzi* invades the host cell by attachment of the motile *T.cruzi* trypomastigotes to the host cell surface and the slow, active process of internalisation takes around 5-10min (de Avalos *et al.*, 2002). During parasite internalisation a PV is formed, as a result of targeted fusion of host cell lysosomes with the plasma membrane and establishment of infection takes about 24h. This includes intracellular invasion, exit from the vacuole, and differentiation to amastigotes. Amastigote replication begins at 24-30hpi and parasite doubling occurs every 12h. In the study by de Avalos (2002) it was noted that independent of the level of *T. cruzi* infection of HFFs, less than 25% of the parasites had escaped from the PV by 6hpi, whereas more than 90% of the intracellular parasites were present in the cytoplasm as amastigotes at 24hpi. The three very different methods of internalisation, intracellular location and survival mechanisms, and times for re-establishment of infection and replication rates will of course play a part in the different gene and protein expression responses seen in the host cell response to infection with these different pathogens.

HFFs are not a typical host cell for *Leishmania* but have been described as such in latent Leishmaniasis (Bogdan *et al*, 2000). The study showed that when skin fibroblasts were incubated with *L.major* promastigotes at a parasite to cell ratio of 3-5:1, the average infection rate at 48h was 17.8(±9) and 24.7(±12.4) % in unstimulated or cytokine stimulated cells respectively (mean± standard deviation). The authors stated that most of the infected cells contained one or two parasites, rarely, three parasites were found within one cell. The presence of intracellular amastigotes in the fibroblasts was confirmed by immunoperoxidase staining of intracellular parasites and transmission electron microscopy. We counted intracellular amastigotes in HFFs using Geimsa staining and light microscopy. It would have been useful to carry out immunolocalisation studies using

antibodies specific to amastigotes for the unambiguous identification of the presence of intracellular amastigotes in the HFFs. Studying the images of the HFFs incubated with the *L.major* promastigotes and in reference to the study by Bogdan *et al* (2001) it is possible to conclude that HFFs can be a host cell for *Leishmania* parasites. It is most likely that the HFFs act as reservoir cells for the parasite and that the parasite does not replicate within the cells. When HFFs were incubated with *L.major* promastigotes at a parasite: cell ratio of 10:1 for 48h, the average infection rate was 6.45 (±6.34) (mean±standard deviation).

Using proteomics, it was possible to investigate the host response to *Leishmania* and to compare this response to *Toxoplasma gondii* infected cells. However, this system is not perfect as it may not be fully representative of what is happening in the natural biological situation. As the HFFs are a mortal cell line and can only be used in the lab up to passage number 20, the cells should not be adapted by the cell culture process, and remain representative of fibroblasts found in the body. The *Leishmania* parasites again were all less than serial passage number 20 so should not have adapted due to the culture conditions. Blader *et al* (2001) investigated the host response of *Toxoplasma gondii* infected cells using microarrays and compared the transcriptional profiles to those of *Trypanosoma cruzi* infected cells. They found very little correlation between the gene expression of the *T.gondii* and *T.cruzi* infected cells. The genes induced in *T.cruzi* infected cells were classified into functional categories. There was a notable absence of several functional groups of genes which had previously been shown to be modulated in *T.gondii* infected HFFs, including apoptosis, carbohydrate metabolism, cell cycle and cytokines (de Avalos, 2002). Proteins in these categories were also not identified in the *Leishmania* proteomic studies. Comparing the gene expression results from *T.cruzi* infected cells with the *T.gondii* and *L.major* infected cell protein responses shows that the host cell responses are different for each intracellular pathogen. This is a tentative conclusion with respect to the *Leishmania* host protein response as perhaps not enough proteins were identified from the infected cells to allow a genuine comparison.

Although very little can be concluded about the protein expression of *Leishmania* infected cells, the fact that much fewer changes were seen in these pathogen infected cells as compared to *Toxoplasma gondii* infected cells shows that the host response is specific to individual pathogens and was not just a stress response of the cell to the experimental conditions. The most important conclusion to be taken from the work in this chapter is that HFF cells do not produce an arbitrary protein expression response. The protein profile seen in Chapters III and IV in *Toxoplasma gondii* infected cells can therefore be defined as specific to the presence of the parasite.

Chapter VII Analysis of the phosphoproteome of *Toxoplasma gondii* infected cells

7.1 Introduction

Protein phosphorylation is one of the most common post translational modifications (PTM) found in the cell (Mann and Jensen, 2003). Phosphorylation is responsible for the regulation of many cellular events, including transcription, translation, cell growth, division and differentiation, metabolism, cytoskeleton dynamics, cell movement, apoptosis, signalling events and memory (Stensballe *et al.*, 2001; Raggiaschi *et al.*, 2005). Therefore the study of protein phosphorylation in the parasite infected cell will give a more in depth understanding of the host cell response to infection rather than just looking at the steady state proteome. Changes in protein abundance may not reflect protein activity in the cell due to PTM therefore inferring protein function purely on the basis of changes in abundance may provide a very limited view of the proteome (Guerrera and Kleiner, 2005). Krebs and Fisher first described protein phosphorylation in 1955 and showed it was essential for the conversion of glycogen phosphorylase from its inactive to active form (Raggiaschi *et al.*, 2005).

It is estimated that approximately one-third of the 10,000 or so protein species in a typical mammalian cell are phosphorylated at any given time, many with more than one phosphate (Alberts et al., 1999). Protein phosphorylation involves the enzyme-catalysed transfer of the terminal phosphate group of an ATP molecule to the hydroxyl group on a serine, threonine or tyrosine side chain of the protein. This reaction is catalysed by a protein kinase. The reverse reaction, dephosphorylation, is catalysed by a protein phosphatase as seen in Figure 7-1. There are some 500 kinases and 100 phosphatases encoded in the human genome accounting for approximately 2% of the total number of genes. This high number implies an important role for the regulation of protein phosphorylation (Mann *et al.*, 2002).

Protein phosphorylation can affect the protein in two ways; (1) the addition of a double negatively charged phosphate group to a protein can cause a major conformational change in the protein which can then affect the binding of ligands elsewhere on the protein surface and change the activity of the protein through an allosteric effect. De-phosphorylation returns the protein to its original confirmation and restores its original activity. (2) An

attached phosphate group can form part of a structure that is directly recognised by binding sites of other proteins. Protein phosphorylation and dephosphorylation have a major role in the formation and disassembly of protein complexes in the cell.

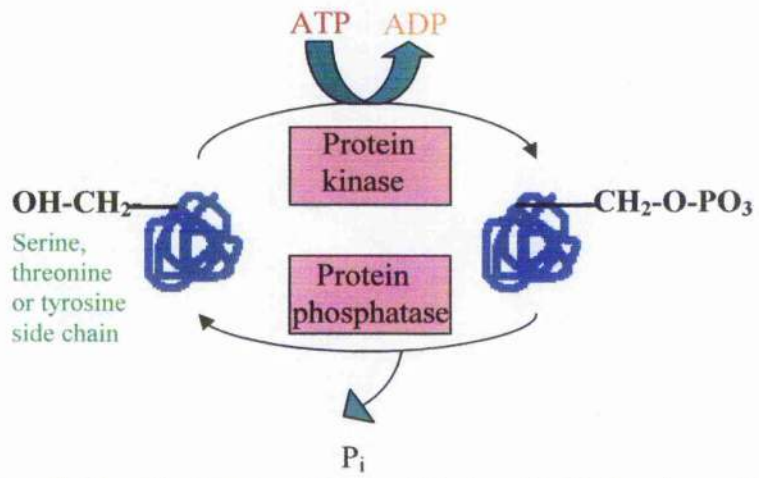


Figure 7-1 Protein phosphorylation and dephosphorylation. A protein kinase enzyme transfers a phosphate group (PO_3) from ATP to an amino acid side chain (serine, threonine or tyrosine) of the protein. The protein can then be dephosphorylated by removal of the phosphate group by a protein phosphatase.

Studying the cellular phosphoproteome using proteomics allows for the potential analysis of all phosphoproteins in the cell at the one time. Several techniques are now available to study the phosphoproteome, for example; MS analysis, radioactive labelling with P^{32}, western blotting employing phosphospecific antibodies, purification of phosphorylated proteins using affinity columns, phosphoprotein specific stains for use with 1-DE and 2-DE gels and treatment of proteins with a phosphatase followed by MS.

Addition of a phosphate group results in an increase in mass of 80Da which can be detected by MS. However, analysis of phosphopeptides by MS is not straight forward. For both MALDI and MS/MS signals from phosphopeptides themselves are usually suppressed because of the abundance of non-phosphorylated species. PMF by MALDI does not yield direct sequence information therefore it is not possible to identify the exact phosphorylated residue. Purified protein or peptides are usually required for analysis of phosphorylation by MS/MS which involves prior purification steps by one or two-dimensional electrophoretic gels or by high performance liquid chromatography. Phosphopeptides are negatively charged and due to the fact that electrospray ionisation is generally performed in the positive mode, ionisation of phosphopeptides can be weak. Furthermore phosphopeptides tend to be hydrophilic molecules and therefore do not bind well to the peptide purification columns often used before analysis (Stensballe *et al.*, 2001; Mann, 2002). It is possible to

identify phosphopeptides using a triple quadrupole MS in the negative ion mode. On fragmentation by collision-induced dissociation (CID) in a MS/MS, phosphopeptides not only produce sequence-specific fragments but also fragments that are specific for phosphate groups. These phosphate-specific fragment ions serve as characteristic reporter ions for phosphorylated peptides in precursor ion scanning experiments by MS/MS. Peptides carrying a phosphate group can therefore be easily identified by precursor-ion scanning because of the loss of phosphate (PO_3^-) under alkaline conditions (Mann, 2002). MALDI-TOF MS in combination with phosphatase treatment has previously been used to identify phosphopeptides. Phosphopeptides were identified based on a characteristic mass shift owing to loss of phosphate (80Da or multiples of) after treatment with phosphatase (Mann, 2002).

Despite these techniques, analysis of phosphoproteins is still quite difficult for several reasons; (1) many of the proteins which require phosphorylation for activation (e.g. those involved in signalling) are present at low copy numbers in the cells therefore enrichment of the proteins is required, (2) often only a small percentage of proteins are phosphorylated at a given time point, (3) phosphatase enzymes must be inhibited during sample preparation to inhibit dephosphorylation and (4) each protein can often exist in several different phosphorylated forms (Mann *et al.*, 2002).

The aim of this chapter was to profile the phosphoproteome of pathogen infected cells to determine the protein phosphorylation status due to the presence of an intracellular parasite. Often the fact that a protein is phosphorylated indicates that it is activated in the cell (Mann *et al.*, 2002). Knowledge of the phosphorylation status of the proteins in the pathogen infected cell will lead to a better understanding of which proteins are involved and activated in the host pathogen response. Phosphorylation events are closely linked with cell signalling and apoptosis. Host cell apoptosis is known to be inhibited by the presence of *Toxoplasma gondii* within the cell, therefore identification of phosphorylated proteins may identify proteins involved in modulating the apoptotic cascade or anti-apoptosis pathways. The infected cell undergoes both suppression and activation of transcription, which is controlled by signalling events. Phosphorylation is also known to play a key role in the immune system (Raggiaschi *et al.*, 2005) therefore it would be expected to play a part in the infected host cell response.

Three techniques were employed. The first involved purification of the subset of phosphorylated proteins in *Toxoplasma gondii* infected and non-infected cells using the PhosphoProtein Purification Kit (QIAGEN). This technique involves passing the cell

lysate through a column that selectively binds phosphorylated proteins. The unphosphorylated flow through is collected and the phosphorylated fraction eluted from the column. The phosphorylated and unphosphorylated fractions of infected and non-infected cells can then be compared by 1-DE and 2-DE.

The second method used Pro-Q® Diamond phosphoprotein gel stain (Molecular Probes) which selectively stains phosphorylated proteins in 1-DE and 2-DE gels. The Pro-Q® Diamond phosphoprotein gel stain can detect as little as 1-16ng of phosphoprotein per band, depending on the phosphorylation state of the protein.

The final technique compared the phosphorylated and dephosphorylated proteins in *T.gondii* infected and non-infected cells using DIGE. Proteins were dephosphorylated using Shrimp Alkaline Phosphatase (SAP) (Promega) and then separated and compared using the DIGE protocols discussed in Chapter IV. This technique using SAP has previously been used to dephosphorylate proteins for use with 1-DE gels (Jonas and Privalsky, 2004) and was also used in a study by Stensballe (2001) where calf alkaline phosphatase was used to dephosphorylate peptides prior to MS analysis, resulting in unambiguous identification of phosphopeptides by observation of 80Da mass shifts. Raggiaschi *et al* (2005) used λ-phosphatase in concert with the gel-to-gel technique to compare the phosphorylated and non-phosphorylated extracts from rat cortical neurons.

7.2 Materials and Methods

7.2.1 Time course infection experiments

To investigate the phosphoproteome of *Toxoplasma gondii* infected cells several different time points post invasion were analysed to determine the degree of phosphorylation during invasion. Time points post invasion investigated were 12h, 16h and 24h. Cells were harvested as previously described in Chapter II. The only exception to the normal protocol was that the cells were washed in FS HEPES/Saline solution and not PBS. Gel bands or gel spots chosen for MS analysis were analysed by the MS methods outlined in Chapters II and III.

7.2.2 Analysis of the phosphoproteome using the QIAGEN PhosphoProtein Purification Kit

To purify phosphoproteins from *Toxoplasma gondii* infected HFFs (16h infection, 10:1 infection ratio) and non-infected HFFs, the PhosphoProtein Purification Kit protocol from QIAGEN was followed. The starting material was $1x10^7$ cells. 875µl CHAPS stock solution (10%w.v) was added to 35ml of PhosphoProtein Lysis Buffer and 75µl CHAPS Stock Solution (10%w/v) to 3ml of PhosphoProtein Elution Buffer to yield a final concentration of 0.25%(w/v) CHAPS in both buffers. One protease inhibitor tablet and 10µl benzonase stock solution were added to a 5ml aliquot of the PhosphoProtein Lysis buffer containing CHAPS in which the cell pellet was resuspended and then incubated for 30min at 4°C. During this time the sample was vortexed briefly every 10min. After 30min incubation, the cell lysate was centrifuged at 10,000 g, 4°C for 30min. During centrifugation, a PhosphoProtein Purification Column was prepared by detaching the top cap, breaking off the bottom closure, and allowing the storage buffer to flow out. Application of 4ml PhosphoProtein Lysis Buffer containing 0.25% CHAPS equilibrated the column and the buffer was allowed to flow out. The cell supernatant was harvested and the protein concentration determined using the Bradford method. The protein concentration of the lysate was adjusted to 0.1mg/ml by adding PhosphoProtein Lysis Buffer containing 0.25% CHAPS, yielding a final volume of 25ml of lysate. The lysate was poured into the upper reservoir of the PhosphoProtein Purification column which allowed the phosphorylated protein to bind to the affinity column. The flow through fraction was collected for analysis of unphosphorylated proteins in the lysate. PhosphoProtein Lysis Buffer (6ml) containing 0.25% CHAPS was added to wash the column through.

PhosphoProtein Elution Buffer (500µl) containing 0.25% CHAPS was then added to the column and the eluted fraction collected. This elution step was repeated four times. The protein concentration in all eluted fractions was determined using the Bradford method. Nanosep ultrafiltration columns were used to concentrate the eluted protein fractions. The protein fraction was placed into the sample reservoir of the Nanosep Ultrafiltration column and centrifuged at 10,000 g for 10min giving a final volume of 50µl.

The unphosphorylated and phosphorylated fractions were then compared using 1-DE and 2-DE methods as detailed in Chapter II. The molecular weight ladder used in all of the 1-DE experiments was a broad range protein molecular weight marker from Promega. For the 2-DE, the samples were concentrated and desalted using Nanosep ultrafiltration columns. 50µl of the phosphorylated or unphosphorylated fraction were placed in the upper chamber of the column and 450µl of 10mM TrisCl, pH7.0 added and the volume reduced to 50µl by centrifugation. This was repeated five times, each time adding another 450µl of 10mM TrisCl, pH7.0 to the 50µl sample. These washes removed any contaminants that would be incompatible with IEF. Gels were stained with colloidal Coomassie.

7.2.3 Pro-Q® Diamond Phosphoprotein Stain

To visualise the phosphoproteome of *Toxoplasma gondii* infected cells and non-infected HFFs, 1-DE and 2-DE gels were stained with Pro-Q Diamond phosphoprotein stain. The stain was purchased from Invitrogen (formerly Molecular Probes).

Cells infected for 12h (5:1 infection ratio) and 16h (10:1 infection ratio) and non-infected cells were used in the ProQ® Diamond stain experiments. Sample preparation using LB3 was carried out as detailed in Chapter II. Any salts or lipids that would not be compatible with the Pro-Q stain were removed from the cell samples by a methanol/chloroform precipitation. To a 150µl sample (containing ~150-300µg protein), 600µl of methanol (MeOH) were added and the sample vortexed. 150µl of chloroform were added and the sample vortexed. 450µl of ddH_2O were added and sample vortexed and then centrifuged at 12,000 g for 5min. The upper phase was discarded and the white precipitation disc that formed between the upper and lower phases kept. 450µl of MeOH was added, the sample vortexed and then centrifuged at 12,000 g for 5min. The supernatant was discarded and the pellet dried in a vacuum centrifuge for 10min or left to air dry. The pellet was then resuspended in LB3 ready for electrophoresis.

2-DE was carried out using the 2-DE protocols discussed in Chapter II. The same quantity of protein was loaded onto the IEF strip as would typically be used for a colloidal Coomassie dye-stained gel. On completion of the second dimension separation the gels were stained with the Pro-Q® Diamond Phosphoprotein gel stain according to the manufacturer's protocol. The following volumes correspond to the volumes required to stain a 7cm mini-gel. In brief; the gel was immersed in 100ml of fix solution (10% trichloroacetic acid/50% MeOH) and incubated at r.t for at least 30min with gentle agitation. This step was repeated once more to ensure all SDS was washed out of the gel. The gel was washed 3x in 100ml of ddH_2O with gentle agitation for 10min each. The gel was then incubated in the dark in 50ml of Pro-Q® Diamond phosphoprotein gel stain with gentle agitation for 75-120min. The gel was destained by incubating in the dark in 80ml of destain solution (Molecular Probes propriety destain solution) at r.t for a total of 3h with two changes of destain solution. The gel was imaged using the Typhoon scanner. Pro-Q® diamond phosphostain has an excitation maximum at approximately 555nm and an emission maximum at approximately 580nm.

7.2.4 Phosphatase treatment

To analyse the phosphoproteome of *Toxoplasma gondii* infected cells, shrimp alkaline phosphatase (SAP) was used to dephosphorylate any phosphorylated proteins and the treated and mock treated samples from both infected and non-infected cells were compared using DIGE. SAP was purchased from Promega (catalogue # M820A). Protein lysates were prepared from 24h infected (10:1 infection ratio) and non-infected cells as previously described for DIGE in Chapter IV. Cells were washed in FS HEPES/Saline solution during harvesting and not PBS as this may have interfered with the natural phosphorylation status of the proteins. The DIGE lysis buffer contains 8M urea which is too high a concentration of urea for SAP activity (Biotec ASA data sheet). Therefore the lysis buffer was diluted four fold to gain a final concentration of 2M urea. The cell lysates (infected and non-infected) were halved and half of each sample was treated with 10units (in 10µl) of SAP to dephosphorylate the phosphorylated proteins. The remainder of each sample was treated with an equal volume of enzyme storage buffer (25mM Tris-HCl, 1mM $MgCl_2$, 0.1mM $ZnCl_2$ and 50% glycerol) to be used as a control.

The samples were then labelled with the appropriate CyDye (Table 7-1) and the DIGE protocol followed (as in Chapter IV) with proteins separated on pH4-7 immobiline drystrips. The addition of the rehydration buffer prior to IEF adjusted the urea concentration back to 8M.

	Cy3	Cy5	Cy2
1	Infected phosphatase treated	Infected mock treated	pooled standard
2	Non-infected phosphatase treated	Non-infected mock treated	pooled standard

Table 7-1 Details of CyDye labelling for each sample in the SAP-DIGE experiment. Both phosphatase treated infected and non-infected samples were labelled with Cy3. Both mock treated infected and non-infected samples were labelled with Cy5. A pooled internal standard labelled with Cy2 was also run on both gels.

The DeCyder DIA module was used to analyse the protein expression differences between the phosphorylated and unphosphorylated samples from *Toxoplasma gondii* infected and non-infected cells. The Cy3 labelled SAP treated infected sample was compared with the Cy5 labelled infected mock treated sample. A two fold or greater difference in protein spot volume was marked as significant. The same analysis was also used to compare the Cy3 labelled SAP treated non-infected cells with the Cy5 labelled non-infected mock treated sample. Lists of the significant differentially expressed spots from both of the comparisons were compiled. These lists were then collated and compared. All of the spots of interest were picked using the Ettan Spot Handling Work Station (Amersham Biosciences) and subjected to MALDI MS analysis. Spots which were not identified using MALDI were then analysed by MS/MS.

7.3 Results

7.3.1 QIAGEN PhosphoProtein Purification Kit

The QIAGEN PhosphoPurification Kit is reported to separate the phosphorylated from the unphosphorylated cellular protein fraction. HFFs infected with *T.gondii* for 16h and non-infected HFFs were separated into phosphorylated and unphosphorylated fractions and compared using 1-DE and 2-DE methods. Proteins that were differentially expressed in the phosphorylated and unphosphorylated extracts between infected and non-infected samples were detected by both 1-DE and 2-DE methods and identified by MS.

Cells were washed in filter sterile HEPES/Saline solution instead of PBS as PBS could have interfered with the phosphorylation status of the proteins and also with the purification column.

Clear differences could be seen between the phosphorylated and unphosphorylated fractions after using the QIAGEN kit. Figure 7-2 shows a 1-DE gel where 5µg of unphosphorylated and phosphorylated non-infected cells and unphosphorylated and phosphorylated infected cells were separated by 1-DE. An extract from total HFF cells was also run (lane 2). The gel was stained with colloidal Coomassie. All of the bands visible in the phosphorylated lanes (5 and 6) should contain only phosphorylated proteins. This figure clearly shows differences between phosphorylated and non-phosphorylated extracts from the same biological sample.

Figure 7-3 is of another 1-DE gel where 5µg of unphosphorylated and phosphorylated extracts from non-infected cells and unphosphorylated and phosphorylated extracts from infected cells were compared and the gel was stained with colloidal Coomassie. This figure shows clearly the differences in the phosphoproteome of *T.gondii* infected cells as compared to non-infected cells.

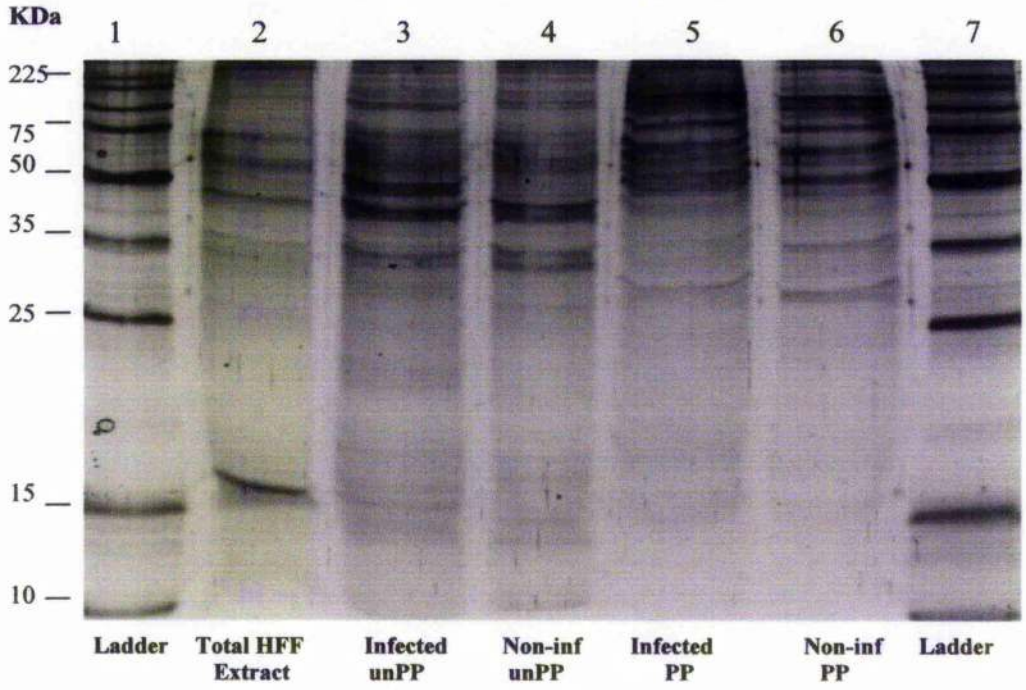


Figure 7-2 1-DE gel comparing the phosphorylated and unphosphorylated proteins from 16h infected and non-infected cells. Phosphorylated and unphosphorylated proteins from *Toxoplasma gondii* infected and non-infected cells were separated using the QIAGEN Phosphoprotein purification kit. Fiveμg of protein were loaded per lane and total HFF extract was run in lane 2. Clear differences can be seen between the phosphorylated and unphosphorylated extracts, the phosphorylated proteins in infected and non-infected samples and in the unphosphorylated proteins in infected and non-infected samples. Molecular weight markers were run in lanes 1 and 7.

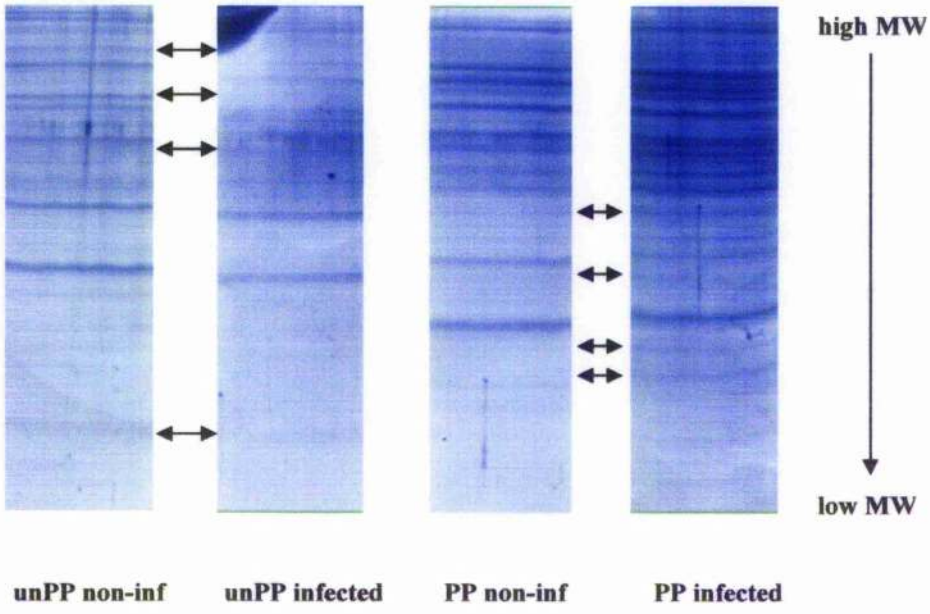


Figure 7-3 1-DE gel comparing the phosphorylated and unphosphorylated protein fractions from 16h infected and non-infected cells. Phosphorylated and unphosphorylated proteins from *Toxoplasma gondii* infected and non-infected cells were separated using the QIAGEN Phosphoprotein purification kit. 5μg of protein were loaded per lane. Clear differences can be seen between the phosphorylated and unphosphorylated extracts, the phosphorylated proteins in infected and non-infected samples and the unphosphorylated proteins in infected and non-infected samples. Differences between extracts are marked by an arrow.

Based on the differences seen in the 1-DE gel, 2-DE was carried out. Fifteen μg of phosphorylated extract from infected and non-infected cells were loaded onto 7cm pH 3-10NL IPG strips. Proteins were separated using 2-DE and the gels were stained with colloidal Coomassie. Figure 7-4 images A and B show the phosphorylated proteins from infected and non-infected cells respectively. Ten spots on each gel that were not present on the opposite image (therefore differentially expressed during infection) were chosen for MS analysis. The twenty gel plugs were digested into their component peptides with trypsin using the Ettan Spot Handling Work Station and analysed by MS/MS. As well as searching the peptide masses against the databases with the fixed carbamidomethyl modification and the variable methionine oxidation modification, the data was also searched against phospho (STY) variable modifications. Images C and D show the annotated gel images with the proteins that were identified using MS.

This experiment was repeated several times. Unfortunately the gels were often of poor quality with bad reproducibility and separation and low numbers of spots resolved. In general, the QIAGEN phosphoprotein purification columns gave very poor results when used in conjunction with 2-DE. It was not possible to perform any large format (24cm) 2-DE gels as the yield of phosphorylated proteins using the QIAGEN kit was very low. One explanation for the poor resolution and yield may be that the eluting buffer from the QIAGEN kit interfered with 2-DE. Cleaning the sample in 10mM TrisCl pH7.0 buffer five times gave no better results. To try to further purify the sample and remove any contaminants, the 2-DE clean-up kit (Amersham Biosciences) was used, but this did not yield any better results.

The results from the QIAGEN phosphoprotein experiment can be seen in Table 7-2. The peptides were analysed by MS/MS and searched against the NCBI and ToxoDB databases. One of the proteins identified as up regulated from the phosphorylated fraction was vimentin. The MS analysis showed that one of the peptides from vimentin was phosphorylated, as can be seen in Table 7-3. Figure 7-5 shows the ion series of the phosphorylated peptide. Neither the b or y ion series are complete and the peptide ion score is quite low.

The eight individual human proteins identified by MS from these gels all have the potential to be phosphorylated. The up regulated proteins included caldesmon, GNAT2, tropomyosin 1 and vimentin. Of interest is vimentin, one of the most prominent phosphorylated proteins in the cell, and caldesmon, which will be discussed later in the chapter.

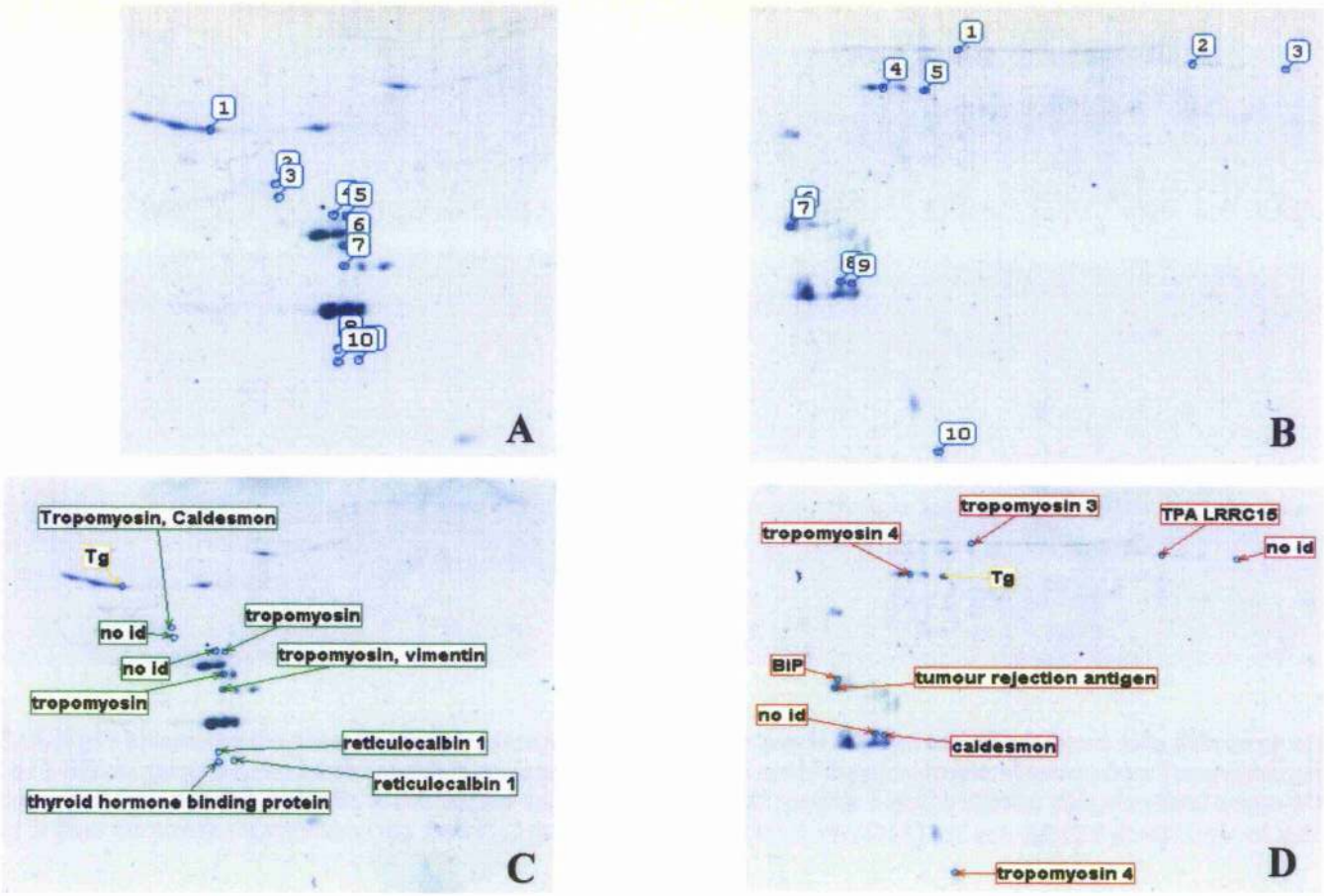


Figure 7-4 2-DE mini gels comparing the phosphorylated extracts from *Toxoplasma gondii* infected and non-infected cells. Fifteen μg of sample was loaded onto 7cm pH3-10NL IPG strips, separated by 2-DE and the gels stained with colloidal Coomassie. Images A and C are of the phosphorylated extract from *Toxoplasma gondii* infected cells. Images B and D are of the phosphorylated extract from non-infected cells. A and B show the spots chosen for MS analysis from the infected phosphorylated extract (A) and from the non-infected phosphorylated extract (B). C and D show the proteins identified by MS from the infected cell phosphorylated extract (A) and non-infected phosphorylated extract (B) respectively.

Sample no.	Protein name	Can be pp?	No of hits	Score	MASCOT	Gene name	Organism	Mass	pI	s.c	P.M	gI number
I1	TgTwinScan_2537		1 of 1	83	http://fun-	Tubulin beta	T.g	50461	4.69	2%	7	
I2	TPM4	YES	1 of 3	216	http://fun-	TPM4	H	28619	4.77	16%	5	gi|10441386
I2	CALD1 protein	YES	2 of 3	170	http://fun-	CALD1	H	21078	5.01	14%	3	gi|38303917
I2	Guanine nucleotide-binding protein 2	YES	3 of 3	46	http://fun-	GNAT2	H	67654	5.54	5%	3	gi|2135319
I5	Tropomyosin 1 alpha chain	YES	1 of 1	495	http://fun-	TPM1	H	32746	4.69	28%	13	gi|136092
I6	Tropomyosin 1	YES	1 of 1	199	http://fun-	TPM1	H	34980	4.81	20%	6	gi|7441391
I7	Tropomyosin 1	YES	1 of 2	83	http://fun-	TPM1	H	34980	4.81	9%	3	gi|7441391
I7	Vimentin	YES	2 of 2	67	http://fun-	VIM	H	53710	5.06	5%	3	gi|37852
N4	Tropomyosin 4	?	1 of 2	396	http://fun-	TPM4	H	28619	4.67	31%	10	gi|54696136
N4	ALL-1 protein	?	2 of 2	49	http://fun-	MLL	H	440237	9.21	1%	5	gi|1490271
N5	TgTigrScan_1642		1 of 1	520	http://fun-	HSP70 T.g	T.g	73379	5.23	13%	24	
N6	BiP	?	1 of 1	729	http://fun-	HSPA5	H	72185	5.03	34%	19	gi|1143492
N7	Tumor rejection antigen (gp96) 1	YES	1 of 1	366	http://fun-	TRA1	H	92696	4.76	12%	9	gi|44890631

Table 7-2 The results table from the QIAGEN phosphoprotein 2-DE experiment. Proteins which had a phosphate group present are seen in blue. *Toxoplasma gondii* proteins are in pink. Column 3 indicates whether the protein can be phosphorylated *in vivo*. All of the proteins identified in this experiment should be able to be phosphorylated as they were prefractionated using the QIAGEN kit into phosphorylated extracts. Table contains information on sample identification, protein name, whether the protein can be phosphorylated, the number of significant MASCOT hits per gel spot, MOWSE score, MASCOT url, gene name, organism, mass, pI, sequence coverage, peptides matched and gI number.

gi|37852	Mass: 53710	Score: 67		Queries matched: 3		
Vimentin						
Query	**Observed**	**Mr(expt)**	**Mr(calc)**	**Score**	**Expect**	**Rank**	**Peptide**
147	**547.29**	**1092.56**	**1092.52**	**57**	**0.0015**	**1**	**FADLSEAANR**
259	**552.37**	**1654.09**	**1654.71**	**(6)**	**2.1e+02**	**8**	**TYSLGSALRPSTSR + 2 Phospho (ST)**
269	**579.37**	**1735.09**	**1734.68**	**13**	**47**	**1**	**TYSLGSALRPSTSR + 3 Phospho (ST)**

Table 7-3 Identification of a phosphorylated protein by MS/MS. Vimentin was identified as one of the differentially expressed spots in a 16h infection experiment. The peptides were analysed by MS/MS and searched against the NCBI and ToxoDB databases. As well as searching the peptide masses against the databases with the fixed carbamidomethyl modification and the variable oxidation (M) modification, the data was also searched against Phospho (STY) variable modifications. The score for the phosphorylated peptides (query 259 and 269) is very low. The overall MOWSE score is 67. The significance cut-off was a score of greater than 42 (p-value<0.05, 95% confidence interval).

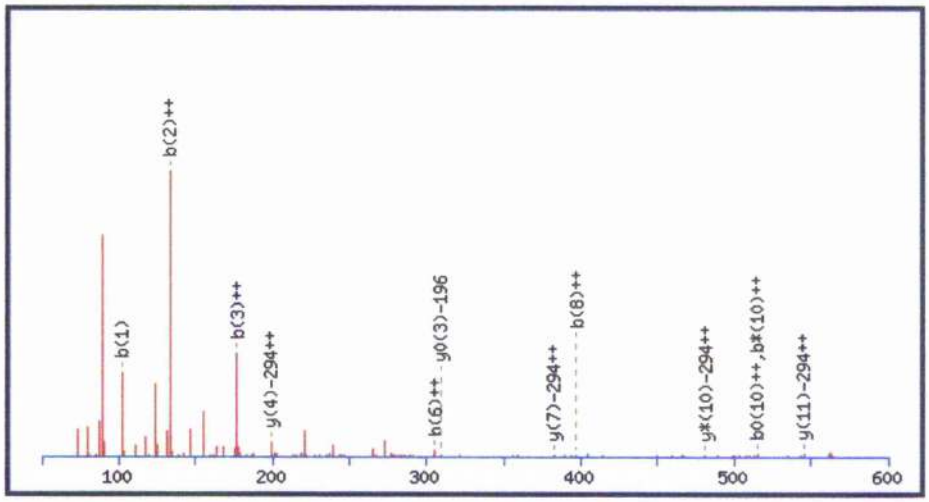
{MATRIX SCIENCE} Mascot Search Results

Peptide View

MS/MS Fragmentation of TYSLGSALRPSTSR
Found in gi|37852, vimentin [Homo sapiens]

Match to Query 269: 1735.092166 from(579.371331,3+)
Elution from: 25.58 to 25.58 period: 0 experiment: 1 cycles: 1
From data file C:\morag\Morag1389.wiff

Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 50 to 600 Da



Monoisotopic mass of neutral peptide Mr(calc): 1734.68
Fixed modifications: Carbamidomethyl (C)
Variable modifications:
S11 : Phospho (ST), with neutral loss 97.98
T12 : Phospho (ST), with neutral loss 97.98
S13 : Phospho (ST), with neutral loss 97.98
Ions Score: 13 **Expect:** 47
Matches (Bold Red): 12/138 fragment ions using 24 most intense peaks

#	b	b^{++}	b^*	b^{*++}	b^0	b^{0++}	Seq.	y	y^{++}	y^*	y^{*++}	y^0	y^{0++}	#
1	**102.05**	51.53			84.04	42.53	T							14
2	265.12	**133.06**			247.11	124.06	Y	1340.71	670.86	1323.68	662.34	1322.70	661.85	13
3	352.15	**176.58**			334.14	167.57	S	1177.64	589.33	1160.62	580.81	1159.63	580.32	12
4	465.23	233.12			447.22	224.12	L	1090.61	**545.81**	1073.59	537.30	1072.60	536.80	11
5	522.26	261.63			504.25	252.63	G	977.53	489.27	960.50	**480.75**	959.52	480.26	10
6	609.29	**305.15**			591.28	296.14	S	920.51	460.76	903.48	452.24	902.50	451.75	9
7	680.32	340.67			662.31	331.66	A	833.47	417.24	816.45	408.73	815.46	408.24	8
8	793.41	**397.21**			775.40	388.20	L	762.44	**381.72**	745.41	373.21	744.43	372.72	7
9	949.51	475.26	932.48	466.75	931.50	466.25	R	649.35	325.18	632.33	316.67	631.34	316.17	6
10	1046.56	523.79	1029.54	**515.27**	1028.55	**514.78**	P	493.25	247.13	476.23	238.62	475.24	238.12	5
11	1115.58	558.30	1098.56	549.78	1097.57	549.29	S	396.20	**198.60**	379.17	190.09	378.19	189.60	4
12	1198.62	599.81	1181.59	591.30	1180.61	590.81	T	327.18	164.09	310.15	155.58	**309.17**	155.09	3
13	1267.64	634.33	1250.62	625.81	1249.63	625.32	S	244.14	122.57	227.11	114.06	226.13	113.57	2
14							R	175.12	88.06	158.09	79.55			1

Figure 7-5 MASCOT results page of the phosphorylated peptide in query 269. This peptide was identified as containing a phospho residue. The b and y ion series can be seen and are not very complete. This is not a very confident identification of a phosphopeptide.

7.3.2 Pro-Q® Diamond phosphoprotein gel stain

Proteins from *Toxoplasma gondii* infected and non-infected cells were separated using 2-DE and the gels stained with ProQ® Diamond phosphoprotein gel stain which is reported to selectively stain phosphorylated proteins. Figure 7-6 shows an image of a ProQ stained mini-gel gel profiling a non-infected HFF extract. As can be seen, the protein resolution on the gel is very poor. Very few proteins were detected by this method of staining.

Several attempts to stain the phosphorylated proteins in 2-DE mini gels using the ProQ phosphoprotein diamond stain resulted in blank images, i.e. no proteins detected. As approximately one third of the protein species in the human proteome is thought be phosphorylated (Alberts *et al.*, 1999), it would be expected that many phosphorylated proteins would be detected using the stain. To confirm there were in fact proteins present on the mini gels, the gels were fixed and then stained in colloidal Coomassie. Figure 7-7 A and B shows images of two colloidal Coomassie stained mini gels (infected and non-infected extracts respectively) which when stained with the ProQ® Diamond stain, according to the manufacturers instructions, were blank, but when stained with colloidal Coomassie (total protein stain) proteins could be seen on the gel.

In case mini gels were not providing large enough quantities of protein per spot for the stain to be most effective, a 24cm 2-DE experiment was carried out with fresh stain and destain supplied by Molecular Probes. The 12h *Toxoplasma gondii* infected and non-infected HFF cell extracts (320µg) were run on 24cm pH3-10NL IPG strips. The gels were stained according to the ProQ® Diamond protocol and imaged using the Typhoon scanner. Again, the results were negative and the images were blank after the fluorescent gel scan. The gels were then fixed and stained with colloidal Coomassie to confirm that there was indeed protein present on the gels. Figure 7-7 C and D shows the infected and non-infected cell extracts respectively after staining with colloidal Coomassie. These images confirm that there were indeed proteins present on the gels.

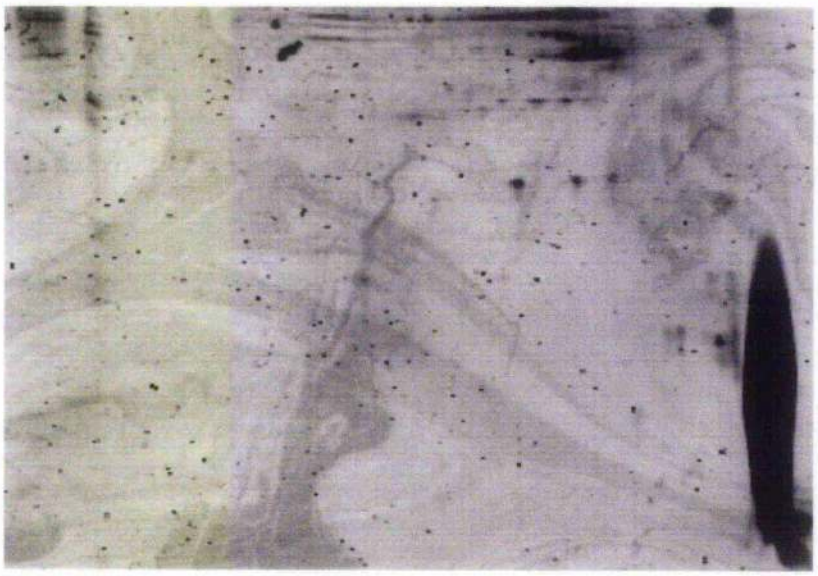


Figure 7-6 HFF sample extract separated by 2-DE on a 7cm pH 3-10NL IEF strip and gel stained with ProQ Diamond gel stain. This is a whole cell extract stained with ProQ Diamond gel stain, a much larger number of protein spots should be visible.

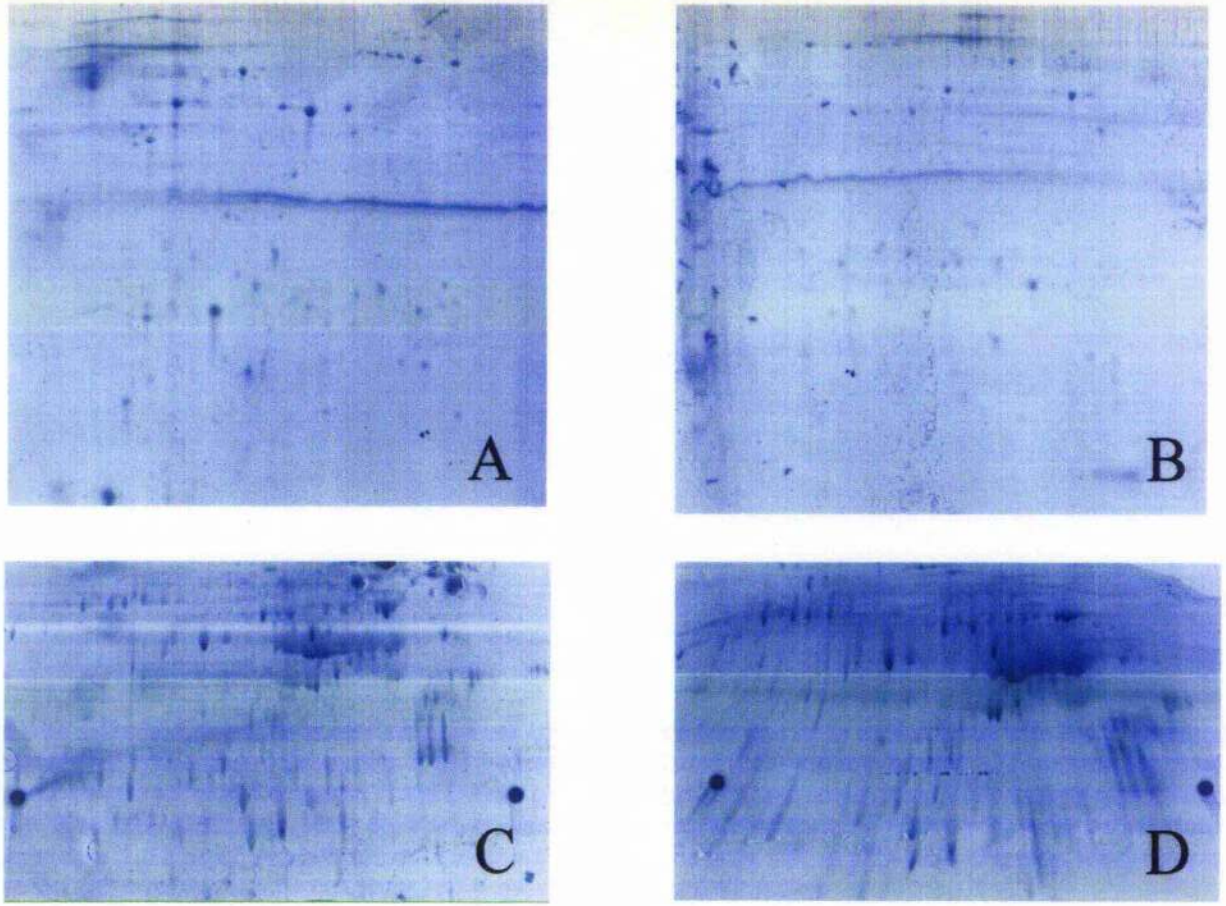


Figure 7-7 2-DE gels stained with colloidal Coomassie to confirm presence of protein. The four gels when stained firstly with ProQ® Diamond phosphoprotein gel stain were blank. Gels were destained and restained with colloidal Coomassie. A and B show gels of infected (24h) and non-infected HFFs extracts separated on 7cm pH3-10NL IPG strips, separated by 2-DE and the gel stained with colloidal Coomassie. C and D show gels of infected (12h) and non-infected HFFs extracts separated on 24cm pH 3-10NL IPG strips, separated by 2-DE and stained with colloidal Coomassie. Many spots can be seen on these images, resolution is however poor perhaps due to too high a voltage or a discontinuous buffer system in the second dimension. D. 320μg of non-infected cell extract separated by 2DE using a 24cm pH3-10NL IPG strip. Many spots can be seen on this image, resolution is however poor perhaps due to too high a voltage or a discontinuous buffer system in the second dimension.

7.3.2.1 Combination of QIAGEN + Pro-Q® Diamond phosphoprotein gel stain

Infected and non-infected cell extracts were separated into their phosphorylated and unphosphorylated fractions, separated by 1-DE and 2-DE and the gels stained with ProQ Diamond Stain. The premise behind this experiment was that the QIAGEN columns would enrich the phosphorylated extract, all the proteins present on the gel would be phosphorylated and therefore be detected by the ProQ® Diamond stain.

7.3.2.1.1 1-DE

Figure 7-8A shows the phosphorylated protein extract from *Toxoplasma gondii* infected and non-infected cells run on 1-DE gels and stained with Pro-Q Diamond stain. Figure 7-8B shows the same phosphorylated protein extracts stained with colloidal Coomassie. Protein (5µg) was loaded in each lane, however, there did appear to be a difference in the amount of protein loaded between the non-infected and infected samples when stained with ProQ stain. The gels were destained and restained with colloidal Coomassie and here the protein loading appears more equal, therefore showing that the ProQ stain is not very quantitative.

Four matching gel bands from each extract were cut out and analysed by MS/MS. The peptides were analysed by MS/MS and the resulting peptide fragmentation data searched against the NCBI homosapiens database and the ToxoDB database with fixed carbomidomethylation modification and variable oxidation of methionine and phosphorylation of threonine, tyrosine and serine residues modifications. The bands chosen and the MS results can be seen in Figure 7-9. Thirteen human and one *Toxoplasma* protein were identified from the eight gel bands. The matching gel bands should contain the same proteins. No phosphorylated peptides were identified. Matching bands 7+8 both contained caldesmon. The volume of each band was measured (Figure 7-9 bottom right table). As can be seen from this table, the bands were all stronger in the non-infected extract. The protein concentration of the two samples was determined using the Bradford assay therefore it is known that equal amounts of protein were loaded onto the 1-DE gel. It is possible that there was there was leakage in the loading well as it appears that more protein was present in the non-infected extract. Due to this it was not possible to determine quantitative differences between these phosphorylated extracts from the non-infected and infected cells.

The identified proteins were searched in harvester and the literature and these proteins can all be phosphorylated. Of interest was calmodulin where phosphorylation results in a decreased activity of the protein. Calmodulin mediates control of a number of protein

kinases and phosphatases. An increase in phosphorylation of lamin A/C occurs before envelope disintegration and probably plays a role in regulating lamin associations. Protein tyrosine phosphatase is a signalling molecule which regulates cell growth, differentiation, mitosis and apoptosis.

7.3.2.1.2 2-DE

The methods of purification of the phosphoproteome using the QIAGEN kit and the phospho stain Pro-Q® Diamond were used in concert. Using sample from a 12h *Toxoplasma gondii* infection experiment, the phosphorylated component of infected and non-infected cells was isolated using the QIAGEN phosphoprotein purification kit. The samples were then separated using 2-DE and when stained with Pro-Q® Diamond stain, according to the manufacturers instructions, were seen to be blank (Figure 7-10 A and B). The gels were then fixed and stained with colloidal Coomassie to determine whether there were proteins present on the gel (Figure 7-10 C and D). As can be seen in C and D, there were proteins present on the gel. Very few spots are present and there was streaking on the gel. This is again likely due to the fact that components of the QIAGEN kit buffer are incompatible with 2-DE (as seen in section 7.3.1) As only phosphorylated proteins should be present on this gel the ProQ® Diamond stained gels (A and B) should not have been blank.

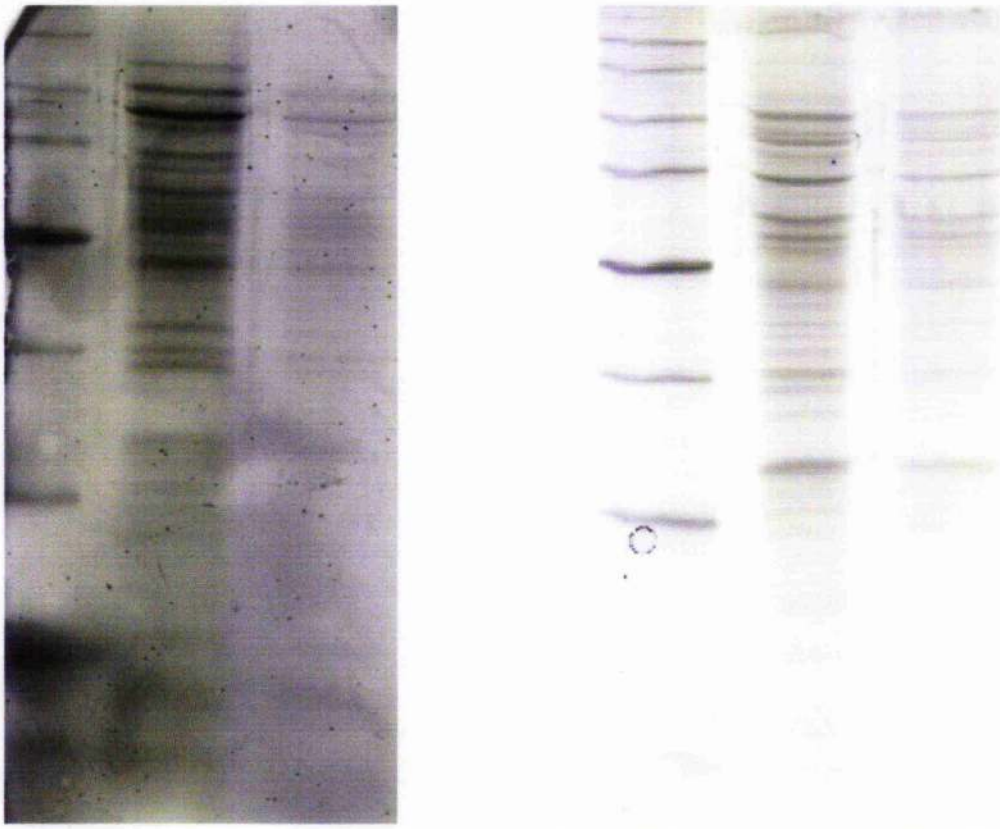
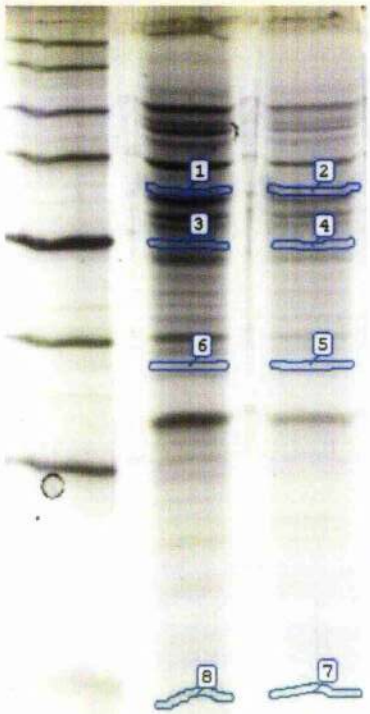


Figure 7-8 1-DE separation of the phosphorylated component of infected and non-infected proteins. The phosphorylated proteins were purified using the QIAGEN columns and 5μg of protein were loaded in each lane. A. Phosphorylated proteins stained with Pro-Q® Diamond phosphoprotein gel stain. There appears to be a difference in the amount of protein loaded between the non-infected and infected samples. B. Phosphorylated proteins stained with colloidal Coomassie. Differences in protein loading are not as apparent with this stain as compared to A. As these samples were previously run through the QIAGEN phosphoprotein column, the hope is that only phosphorylated proteins will be present on the gels. Pro-Q® Diamond stain is supposed to only stain phosphorylated proteins therefore images A and B should be identical. The non-infected samples on both images appear to be quite similar. The infected sample on image A is very weak, is this due to a suppression of phosphorylation in infected cells, a loading issue or a staining issue.



	Non-infected	Infected
1+2	no identification	PDI-associated 4 Lamin A/C isoform 2
3+4	Tubulin alpha Thyroid hormone binding protein precursor Eukaryotic translation initiation factor	no identification
6+5	Eukaryotic translation elongation factor 1 Tropomyosin skeletal muscle Tubulin beta	Protein-tyrosine phosphatase TgGImHMM_1337 ribosomal protein
8+7	Calmodulin Ribosomal protein L22 pro-protein Smooth muscle myosin alkali light chain	Calmodulin

	Non-infected	Infected
1+2	204	162
3+4	207	160
6+5	227	206
7+8	245	230

Figure 7-9 Proteins identified from non-infected and *T.gondii* infected cells which had undergone purification with the QIAGEN kit. The gel image on the left shows which gel bands were selected for MS analysis. The gel bands were excised from the gel and the proteins digested with trypsin. The peptides were analysed by MS/MS and the resulting peptide fragmentation data searched against the NCBI homosapiens database and the ToxoDB database. The MASCOT results can be seen in the table on the top right. The table on the bottom right shows the densitometry results, every band in the non-infected gel is stronger; therefore it would appear that the protein loading of samples was not equal.

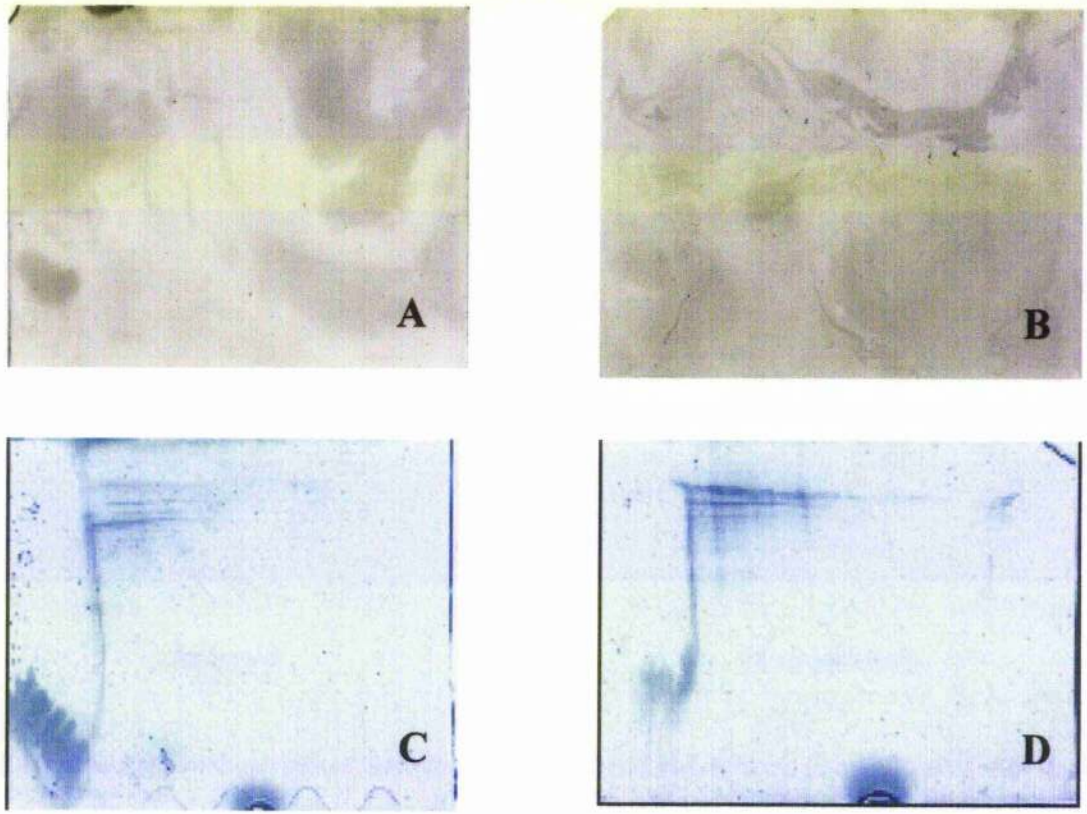


Figure 7-10 Comparison of the phosphorylated extracts of the infected (A and C) and non-infected (B and D) cellular proteome. A and B are images of the gels stained with Pro-Q® Diamond phosphoprotein gel stain and C and D of the gels stained with colloidal Coomassie. The gels stained with Pro-Q Diamond stain were blank. The gels were then fixed and stained with colloidal Coomassie to confirm that protein was present on the gel. C and D confirm that protein is present on the gel, although the proteins are not very resolved. This may have been due to incompatibility of the buffer in the QIAGEN purification kit with 2-DE.

7.3.3 *Phosphatase treatment to investigate the phosphoproteome of T.gondii infected cells*

Shrimp alkaline phosphatase (SAP) was used to dephosphorylate proteins from *T.gondii* infected and non-infected cells. The phosphorylated and dephosphorylated fractions were then analysed by DIGE. SAP is most effective in an alkaline environment similar to that required for efficient CyDye labelling. SAP dephosphorylation of proteins has previously been used by Jonas (2004) and Stensballe (2001) in 1-DE and MS studies respectively. To date there are no published studies where phosphatase treatment has been combined with DIGE.

Figure 7-11 profiles the hypothesised protein positional changes post treatment with SAP. A. The loss of the phosphate groups would cause the protein to migrate in the direction of the acidic end of the IPG strip. This positional change would result in protein volumes both increasing and decreasing after SAP treatment. The change in mass (loss of 80Da per phosphate group) would probably not be significant enough to be detected by DIGE. B. If several different phosphorylated isoforms of the same protein exist, dephosphorylation will result in all isoforms returning to the native protein state, resulting in an increase in volume of the original protein and several decreases in volume due to the loss of the isoforms in the treated sample.

T.gondii infected and non-infected cell extracts were halved and one half treated with SAP, the other mock treated. The four extracts were then separated on two DIGE gels and the spot maps compared.

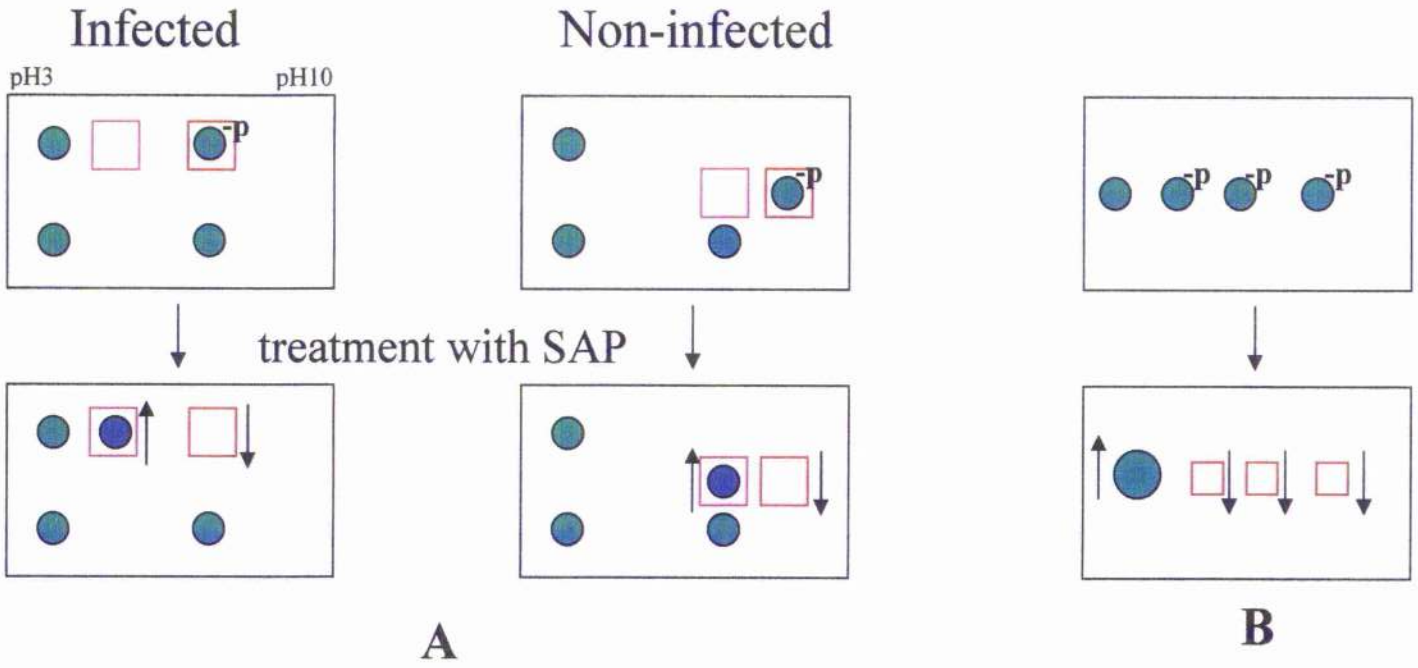


Figure 7-11 Diagram showing the protein positional changes post treatment with SAP. A. The loss of the phosphate groups would cause the protein to migrate in the direction of the acidic end of the IPG strip. The change in mass (loss of 80Da per phosphate group) would probably not be significant enough to be detected by DIGE. The positional pH change would result in protein volumes both increasing and decreasing after SAP treatment. B. If several different isoforms of the same protein exist, dephosphorylation will result in all isoforms returning to the native protein state, resulting in an increase in volume of the original protein and several decreases in volume due to the loss of the isoforms in the treated sample.

The protein expression differences between the phosphorylated and unphosphorylated samples from *Toxoplasma gondii* infected and non-infected cells were analysed using the DIA module of DeCyder. The Cy3 labelled SAP treated infected sample was compared with the Cy5 labelled non treated infected sample. Differences in spot volume within this comparison are indicative of a loss or gain of phosphate group. In the infected sample, spots whose volume was down regulated after treatment were phosphorylated. Spots whose volume increased after treatment were most probably the native protein spot. The volume of these spots increased as the phosphate groups were removed from the isoforms of the protein therefore clustering all the isoforms as the native protein. A screenshot of this comparison of infected cells mock-treated versus infected cells treated with SAP can be seen in Figure 7-12. Any matched spots whose volume differed by 2-fold or greater were deemed to be significant. Figure 7-13 is a screen shot of the BVA module where the differentially expressed spots of interest were matched to the preparative gel. The top left image shows a spot on the preparative gel and the matching spot on the Cy3 labelled infected treated cells is highlighted in pink on the image on the right.

The same analysis was also used to compare the Cy5 labelled mock-treated non-infected cells with the Cy3 labelled SAP treated non-infected sample (Figure 7-14 and Figure 7-15).

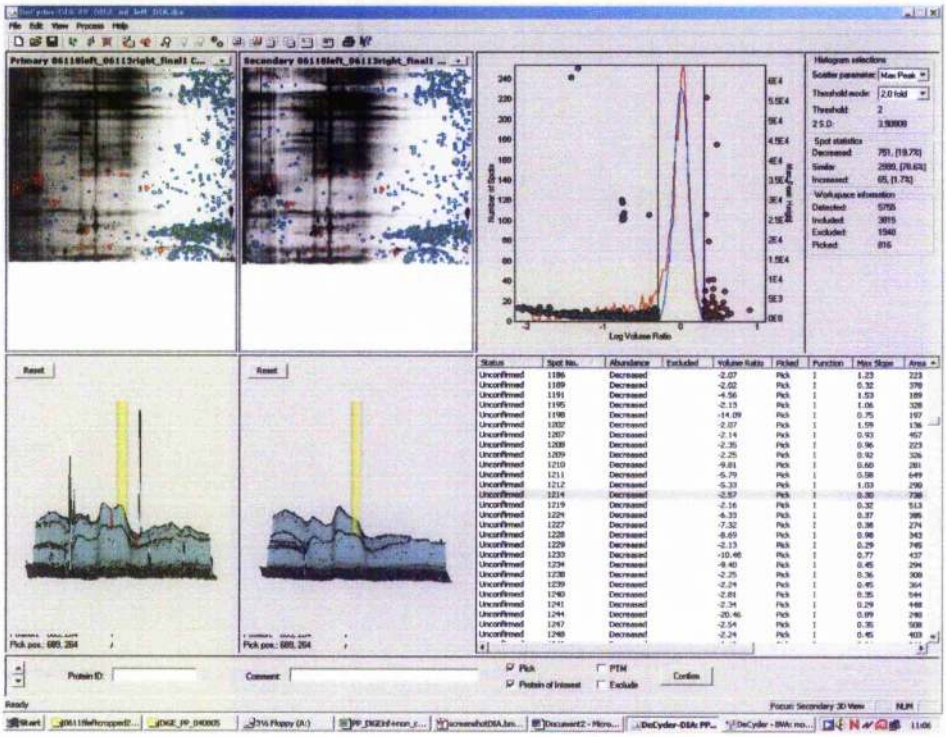


Figure 7-12 Screenshot of the DeCyder DIA module. The gel images show the Cy5 labelled infected cells mock-treated on left and the matched Cy3 labelled infected cells treated with SAP on the right. The blue spots on the right of the image are most probably noise and not biological differences between samples.

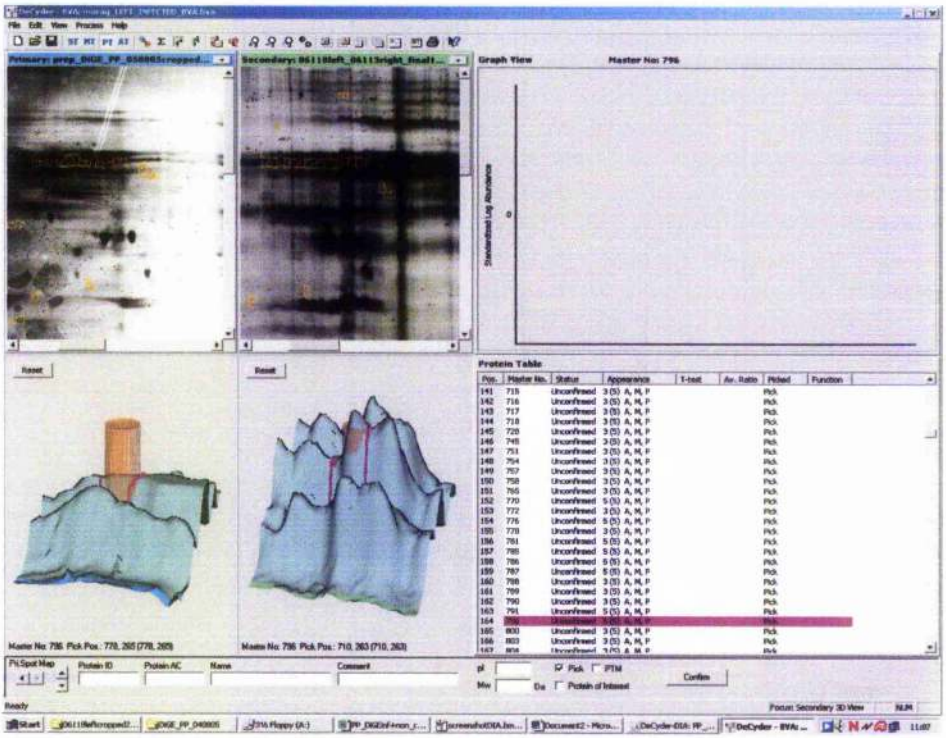


Figure 7-13 Screenshot of the DeCyder BVA module comparing the prep gel with the Cy5 labelled infected cell gel. The top left image shows a spot on the preparative gel and the matching spot on the Cy3 labelled infected treated cells highlighted in pink on the image on the right.

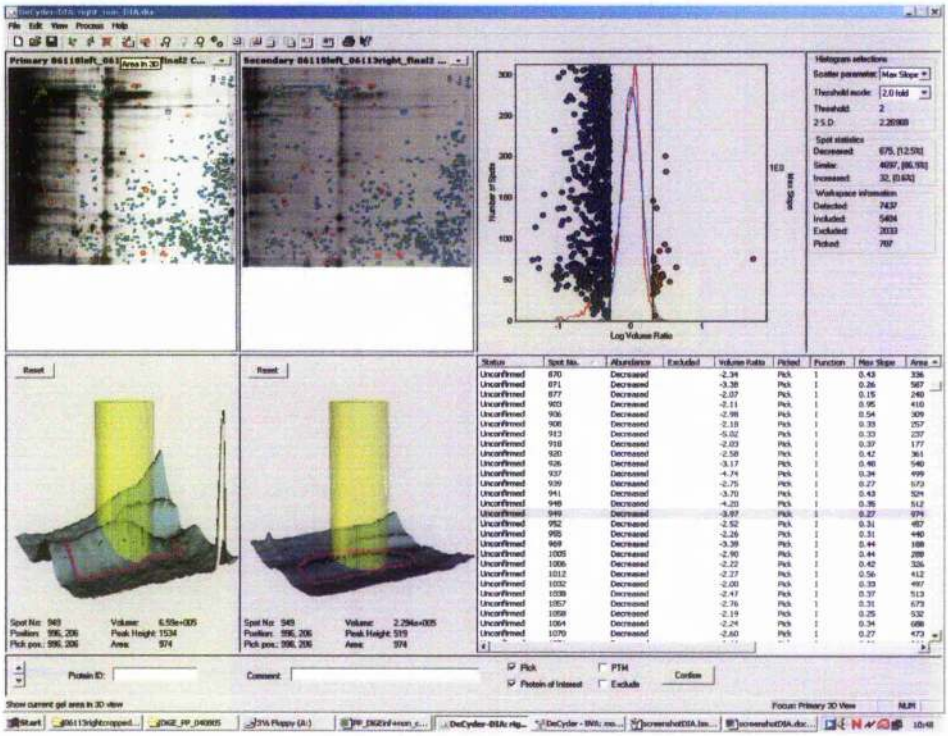


Figure 7-14 Screenshot of the DeCyder DIA module. The gel images show the Cy5 labelled non-infected cells mock-treated on the left and the matched Cy3 labelled non-infected cells treated with SAP on the right. Again, the blue spots on the far right are most probably noise and not true biological differences.

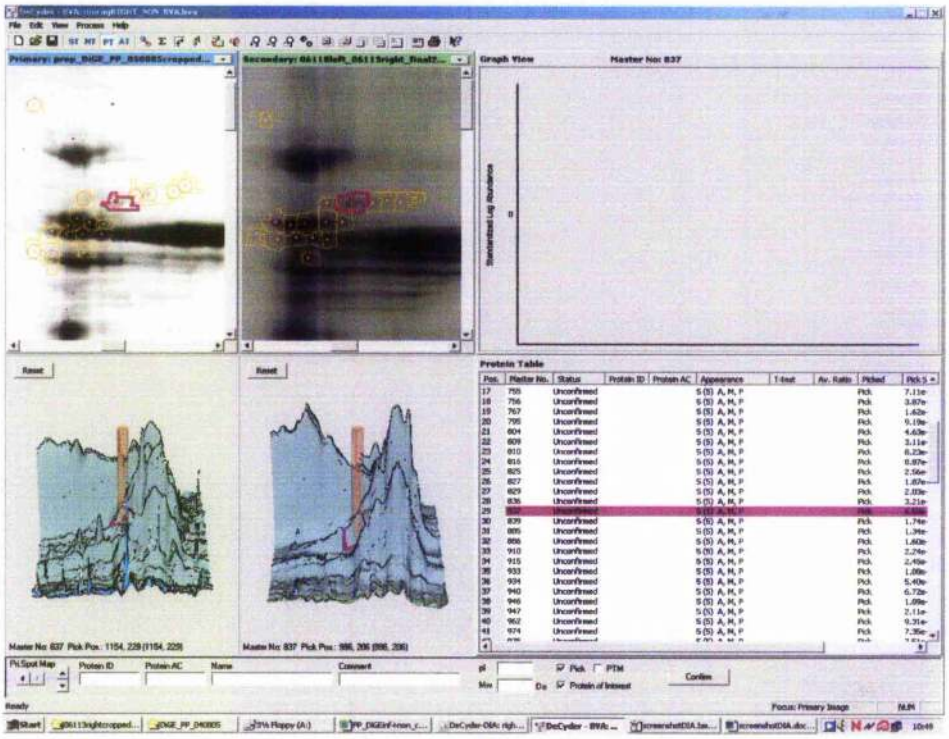


Figure 7-15 Screenshot of the DeCyder BVA module comparing Cy3 labelled non-infected cells treated with SAP versus Cy5 labelled non-infected cells mock-treated. The top left image shows a spot on the preparative gel and the matching spot on the Cy3 labelled non-infected treated cells highlighted in pink on the image on the right.

Lists of the significant spots from both of the comparisons were compiled. This provided information about which proteins were potentially phosphorylated in the *Toxoplasma gondii* infected and non-infected cells. These lists were then collated and compared. Spots which were present in both comparisons and whose volume changes differed indicated that they were phosphorylated or dephosphorylated during the invasion process.

	Sample 1	Sample 2	Number down regulated	Number up regulated	Number matched to prep gel	Number down regulated spots picked	Number up regulated spots picked
DIGE gel 1	Infected, mock treated	Infected, SAP treated	751	65	52	48	4
DIGE gel 2	Non-infected mock treated	Non-infected, SAP treated	675	32	104	97	7

Table 7-4 Number of differentially expressed spots detected on the two DIGE gels in the SAP experiment.

In the comparison between infected SAP treated proteins versus infected mock treated proteins using the DIA module 751 spots had a decreased volume following SAP treatment and 65 spots had an increase in volume following treatment (2 fold volume difference or greater). These spots of interest were then matched to the preparative gel in the BVA module. 52 spots were matched to the prep gel and picked, of which 4 were up regulated and 48 down regulated in volume after SAP treatment.

In the comparison between non-infected SAP treated proteins versus non-infected mock-treated proteins using the DIA module, 675 spots had a decreased volume following SAP treatment and 32 spots had an increased volume following treatment (2 fold difference or greater). These spots of interest were then matched to the preparative gel in the BVA module. 104 spots were matched to the preparative gel and picked, of which 7 were up regulated and 97 down regulated in volume following SAP treatment.

All of the spots of interest were picked using the Ettan Spot Handling workstation (Amersham Biosciences) and subjected to MALDI analysis. Spots which were not identified using MALDI but some spectra was seen were then analysed using MS/MS. The results from the MS are in Table 7-5. The table profiles all the information about the spots identified using MS. Columns running from left to right are; infected or non-infected sample, BVA spot number, DIA spot number, decreased or increased spot volume, volume

ratio between Cy3 and Cy5 matched spots, number of proteins identified by MS per spot, gene name, score, protein name, mass, pI, sequence coverage, gI number and type of MS used.

The first comparison was of infected treated cells (all proteins dephosphorylated) versus infected non treated cells (proteins in native state, phosphate groups present). Any decreases in volume indicate that the volume was greater pre-treatment and therefore the decrease in spot volume could be attributed to loss of phosphate groups post treatment. This indicates that the proteins whose volume decreases after treatment were phosphorylated in the infected cell.

The second comparison was of non-infected treated cells (all proteins dephosphorylated) versus infected non-infected non treated cells (proteins in native state, phosphate groups present). Any decreases in volume indicate that the volume was greater pre-treatment and therefore the decrease in spot volume could be attributed to loss of phosphate groups post treatment. This indicates that the proteins whose volume decreases after treatment were phosphorylated in the non-infected cell.

The differentially expressed spots from the infected and non-infected spot maps were then compared. Any spots which had the same co-ordinates in both studies were assumed to be the same proteins. Nine spots were matched between the studies, of these two were identified by MS. They were both down regulated in expression post treatment with SAP and were present in equal volumes in both the infected and non-infected cells pre and post treatment, and so were excluded from the results.

Spots whose volume differed in either the infected or non-infected comparison and were not matched between studies were assumed to be differentially expressed due to the presence of the parasite. The 153 spots of interest were excised from the prep gel using the Ettan Spot Handling Work Station and analysed by MALDI. Of these, fifteen were identified. Spots which had shown some spectra in the MALDI analysis but were not identified were then analysed by MS/MS. Thirty three spots were analysed and of these six proteins were identified.

An example of a spot whose volume differed post SAP treatment can be seen in Figure 7-16.

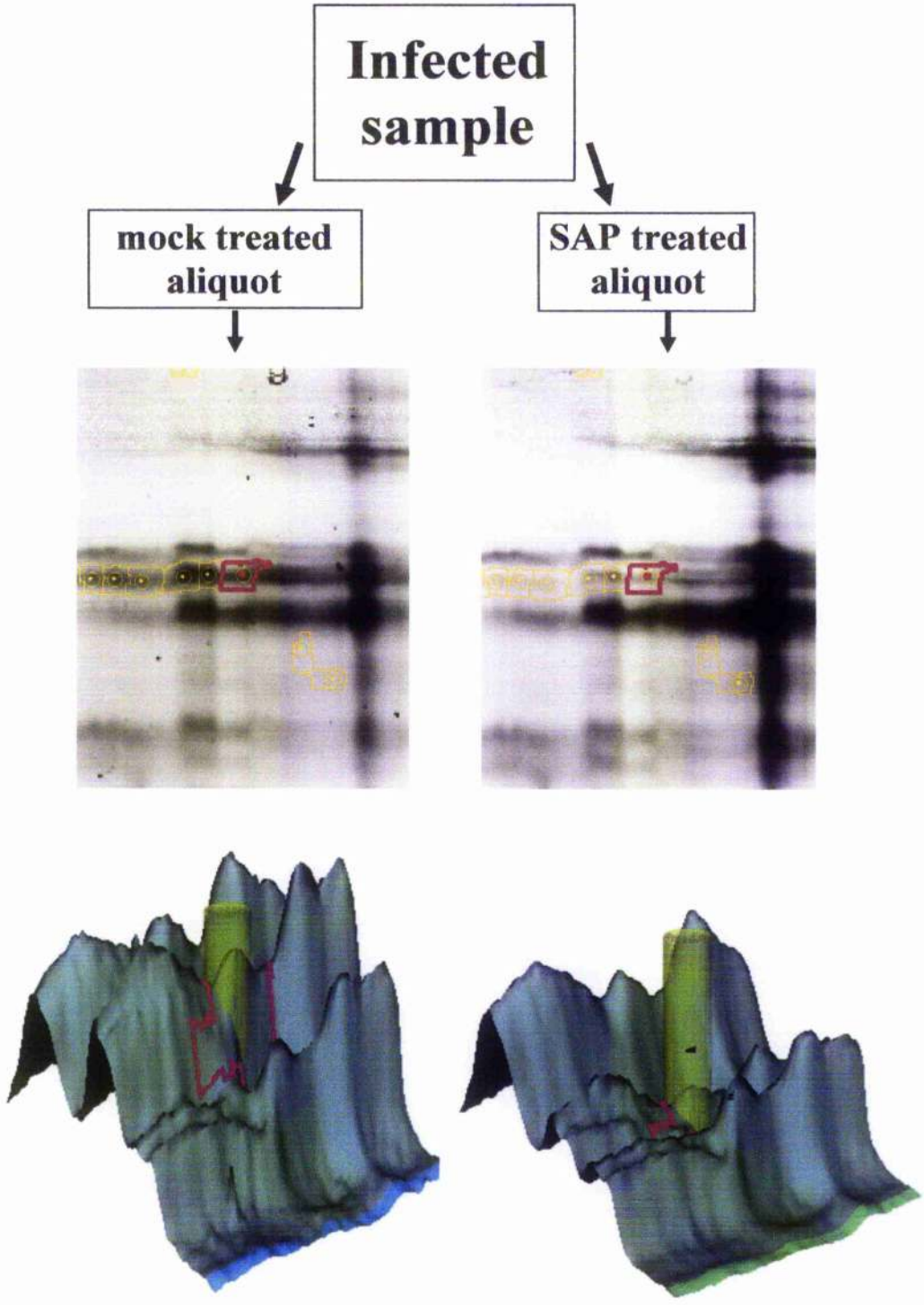


Figure 7-16 Detection of phosphoproteins using shrimp alkaline phosphatase (SAP). The cell extract from infected cells was divided into 2 aliquots, one of which was treated with SAP. Both samples were then separated using DIGE. Phosphorylated proteins were identified by comparing spots on the 2-DE images which have migrated differently due to loss of phosphate groups. The top images compare the spot on the gel, the 3D images show the volume difference very clearly, the protein volume was greater pre-SAP treatment. This protein was identified by MS as vimentin.

	BVA no.	DIA no.	spot volume	volume ratio	number of spectra	gene name	score	protein	mass	pI	s.c	gI number	MS
infected	629	989	Decreased	-2.51	1 of 1	HSPD1	365	Chaperonin (HSP60)	61157	5.7		306890	MS/MS
infected	632	993	Decreased	-5.17	1 of 1	LMNA	104	LMNA protein	53222	7.7		21619981	MS/MS
infected	776	1191	Decreased	-4.56	1 of 1	VIM	68	Vimentin	53710	5.06	18	37852	MALDI
infected	781	1198	Decreased	-14.09	1 of 1	VIM	444	Vimentin	53710	5.06	25	37852	MS/MS
infected	785	1202	Decreased	-2.07	1 of 1	VIM	191	Vimentin	53676	5.06	34	7576229	MALDI
infected	786	1207	Decreased	-2.14	1 of 1	VIM	126	Vimentin	53710	5.06	26	37852	MALDI
infected	787	1208	Decreased	-2.35	1 of 1	VIM	481	Vimentin	53710	5.06	31	37852	MS/MS
infected	791	1212	Decreased	-5.33	1 of 1	VIM	217	Vimentin	53676	5.06	38	7576229	MALDI
infected	796	1214	Decreased	-2.57	1 of 1	VIM	134	Vimentin	53676	5.06	28	7576229	MALDI
infected	805	1228	Decreased	-8.69	1 of 1	VIM	171	Vimentin	53710	5.06	35	37852	MALDI
infected	810	1234	Decreased	-9.4	1 of 1	VIM	229	Vimentin	53710	5.06	44	37852	MALDI
infected	811	1238	Decreased	-2.25	1 of 1	VIM	211	Vimentin	53710	5.06	39	37852	MALDI
infected	812	1239	Decreased	-2.24	1 of 1	VIM	189	Vimentin	53710	5.06	43	37852	MALDI
infected	813	1240	Decreased	-2.81	1 of 1	VIM	257	Vimentin	53676	5.06	48	7576229	MALDI
infected	819	1244	Decreased	-20.46	1 of 1	VIM	157	Vimentin	53710	5.06	33	37852	MALDI
infected	820	1247	Decreased	-2.54	1 of 1	VIM	117	Vimentin	53710	5.06		37852	MALDI
infected	821	1248	Decreased	-2.24	1 of 1	VIM	135	Vimentin	53710	5.06	29	37852	MALDI
non-infected	836	948	Decreased	-4.2	1 of 1	VIM	234	Vimentin	53710	5.06	45	37852	MALDI
infected	845	1277	Decreased	-2.07	1 of 1	VIM	153	Vimentin	53676	5.06	33	7576229	MALDI
non-infected	933	1057	Decreased	-2.76	3 of 3	CALD1	174	Caldesmon	62715	6.18	9	179830	MS/MS
non-infected	933	1057	Decreased	-2.76	2 of 3	NUCB1	155	Nucleobindin	53698	5.15	13	189308	MS/MS
non-infected	933	1057	Decreased	-2.76	1 of 3	P4HB	345	Protein disulfide isomerase	57043	6.1	21	860986	MS/MS
non-infected	947	1071	Decreased	-2.66	1 of 1	-	116	Thyroid hormone binding protein	57468	4.82	24	339647	MALDI
non-infected	975	1099	Decreased	-2.69	1 of 1	VIM	71	Vimentin	53710	5.06	15	37852	MALDI
non-infected	1044	1169	Decreased	-3.03	1 of 1	VIM	191	Vimentin	53676	5.06	35	7576229	MALDI
non-infected	1063	1188	Decreased	-3.67	1 of 1	CALR	102	Calreticulin	48283	4.29	7	30583735	MS/MS

Table 7-5 Results table from the DIGE phosphorylation experiment. Table legend in text.

Protein	INF/NON-INF	Function
Caldesmon	1 non-infected	mitosis
Calreticulin	1 non-infected	protein folding
Nucleobindin	1 non-infected	protein fate
Protein disulfide isomerase	1 non-infected	structural
Thyroid hormone binding protein	1 non-infected	signal transduction/cellular communication
Chaperonin (HSP60)	1 infected	energy metabolism
LMNA protein	1 infected	unknown
Vimentin	16 infected, 2 non-infected	structural

Table 7-6 SAP-DIGE results

Proteins in blue (Table 7-6) are those spots which had a decreased volume in non-infected cells post treatment with SAP. Therefore these proteins were phosphorylated in non-infected cells. As these exact spots were not seen in the infected comparison we can conclude that they were dephosphorylated in infected cells due to the presence of the parasite. Caldesmon, calreticulin, nucleobindin, protein disulfide isomerase and thyroid hormone binding protein precursor were all phosphorylated in the non-infected cell. The presence of the parasite caused these proteins to be dephosphorylated in the cell. The presence of the parasite causes down regulation of caldesmon (seen in the gel-to-gel and DIGE experiments in Chapters III and IV) in the infected cell therefore the dephosphorylation of this protein may lead to its inactivation in the infected cell. Caldesmon is an actin and myosin binding protein involved in mitosis. It is possible that dephosphorylation of this protein in the infected cell causes modulation of mitosis.

Proteins in pink (Table 7-6) are proteins which had a decreased volume in infected cells post treatment with SAP, therefore were phosphorylated in the infected cell due to the presence of the parasite. It can be concluded that vimentin, chaperonin and LMNA protein were phosphorylated in the infected cell and not in the non-infected cell.

7.4 Discussion

This chapter focused on the study of the phosphoproteome in *Toxoplasma gondii* infected cells. Elucidation of which proteins are phosphorylated during the invasion response will lead to a better understanding of the host pathogen response and which pathways and proteins are activated/inhibited due to the presence of the parasite. The detailed characterisation of the phosphoproteome of the cell along with other types of PTM is essential in order to gain a true understanding of protein function and regulatory events in cells and organisms. Phosphorylation is the most common form of PTM and is an essential process for the functioning of many proteins and pathways within the cell.

Three techniques were used to investigate the phosphoproteome of *Toxoplasma gondii* infected cells. The QIAGEN PhosphoProtein Purification Kit is reported to separate phosphorylated proteins from unphosphorylated proteins and these fractions from *T.gondii* infected and non-infected cells were analysed by 1-DE and 2-DE. Gels showed clear differences between purified phosphorylated extracts from infected and non-infected cells, although the 2-DE gels were mostly incompatible with the QIAGEN technique. In the 1-DE comparison, eleven human proteins were identified by MS. All had the potential to be phosphorylated, although no phosphopeptides were detected by MS. In the 2-DE experiment, nine differentially expressed human proteins were identified by MS. In the *T.gondii* infected cell; caldesmon, guanine nucleotide binding protein, tropomyosin 1 and vimentin were up regulated. ALL-1 protein, BiP and tumour rejection antigen were all down regulated in the infected cell. Of these only one was identified having a phosphopeptide using MS and five of the remaining eight have the potential to be phosphorylated. Detection of phosphopeptides by MS/MS is inherently difficult. MS/MS identifies a number of peptides from the peptide mixture. If the peptide that contains the phospho modification is not one of the peptides identified then it is not possible to determine if that protein is phosphorylated. The advantage of using the gel based method is that it is possible to see PTM on the gel and therefore be certain that the protein is modified even if it is not possible to identify the modification using MS. This is an advantage that gel based systems have over MUDPIT. This MS based technique will only identify PTM if the specifically modified peptides are identified by MS.

Pro-Q® Diamond phosphoprotein gel stain is reported to selectively stain phosphorylated proteins. This technique provided very poor results, even when used in combination with the QIAGEN PhosphoProtein Purification Kit. Many experimental replicates were carried

out but none were of good enough quality to profile the phosphoproteome of *Toxoplasma gondii* infected cells therefore this technique provided no results. Personal communication with Molecular Probes unfortunately did not help to resolve the problems associated with this technique. A study by Raggiaschi *et al* (2005) used the Pro-Q® Diamond phosphoprotein gel stain to selectively stain the phosphoproteins on 2-DE gels of rat cortical neuron proteins. They also found that the stain did not provide a high enough sensitivity for comprehensive analysis of the phosphoproteome.

A novel technique was developed which investigated the infected cell phosphoproteome by dephosphorylation of proteins using Shrimp Alkaline Phosphatase (SAP) used in combination with DIGE. Phosphorylation of proteins leads to a change of the net charge of proteins and thus the charge variation occurring after phosphatase treatment can be exploited to discriminate phosphorylated from unphosphorylated proteins by a mass shift during 2-DE. Samples were halved, half was treated with SAP and half was mock treated. The proteins were then separated and compared using DIGE. Using SAP to study phosphorylation status has previously been used in concert with 1-DE gels (Jonas and Privalsky, 2004), conventional gel-to-gel (Raggiaschi *et al.*, 2005) and MS analysis (Stensballe, 2001). This is the first example of SAP treatment combined with DIGE to study the phosphoproteome. This technique proved quite successful, with many differences being seen between the SAP treated and mock-treated samples. Using MS twenty four spots were identified as being differently expressed during the invasion response. This amounted to three individual phosphorylated human proteins (chaperonin, LMNA protein and vimentin) being up regulated in the infected cell and the down regulation of four phosphorylated human proteins (caldesmon, calreticulin, nucleobindin and thyroid hormone binding protein). The down regulation of caldesmon seen in this experiment contradicts with the results seen in QIAGEN experiment, however it does agree with the data from the gel-to-gel and DIGE experiments in Chapters III and IV. The up regulation of the phosphorylated isoform of vimentin is common to both the SAP-DIGE and QIAGEN experiments.

Vimentin is known to be one of the most prominent phosphoproteins in the cell and phosphorylation is enhanced during cell division, at which time vimentin filaments are significantly reorganized. In context with the infected cell, we know that vimentin expression is modulated due to the presence of the parasite as seen in the gel to gel and DIGE experiments. Massive structural reorganisation takes place in the parasite infected cell (Coppens and Joiner, 2001). In a normal non-infected cell, vimentin is distributed throughout however, in *T.gondii* infected cells, vimentin is reorganised to around the

parasitophorous vacuole (PV) where it is thought it acts to anchor the PV to the host cell nuclear surface by the network of intermediate filaments and may also help to bring metabolites and nutrients to the parasites (Coppens and Joiner, 2001). Therefore perhaps the increase in the phosphorylated isoform of this protein helps to bring about such modulations in the parasite infected cell. An increased phosphorylation of the lamins occurs during mitosis before envelope disintegration when the lamina matrix is being disassembled. Lamin proteins are thought to be involved in nuclear stability, chromatin structure and gene expression. Perhaps the increase in phosphorylation of LMNA is related to a change in mitosis in the infected cell.

In conclusion, three techniques were exploited to undertake comparative phosphoproteomics. The QIAGEN PhosphoPurification Kit provided some information but was in general incompatible with 2-DE. The Pro-Q® Diamond phosphoprotein gel stain gave very poor results. These approaches are difficult to optimise because of the proprietary nature of the reagents. The SAP-DIGE experiment was much more promising although the technique clearly needs optimised to provide a larger results set. Perhaps running an increased number of replicates would be beneficial but unfortunately was not possible due to time constraints. Further studies of the phosphoproteome of *Toxoplasma gondii* infected cells will lead to an increased understanding of the host-pathogen response and help to elucidate novel and proteins and pathways involved.

The function of all of the modulated proteins in the *Toxoplasma gondii* infected cell identified by the gel-to-gel electrophoresis, DIGE and phosphoproteomic studies will be discussed in the following Chapter.

Chapter VIII Functional analysis of the differentially expressed proteins in *Toxoplasma gondii* infected cells

8.1 Introduction

Using the proteomic techniques discussed previously, 157 proteins were identified as being differentially expressed in *Toxoplasma gondii* infected HFFs. To gain a more thorough understanding of the host response to infection it is paramount to determine whether these proteins are modulated in expression as a result of the parasite inducing modulation for its own benefit, or alternatively, the changes may be due to the host cell inducing a protective response to the parasite. Also of consideration is whether the proteins are modulated as a bystander effect from the first two classes of modulation. Proteins which are specifically induced or repressed in *T.gondii* infected cells are particularly interesting for further studies as they might be used to dissect specific host-parasite interactions.

Pathogens have evolved multiple strategies to avoid host defence mechanisms including host cell invasion, host mimicry and antigenic variation. The better evolved the pathogen is in this area, the more successful it will be. Invasion involves physiological changes in both the host cells and invading micro-organisms. These changes are relayed by signal transduction systems and are mirrored by gene and protein expression changes in the host and the pathogen. In general, when an intracellular pathogen invades a host cell, host immune response genes are activated (Cummings and Relman, 2000; Blader *et al.*, 2001) and dependant on the pathogen and its intracellular lifestyle, the host gene and protein response can be relatively little or can involve extensive modulation. A common theme in intracellular parasitism is modulation of host cell signalling and transcription to create a replicative niche. While much progress has been made regarding viral and bacterial pathogens, few host cell pathways are known to be activated by *Toxoplasma* and even less is known about the roles they play in regulating parasite growth (Spear *et al.*, 2006). This study investigated the host cell proteome in response to *T.gondii* infection.

For the purpose of this study, protein functions were determined using the following websites: www.hprd.org and www.harvester.embl.de. All of the proteins modulated in the study were assigned a putative functional classification based on the schema put forward by the Munich Information Centre for Protein Sequences (http://mips.gsf.de/projects/funcat). FunCat is an annotation scheme for the functional description of proteins from prokaryotes, unicellular eukaryotes, plants and animals.

Several papers look at the functions of proteins identified using proteomics and discussing the functional roles ((Bevan *et al.*, 1998; Alfonso *et al.*, 2005)) often making their own system of functional classification (Jiang *et al.*, 2005). However, to date, there is no single method of functional classification. Pathways were investigated using the KEGG Pathway Database (http://www.genome.jp/kegg/pathway.html).

The results section of this chapter is divided into 19 sections corresponding to 19 of the functional categories put forward by MIPS. A further two categories highlight the modulated proteins specifically involved in the host cell apoptotic host cell response and a category to discuss the modulated proteins associated with the mitochondria. Each section profiles which proteins were modulated in the *Toxoplasma gondii* infected cell. What role these modulated proteins may play in the infected cell is put forward in the discussion. The conclusions presented in this chapter are hypothetical. The proteins have not been assigned a definitive role or place within the pathogen-infected cell, but merely assigned a putative function and hypothesised to how it may fit into the host-pathogen response.

Grouping the modulated proteins by function will highlight clusters of functionally similar proteins. This will in turn show to what degree host cell functions are modulated due to the presence of the parasite. Of special interest are modulated pathways in the infected cell. If several proteins in the one pathway are modulated, it is possible to hypothesise that this pathway plays an important role in the host-pathogen response.

As discussed in the preceding chapters, *Toxoplasma gondii* is known to cause host cell modification upon invasion and infection. Studies by Coppens and Joiner (2001); Gail et al. (2004) and Blader et al., (2001) have implicated modulation of host genes and proteins in the response to invasion with *Toxoplasma gondii.* This study represents the first global proteomics study to look at potentially the entire host cell protein response to invasion with an apicomplexan parasite.

This chapter aims to bring together all the results from the previous chapters which investigated the host cell protein response to infection with *Toxoplasma gondii*, and to present this data in the context of what functions the modulated host cell proteins play in the pathogen infected cell.

8.2 Materials and Methods

Each of the 157 proteins which were identified by 2-DE as modulated in the parasite infected cell was individually investigated to determine its function. This was done firstly using the harvester search engine from EMBL (www.harvester.embl.de). The Bioinformatic Harvester search engine cross-links public bioinformatic databases and prediction servers to provide fast access to protein specific bioinformatic information. A ranking search engine retrieves the most relevant information based on the query. Harvester collates the information from these databases to provide a universal database profiling information about protein sequence, protein domains and protein localisation. Harvester collects information from the following public databases and prediction servers; Uniprot, SOURCE, Genome Browser, BLAST, SMART, SOSUI, PSORT II, CDART, MapView, NCBI-BLAST, STRING, GenomeBrowser and EMBL, and the data is collected and assembled into a single webpage. Harvester continuously updates pages older than 21 days (Liebel *et al.*, 2004). Figure 8-1 shows a screenshot from the Harvester search page and the Harvester results page. The Harvester site was searched using the protein name or gene name of the modulated protein.

A second search engine, the human protein reference database (www.hprd.org) was also searched to confirm the function of the protein. HPRD provides information on protein domain architecture, post-translational modifications, interaction networks and disease association for each protein in the human proteome (Peri *et al.*, 2004). Figure 8-2 shows a screenshot from the HPRD website. The HPRD site was searched using the protein name or gene name of the modulated protein.

For further functional confirmation, PubMed and general internet searches were carried out for each of the modulated proteins.

The functional classification system used in this chapter is based on the schema put forward by the Munich information Centre for Protein Sequences (MIPS) (www.mips.gsf.de) termed FunCat. The FunCat annotation scheme consists of 28 main categories that cover general features like cellular transport, metabolism and protein activity regulation. In total the superset of FunCat currently contains 1307 categories. These are not species specific as FunCat is not dedicated to a single organism but rather to allow annotation of a large spectrum of organisms. Therefore not all of the 28 categories have been included in this study, as, for example, there is a FunCat category assigned to cover viral, transposon and plasmid gene products that is not applicable to the human

proteins in this study. An example of a study which assigned genes to functional categories based on the MIPS yeast functional catalogue is by Bevan *et al*, 1998.

For the majority of proteins, the cellular function cannot be entirely described with a single functional category as often the protein function falls into several categories (Ruepp *et al*., 2004). Therefore the approach undertaken in this chapter, which used a minimum of three functional references per protein, allowed for functional assignment into the most relevant functional category for that protein. Where it was not possible to unequivocally assign a single function to a protein, the protein was classed into however many applicable categories and this has been documented and a possible explanation provided.

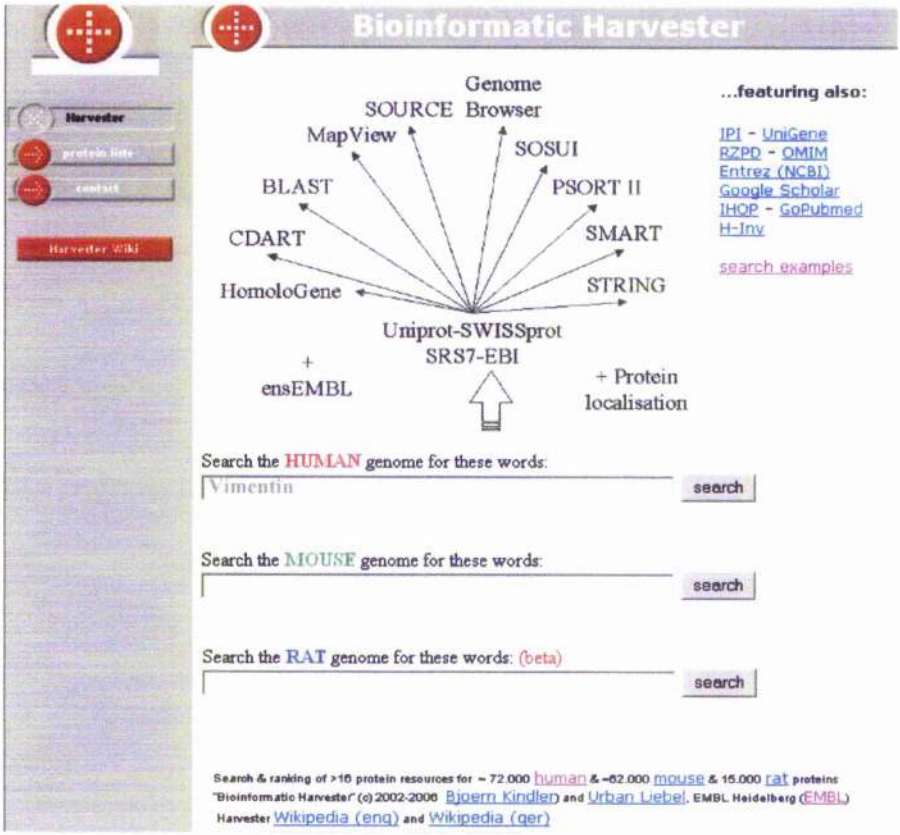


Figure 8-1 A Screenshot of the Harvester search page.

SOURCE GeneReport *H. sapiens*

VIM

Vimentin

UniGene, LocusLink, OMIM, GenAtlas, GeneCard, Ensembl, MapView, Genome Browser

Chromosomal Location	
Chromosome/Cytoband	10p13
Microarray Gene Expression Data	
Data available	Show Gene Expression Data
LocusLink Information	
Locus Link Summary	Along with the microfilaments (actins) and microtubules (tubulins), the intermediate filaments represent a third class of well-characterized cytoskeletal elements. The subunits display a tissue-specific pattern of expression. Desmin (MIM 125660) is the subunit specific for muscle and vimentin the subunit specific for mesenchymal tissue.[supplied by OMIM]
SwissProt Information	
SwissProt Accession No.	P08670 Vimentin (Homo sapiens); 100% similarity over 466 a.a.
Function	vimentins are class-iii intermediate filaments found in various non-epithelial cells, especially mesenchymal cells.
Subunit	homopolymer. interacts with hcv core protein.
Caution	ref.4 (bab71275) sequence differs from that shown due intron retention at the c-terminus.
Ptm	one of the most prominent phosphoproteins in various cells of mesenchymal origin. phosphorylation is enhanced during cell division, at which time vimentin filaments are significantly reorganized.
Tissue Specificity	highly expressed in fibroblasts, some expression in t and b lymphocytes, and little or no expression in burkitt's lymphoma cell lines. expressed in many hormone- independent mammary carcinoma cell lines.
Miscellaneous	interaction: q14194:crmp1; nbexp=1; intact=ebi-353844, ebi-473101;
Similarity	belongs to the intermediate filament family.
SwissProt Copyright	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licenses@isb-sib.ch).

Figure 8-1 B Screenshot of the harvester results page queried with Vimentin. The harvester results page runs for several pages, this screenshot shows information about Vimentin including the gene name, function and PTMs.

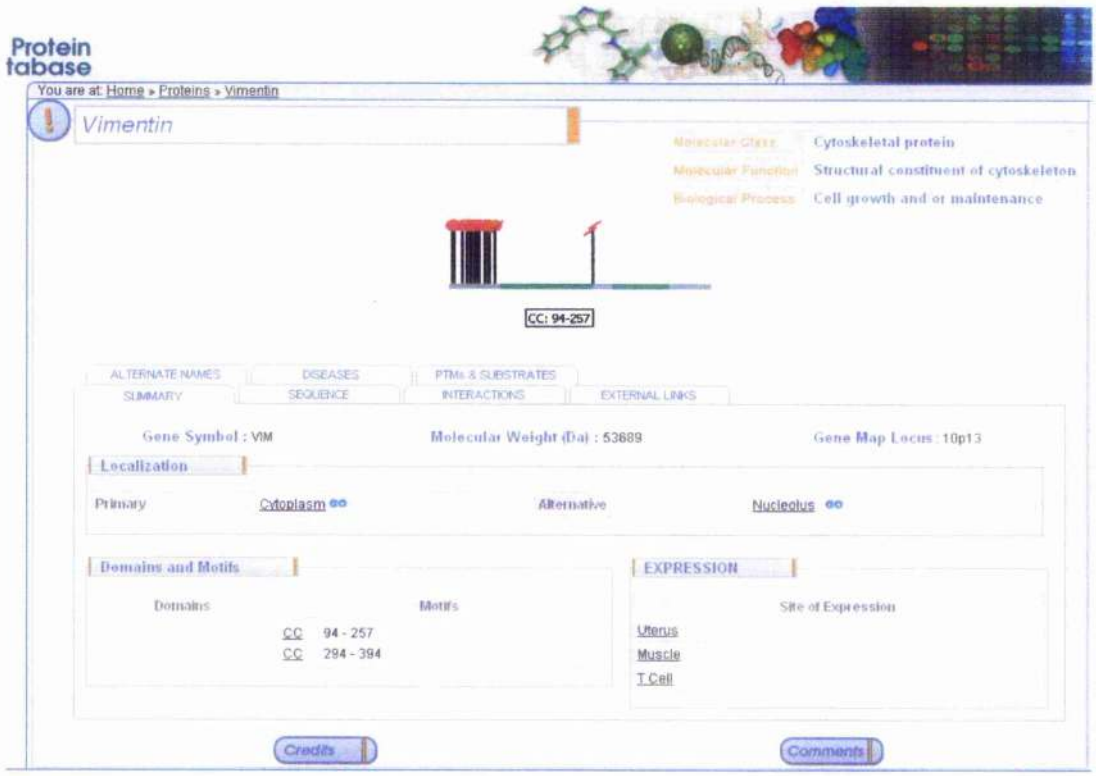


Figure 8-2 A screenshot of the HPRD results webpage. The protein vimentin was queried against hprd. This page serves as the entry point for accessing detailed annotation of proteins in HPRD. Molecular function and biological process for each molecule are indicated at the top right with a graphic representation of the protein domains and motifs and PTMs displayed in the middle. The HUGO gene symbol, the calculated MW and the gene locus are indicated along with the subcellular localisation. All underlined text is linked to PubMed entries that provide further details or to a database source.

8.3 Results

The results section is spilt into 19 functional categories each of which contains a table showing which proteins in that functional category were modulated in the proteomic experiments. What relevance these proteins have in the infected cell will be discussed later in the chapter. The tables include the following information; protein name, gene name, the overall change in expression seen using proteomics (up, down or modulated) and the functional category the protein belongs to. Modulated protein expression indicates that the protein was identified as both up and down regulated in the infected cell therefore the degree of expression is not clear cut. The functional data is summarised in the piechart in Figure 8-3.

8.3.1 Metabolism

Protein name	Gene name	Overall change	Category	Functional category
6-phosphogluconolactonase	**PGLS**	**down**	1	metabolism
Acetyl coenzyme a acyltransferase 2	**ACAT1**	**modulated**	1	metabolism
Adenylate kinase 2 isoform a	**AK2**	**up**	1	metabolism
Aldehyde dehydrogenase 1A3	**ALDH1A3**	**up**	1	metabolism
Aldehyde dehydrogenase X	**ALDH1B1**	**down**	1	metabolism
Aldose Reductase	**AKR1B1**	**up**	1	metabolism
Carbonyl reductase 1	**CBR1**	**down**	1	metabolism
Cargo selection protein TIP47	**M6PRBP1**	**down**	1	metabolism
Cathepsin B	**CTSB**	**down**	1	metabolism
Cathepsin D preprotein	**CTSD**	**modulated**	1	metabolism
Coproporphyrinogen oxidase	**CPOX**	**up**	1	metabolism
Enoyl-coenzyme A hydratase 1	**ECHS1**	**down**	1	metabolism
Glutathione synthetase	**GSS**	**up**	1	metabolism
Glutathione-S-Transferase chain A	**GSTA2**	**down**	1	metabolism
hnrp C-like protein	**HNRPC**	**down**	1	metabolism
hnrp F	**HNRPF**	**modulated**	1	metabolism
hnrp H2	**HNRPH2**	**down**	1	metabolism
hnrp A2/B1 isoform A2	**HNRPA2B1**	**up**	1	metabolism
Vigilin	**HDLBP**	**down**	1	metabolism
lrp protein	**LRP1**	**down**	1	metabolism
MTHFD2 protein	**MTHFD2**	**up**	1	metabolism
N-acetyl-beta-glucosaminidase polypeptide	**HEXB**	**down**	1	metabolism
NDUFS5	**NDUFs5**	**up**	1	metabolism
Nucleoside diphosphate kinase1 -Nm23	**NME1**	**modulated**	1	metabolism
Paraspeckle protein 1	**PSPC1**	**down**	1	metabolism
Phosphoenolpyruvate carboxykinase	**pck1**	**down**	1	metabolism
Protein-L-isoaspartate O-methyltransferase	**PCMT1**	**up**	1	metabolism
Pyridoxine 5'-phosphate oxidase	**PNPO**	**down**	1	metabolism
Pyrophosphatase (inorganic)	**PPA1**	**down**	1	metabolism
RPS12	**RPS12**	**up**	1	metabolism

Table 8-1 Proteins modulated in the pathogen infected host cell involved in metabolism

8.3.2 Energy

Protein name	Gene name	Overall change	Category	Functional category
Aldolase A	**ALDOA**	**up**	2	glycolysis
Aldolase B	**ALDOB**	**up**	2	glycolysis
ATP synthase beta subunit	**ATP5D**	**up**	2	glycolysis
Cytochrome C oxidase subunit Vib	**COX6B**	**down**	2	glycolysis
Dimethyl arginine dimethyl aminohydrolase	**DDAH1**	**down**	2	glycolysis
Enolase 1	**ENO1**	**modulated**	2	glycolysis
Glyceraldehyde 3 phosphate dehydrogenase	**GAPD**	**up**	2	glycolysis
Phosphoglycerate kinase 1	**PGK1**	**modulated**	2	glycolysis
Protein disulfide isomerase	**P4HB**	**up**	2	glycolysis
Pyruvate kinase M2 isozyme	**PKM2**	**modulated**	2	glycolysis
Thioredoxin domain 5 isoform 2	**TXNDC5**	**up**	2	glycolysis
Triose phosphate isomerase	**TPI1**	**modulated**	2	glycolysis

Table 8-2 Proteins modulated in the pathogen infected host cell involved in glycolysis

8.3.3 Cell cycle

Protein name	Gene name	Overall change	Category	Functional category
Dynactin 2	**DCTN2**	**modulated**	3	cell cycle
NEDD5	**NEDD5**	**down**	3	cell cycle
Lamin B2	**LMNB2**	**up**	3	cell cycle
p47	**NSFL1C**	**up**	3	cell cycle
Peroxiredoxin 1	**PRDX1**	**down**	3	cell cycle
Prohibitin	**PHB**	**modulated**	3	cell cycle
S100 calcium binding protein A11	**S100A11**	**down**	3	cell cycle
TPTE2	**TPTE2**	**down**	3	cell cycle
Tubulin Beta 5	**TUBB5**	**down**	3	cell cycle
Ubiquilin 2	**UBQLN2**	**up**	3	cell cycle

Table 8-3 Proteins modulated in the pathogen infected host cell involved in cell cycle

8.3.4 Transcription

Protein name	Gene name	Overall change	Category	Functional category
Brain abundant membrane signal protein	**BASP1**	**up**	4	transcription
CLE7 - C14orf166	**C14orf166**	**up**	4	transcription
hnrp K isoform a	**HNRPK**	**down**	4	transcription
PIAS-Ny protein	**PIAS2**	**up**	4	transcription
Presenelin binding pro	**CSEN**	**up**	4	transcription
Zinc finger protein	**ZNF235**	**up**	4	transcription

Table 8-4 Proteins modulated in the pathogen infected host cell involved in transcription

8.3.5 Translation

Protein name	Gene name	Overall change	Category	Functional category
Elongation factor 1 alpha	**EEFIA2**	**up**	5	translation
Ubiquitin A-52	**UBA52**	**up**	5	translation
WBSCR1	**WBSCR1**	**up**	5	translation

Table 8-5 Proteins modulated in the pathogen infected host cell involved in translation

8.3.6 Protein fate

Protein name	Gene name	Overall change	Category	Functional category
BiP protein, HSPA5 precursor	**HSPA5**	**up**	6	protein fate
CDC37 cell division cycle 37 homolog	**CDC37**	**up**	6	protein fate
Chaperonin containing TCP1 subunit 2beta	**CCT2**	**modulated**	6	protein fate
Chaperonin HSP60	**HSPD1**	**modulated**	6	protein fate
Crocalbin-like protein	**CALU**	**up**	6	protein fate
dnak-type molecular chaperone HSPA6	**HSPA6**	**modulated**	6	protein fate
ERprotein29 precursor	**C12orf8**	**modulated**	6	protein fate
HS70kd pro 42kd ATPase	**HSPA8**	**modulated**	6	protein fate
HSP -MTHSP75	**HSPA9B**	**modulated**	6	protein fate
HSPC054protein	**HSPC054**	**up**	6	protein fate
PDI related protein 5	**PDIR**	**up**	6	Protein fate
Peptidyl prolyl isomerase A (cyclophilin A)	**PPIA**	**down**	6	protein fate
Tumor rejection antigen (gp96) 1	**TRA1**	**down**	6	protein fate
Ubiquitin carboxyl terminal esterase L1	**UCHL1**	**down**	6	protein fate

Table 8-6 Proteins modulated in the pathogen infected host cell involved in protein fate

8.3.7 Protein binding

Protein name	Gene name	Overall change	Category	Functional category
B23 nucleophosmin	**NPM1**	**up**	7	protein binding
Caldesmon	**CALD1**	**down**	7	protein binding
Calponin isoform a	**CNN2**	**down**	7	protein binding
E2IG3 (nucleotide binding protein)	**NS**	**up**	7	protein binding
Heme binding protein 1	**HEBP1**	**down**	7	protein binding
Neuropolypeptide h3	**PBP**	**down**	7	protein binding
THAP domain containing 11	**THAP11**	**up**	7	protein binding
Transgelin	**TAGLN**	**down**	7	protein binding
Transgelin 2	**TAGLN2**	**down**	7	protein binding
Tropomodulin 3	**TMOD3**	**down**	7	protein binding
Tropomyosin 3	**TPM3**	**down**	7	protein binding
UMP-CMP kinase	**UMP-CMPK**	**down**	7	protein binding

Table 8-7 Proteins modulated in the pathogen infected host cell involved in protein binding

8.3.8 Protein activity

Protein name	Gene name	Overall change	Functional category	Category
Visceral adipose-specific SERPIN	**SERPINA1**	modulated	8	protein activity regulation

Table 8-8 Proteins modulated in the pathogen infected host cell involved in protein activity regulation

8.3.9 Cellular transport

Protein name	Gene name	Overall change	Category	Functional category
ATP binding cassette transporter A13	**ABCA13**	**up**	9	cellular transport
Chloride intracellular channel 4	**CLIC4**	**down**	9	cellular transport
Clathrin, light polypeptide A isoform a	**CLTA**	**down**	9	cellular transport
DAT-1 gene human dopamine transporter	**SLC6A3**	**up**	9	cellular transport
Nuclear chloride ion channel protein	**CLIC1**	**down**	9	cellular transport
Tubulin alpha	**TUBA1**	**modulated**	9	cellular transport
Vacuolar protein sorting 33B	**VPs33B**	**up**	9	cellular transport

Table 8-9 Proteins modulated in the pathogen infected host cell involved in cellular transport

8.3.10Signal transduction

Protein name	Gene name	Overall change	Category	Functional category
Annexin 2	**ANXA2**	**up**	10	signal transduction
Cardiotrophin like cytokine	**CLC**	**modulated**	10	signal transduction
Phospholipase C alpha	**PLCL1**	**up**	10	signal transduction
Progesterone receptor membrane component	**PGRMC1**	**up**	10	signal transduction
Protein serine/threonine kinase	**STK16**	**down**	10	signal transduction
Ran-binding protein 1	**RANBP1**	**down**	10	signal transduction
Reticulocalbin 1 precursor; Rcal	**RCN1**	**down**	10	signal transduction
RNA-binding protein regulatory subunit	**PARK7**	**down**	10	signal transduction
Zyxin	**ZYX**	**down**	10	signal transduction

Table 8-10 Proteins modulated in the pathogen infected host cell involved in signal transduction

8.3.11Cell rescue/defence/stress response

Protein name	Gene name	Overall change	Category	Functional category
Carboxyl terminal LIM domain protein	**PDLIM1**	**modulated**	11	stress response
Heat shock protein 27	**HSPB1**	**modulated**	11	stress response
Methylumbelliferyl-acetate deacetylase	**CHIT1**	**down**	11	stress response
PDI related protein 5	**PDIR**	**up**	11	stress response
Peroxiredoxin 2	**PRDX2**	**down**	11	stress response
Peroxiredoxin 3 isoform a precursor	**PRDX3**	**down**	11	stress response
Peroxiredoxin 5	**PRDX5**	**modulated**	11	stress response
Peroxiredoxin 6	**PRDX6**	**modulated**	11	stress response
STIP1	**STIP1**	**down**	11	stress response
Superoxide dismutase	**SOD2**	**modulated**	11	stress response
Thioredoxin peroxidise	**PRDX4**	**up**	11	stress response
UV excision repair protein RAD23	**RAD23B**	**up**	11	stress response

Table 8-11 Proteins modulated in the pathogen infected host cell involved in the stress response

8.3.12Cell motility

Protein name	Gene name	Overall change	Functional category	Category
Myristoylated alanine-rich C-kinase substrate	**MARCKS**	up	12	cell motility

Table 8-12 Proteins modulated in the pathogen infected host cell involved in cell motility

8.3.13Immune response

Protein name	Gene name	Overall change	Functional Category	Category
Annexin 1	**ANXA1**	up	13	immune response

Table 8-13 Proteins modulated in the pathogen infected host cell involved in the immune response

8.3.14Cell fate

Protein name	Gene name	Overall change	Functional category	Category
Beta galactosidase soluble lectin	**LGALS1**	down	14	cell fate
Porin 31HM	**VDAC1**	up	14	cell fate
Serum albumin precursor	**ALB**	modulated	14	cell fate
Tyrosine 3-monooxygenase	**YWHAH**	down	14	cell fate

Table 8-14 Proteins modulated in the pathogen infected host cell involved in cell fate

8.3.15Structural

Protein name	Gene name	Overall change	Functional category	Category
Actin	**ACTB**	down	16	structural
Actin gamma 1 propeptide	**ACTG1**	down	16	structural
Actinin-1 alpha	**ACTN1**	down	16	structural
Capping protein (actin filament)	**CAPZA1**	down	16	structural
Cofilin 1	**CFL1**	down	16	structural
Collagen type 7	**COL7A1**	down	16	structural
Cytoskeleton associated protein 4	**CKAP4**	up	16	structural
Desmin	**DES**	up	16	structural
Filamin	**FILIP1**	modulated	16	structural
Lamin A/C, isoform 2, 70kDa lamin	**LMNA**	down	16	structural
Lamin B1	**LMNB1**	up	16	structural
Lasp-1 protein	**LASP1**	modulated	16	structural
Microtubule associated protein 1B	**MAP1B**	down	16	structural
Moesin	**MSN**	up	16	structural
Myosin alkali light chain isoform1	**MYL4**	down	16	structural
Myosin, non muscle, heavy chain	**MYH9**	up	16	structural
Tropomyosin 2	**TPM2**	modulated	16	structural
Tubulin beta-1 chain	**TUBB1**	modulated	16	structural
Tubulin epsilon 5	**TUBE1**	down	16	structural
Vilin 2	**VIL2**	up	16	structural
Vimentin	**VIM**	modulated	16	structural

Table 8-15 Proteins modulated in the pathogen infected host cell with structural function

8.3.16Interaction with environment

Protein name	Gene name	Overall change	Functional category	Category
Annexin 5	**ANXA5**	modulated	17	interaction with environment

Table 8-16 Proteins modulated in the pathogen infected host cell involved in interaction with the environment

8.3.17Development

Protein name	Gene name	Overall change	Functional category	Category
Desmoyokin	**AHNAK**	down	18	development

Table 8-17 Proteins modulated in the pathogen infected host cell involved in development

8.3.18Interaction with cellular environment

Protein name	Gene name	Overall change	Functional category	Category
Galactose-specific lectin	**LGALS3**	up	15	interaction with the cellular environment
Myosin alpha heavy chain	**MYH6**	up	15	interaction with the cellular environment
Tropomyosin 1	**TPM1**	down	15	interaction with the cellular environment
Tropomyosin 4	**TPM4**	down	15	interaction with the cellular environment
Vinculin	**VCL**	modulated	15	interaction with the cellular environment

Table 8-18 Proteins modulated in the pathogen infected host cell involved in interaction with the cellular environment

8.3.19Unclassified

Protein name	Gene name	Overall change	Functional category	Category
4 and a half LIM domains 2 isoform 2	**FHL2**	down	19	unclassified
Coiled-coil-helix-coiled-coil-helix domain	**CHCHD3**	up	19	unclassified
Inter alpha (globulin) inhibitor H5	**ITIH5**	down	19	unclassified
Protein disulfide-isomerase A3 precursor	**GRP58**	up	19	unclassified
Protein kinase NYD-Sp9	**MFAPL3**	up	19	unclassified
snf7dc2	**snf7dc2**	down	19	unclassified
Thyroid hormone binding protein	**ETHE1**	down	19	unclassified

Table 8-19 Proteins modulated in the host cell whose function is unknown

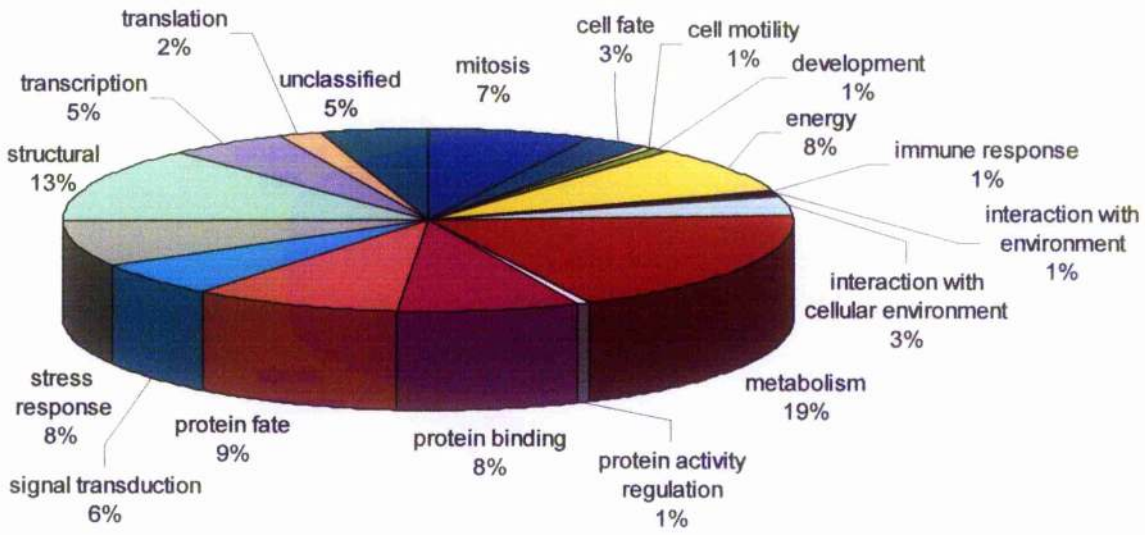


Figure 8-3 Pie chart showing the protein functions modulated due to infection. Metabolic and structural proteins are the most heavily modulated categories in the *Toxoplasma gondii* infected cell.

8.4 Discussion

The proteomic studies identified 157 proteins that were modulated in expression in *Toxoplasma gondii* infected cells. The proteomic techniques used in this study allowed profiling of approximately 3000 spots on a 2-DE gel. Of the potential 3000 protein spots, 157 modulated proteins were identified. These proteins were then classified into functional categories to determine the relevance of their modulation within the pathogen infected cell.

Since *Toxoplasma gondii* is an obligate intracellular parasite, factors within the host cell, e.g. essential nutrients not found extracellularly, are essential for parasite survival (Coppens, 2006). Using microarray and biochemical approaches it has previously been shown that *Toxoplasma gondii* infected HFFs show inhibition of apoptosis, remodelling of host cell organelle distribution, an up regulation of pro-inflammatory cytokines and up regulation of host glycolytic and mevalonate metabolic transcripts (Blader *et al.*, 2001). The results presented in this thesis represent the first attempt to profile the entire proteome of *Toxoplasma gondii* infected cells. The choice of host cell used in the experiments will be partially responsible for some of the protein responses observed. Many studies have investigated the macrophage response to infection with *Toxoplasma gondii* and other intracellular pathogens (Bogdan and Rollinghoff, 1999). Host macrophage responses to *Toxoplasma gondii* include inhibition of acidification of the PV, reduced macrophage survival after infection, induction of cytokines which inhibit or deactivate macrophages, inhibition of antigen presentation and T-cell stimulation by reduced MHC class II expression and reduced expression of co-stimulatory molecules.

Categorising the modulated proteins by function indicates that numerous cellular processes are modulated during *T.gondii* infection at the protein level. These included proteins involved in metabolism (19%); energy metabolism (8%); cell fate, including apoptosis (3%); structural (13%) and stress response (8%). Figure 8-3 shows the percentage of modulated proteins in each category. It is important to note that this type of functional annotation may be misleading, because some proteins may be involved in more than one process but have been grouped arbitrarily into a single class. Where this is the case, it has been documented.

Host proteins modulated during infection represent three functionally distinct classes; (1) proteins required for host defence, (2) proteins required for parasite growth and (3) proteins incidentally regulated as a consequence of modulation in the first two classes. Unless otherwise referenced, the protein function information was taken from

www.harvester.embl.de. The remainder of the discussion will focus on the most biologically relevant protein expression changes seen in the host cell along with a hypothesis of why protein modulation may have occurred.

8.4.1 Metabolism

Metabolism is the largest modulated category in this study indicating that metabolism is paramount in the infected cell for host or parasite survival. The thirty modulated proteins involved in host cell metabolism can be seen in Table 8-1. To what gain would modulation of host cell metabolism be to the parasite or host cell? A decrease in metabolism could be accounted for if the host cell stops dividing upon infection, therefore less metabolic intermediates and end-products would be required, or conversely, an increase in metabolism could be due to the parasite scavenging the host cell metabolic products resulting in the host cell having to increase metabolism to keep up with demand.

Several proteins involved in amino acid metabolism were modulated in the infected cell; cathepsin B (CTSB), cathepsin D (CTSD) and glutathione synthetase (GSS). *Toxoplasma* is known to be auxotrophic for tryptophan and probably requires other essential amino acids (Coppens and Joiner, 2001). Cathepsins are acid proteases active in intracellular protein breakdown and therefore release of amino acids. GSS is involved in glutathione biosynthesis and oxidative stress and expression is increased in the infected cell. Could modulation of the expression of these proteins be parasite driven and result in an increase in the supply of amino acids for the parasite, or could modulation of expression be host driven? If the host cell upon infection with *T.gondii* stops dividing then this could be driven by or result in a decrease of host functions including metabolism. Cathepsins have also been implicated in apoptosis which will be discussed later.

T. gondii is a purine auxotroph and must salvage adenine nucleobases, nucleosides or nucleotides as it is unable to convert guanine to adenine-based substituants. The parasite also takes up host hypoxanthine, xanthine, adenine, inosine and guanine. *T. gondii* can however synthesise pyrimidines denovo and is also able to scavenge host cell pyrimidines efficiently from the host cytoplasm, indicating the presence of specific uptake mechanisms (Sibley, 2002). The metabolic protein adenylate kinase 2 (AK2) is up regulated in the infected cell. AK2 is localized in the mitochondrial intermembrane space and it is speculated that AK2, once released from mitochondria, may affect the cellular levels of ADP and ATP (Kohler *et al*, 1999). AK contributes to this regulation of the homeostasis of the cellular adenine and guanine nucleotide pools by the phosphorylation of AMP to ADP,

which is further phosphorylated to ATP either by glycolytic enzymes or through oxidative phosphorylation (Kohler *et al*, 1999). Five heterogeneous nuclear ribonucleoprotein (hnrps) were found to be modulated in the infected cell. Four of these hnrps were functionally classed into nucleotide metabolism; HNRPC, HNRPF, HNRPH2 and HNRPA2/B1. Hnrps have not been previously implicated in the host-*Toxoplasma* response. Hnrps are described as a major group of nuclear RNA-binding proteins that function in transcription, RNA processing, mRNA translation, and turnover. Hnrps have also been implicated in apoptosis which will be discussed later. There is little information in the literature about the function of these proteins, however as five of these proteins were modulated post infection they perhaps play an important role in the host-pathogen response.

Six proteins involved in carbohydrate metabolism were modulated in the infected cell; aldose reductase (AKR1B1), aldehyde dehydrogenase 1A3 (ALDH1A3), aldehyde dehydrogenase X, mitochondrial precursor (ALDH1B1), hexosaminidase B (HEXB), phosphoenolpyruvate carboxykinase (PCK1) and 6-phosphogluconolactonase (PGLS). Carbohydrate metabolism may be modulated as the cell (1) is no longer carrying out its usual functions therefore carbohydrate is not needed, (2) the parasite requires host cell carbohydrate metabolism to be up regulated to aid its survival or (3) the stress exerted on the host cell due to the presence of the replicating parasite may cause an increase in this host cell process. PCK1 is down regulated in the infected cell. This gene is a main control point for the regulation of gluconeogenesis (production of glucose). The cytosolic enzyme encoded by this gene, along with GTP, catalyzes the formation of phosphoenolpyruvate from oxaloacetate, with the release of carbon dioxide and GDP. This enzyme is specific to gluconeogensis, many of the glycolytic modulated proteins could also be involved in this process and this is discussed later.

AKR1B1 is up regulated in the infected cell. This aldose reductase participates in glucose metabolism and osmoregulation and is supposed to play a protective role against toxic aldehydes derived from lipid peroxidation and steroidogenesis that could affect cell growth/differentiation when accumulated. As will be discussed, host cell steroid metabolism is increased in the infected cell therefore an increase of AKR1B1 fits in with a pro-host response, protecting the cells from the increase in toxic products produced. AKR1B1 is the first enzyme of the polyol pathway of sugar metabolism (Lefrancois-Martinez *et al.*, 2004).

ALDH1B1 is down regulated and ALDH1A3 is up regulated in the infected cell. Aldehydes are involved in the metabolism of corticosteroids, biogenic amines, neurotransmitters, and lipid peroxidation. Aldehyde dehydrogenase is the second enzyme of the major oxidative pathway of alcohol metabolism. Aldehyde dehydrogenase isozymes are thought to play a major role in the detoxification of aldehydes generated by alcohol metabolism and lipid peroxidation. Modulation of these proteins again fits in with the pro-host response.

T. gondii cannot synthesise selected phospholipids de novo or synthesise sterols de novo via the mevalonate pathway, yet cholesterol is the main sterol in parasite membranes and these lipids are scavenged from the host through an elaborate modification of the host cell. In addition to purine salvage, the PV membrane is responsible for recruiting host cell mitochondria and ER immediately following invasion. These organelles form an intimate association with the PVM, covering almost 75% of the vacuolar membrane by 4h post infection and it has been proposed that this parasite driven recruitment of host ER and mitochondria may serve to provide lipids for the PV by situating the host cell lipid biosynthesis machinery in close proximity to the PV (Black and Boothroyd, 2000). Morphological studies indicated that strips of ER and mitochondria come to lie along the cytoplasmic side of the vacuolar membrane, apparently hindering access of lysosomes (Mauel, 1996). However, recent studies by Coppens *et al* (2006) have revealed that host microtubule based invaginations allow for the delivery of host endo-lysosomes to inside the PV. Thus, *T. gondii* is dependent on actively intercepting low-density lipoprotein-(LDL-) derived cholesterol that has transited through host lysosomes and the cholesterol must then be delivered to and across the PVM (Coppens, 2006). Host lysosomes are transported to the PV along host microtubules in a *T. gondii*-driven process. It has been hypothesised that *T. gondii* induces a unique process of microtubule-based invaginations from the host cell cytoplasm into the vacuolar space for acquisition of nutrients internalized into the host cell via the host endocytic pathway (Coppens, 2006). Exogenous cholesterol is then rapidly delivered to the PV and is concentrated in the parasite. LDL-receptor-mediated endocytosis is specifically increased in infected cells compared with non infected cells (Coppens and Joiner, 2001). These changes in sterol metabolism may be a consequence of parasite manipulation of host cholesterol regulatory pathways, or the inability of the host cell to function normally under the stress of the parasite's assault, or a combination of these factors (Coppens, 2006).

Carbonyl reductase (CBR1), high density lipoprotein binding protein (HDLBP), cargo selection protein TIP47 (M6PBRP1) and Irp protein (LRP1) which are all involved in lipid

and sterol metabolism are modulated in the infected cell. As it is known that the parasite scavenges host cell lipids and cholesterol, an obvious hypothesis is that host cell proteins involved in lipid and sterol metabolism would be increased. However, the proteomic studies showed that these four proteins were all down regulated in expression in the parasite infected cell. Perhaps this down regulation is a mechanism of host defence against the parasite, or merely due to the cell shutting down functions. One modulated protein which may be a pro-parasite driven response is HDLBP which plays a role in sterol metabolism and one of its functions may be to protect cells from over accumulation of cholesterol by removing it. The parasite will want as much cholesterol as possible in the host cell so perhaps the parasite inhibits expression of this protein to avoid cholesterol removal from the host cell.

Apicomplexan parasites synthesise fatty acids (FAs) de novo in the apicoplast and can also accumulate host FAs. FAs are building blocks of phospholipids and glycolipids (membranes) and many proteins are modified by the covalent attachment of FAs, which targets them to membranes. The proposed ability to transfer lipids directly from the host cell to the parasite at sites of PVM-organelle association would implicate appropriate transport activities in the PVM, in order to support the needs of intravacuolar parasites (Sinai *et al.*, 1997). Acetyl coenzyme A acyltransferase 2 (ACAT1) and enoyl-coenzyme A hydratase 1 (ECHS1) are modulated in the infected cell and are early components of the fatty acid oxidation pathway. ACAT1 plays an essential role in breaking down proteins from the diet therefore may provide AA for the parasite. Specifically, the enzyme is needed to process isoleucine (an amino acid that is a building block of many proteins). This enzyme is also involved in processing ketones which are produced when fat is broken down in the body for energy. It facilitates the reversible production of two molecules of acetyl-CoA from a molecule called acetoacetyl-CoA. Additional chemical reactions convert these compounds into molecules that are later used for energy (http://ghr.nlm.nih.gov/gene=acat1). This fits in with the increase in energy metabolism seen in the infected cell.

8.4.2 Energy

In general, the glycolytic proteins identified in the *Toxoplasma gondii* infected cells were increased in expression (Table 8-2). Several of these changes have previously been documented by Blader *et al* (2001) and Coppens *et al* (2000). The proteomic studies in this thesis found a glycolytic protein, pyruvate kinase, not previously thought to be modulated in expression in the parasite infected cell.

Glycolysis is a ten step metabolic pathway that converts glucose to pyruvate with a net gain of two molecules ATP per glucose molecule. Oxidative phosphorylation takes place in the inner mitochondrial membrane in human cells, and in the process in which ATP is formed as a result of the transfer of electrons from NADH or $FADH_2$ to O_2 by a series of electron carriers. When glucose is completely oxidised to CO_2 and H_2, 30 molecules of ATP are formed, 26 during oxidative phosphorylation (Berg *et al*, 5th Edition, p 491).

The proteomic analysis identified six enzymes involved in the glycolytic chain in mammalian cells are modulated in expression due to the presence of the intracellular parasite (Figure 8-4); aldolase, enolase 1, glyceraldehyde phosphate dehydrogenase (GAPD), phosphoglycerate kinase, pyruvate kinase and triose phosphate isomerase. Aldolase and GAPD were significantly up regulated, ≥ 2fold in the DIGE experiment. It has previously been documented (Blader *et al*, 2001) that it is not thought that *Toxoplasma* induces up regulation of oxidative phosphorylation. The proteomic results show that three enzymes, near the end of the oxidative phosphorylation pathway (ATP5D, COX6B and PPi), are modulated in the *Toxoplasma gondii* infected cell and this pathway can be seen in Figure 8-5.

The increased expression of host cell glycolytic proteins may be due in part to the fact that *Toxoplasma* is auxotrophic for purines (as mentioned in the metabolism section). The PV allows passive diffusion of host cell ATP from the host cell cytoplasm into the vacuolar lumen of the PV. Here, the ATP is hydrolysed into its constituent ADP and AMP. The AMP is cleaved to adenosine and transported into the parasite (Coppens and Joiner, 2001). To compensate for the parasite scavenging the host cells own glycolytic reserves, the cell may have to increase the glycolytic pathway as a mechanism of survival. The redistribution of host cell mitochondria to around the PV allows for easier access by the parasite to the end products of glycolysis and oxidative phosphorylation.

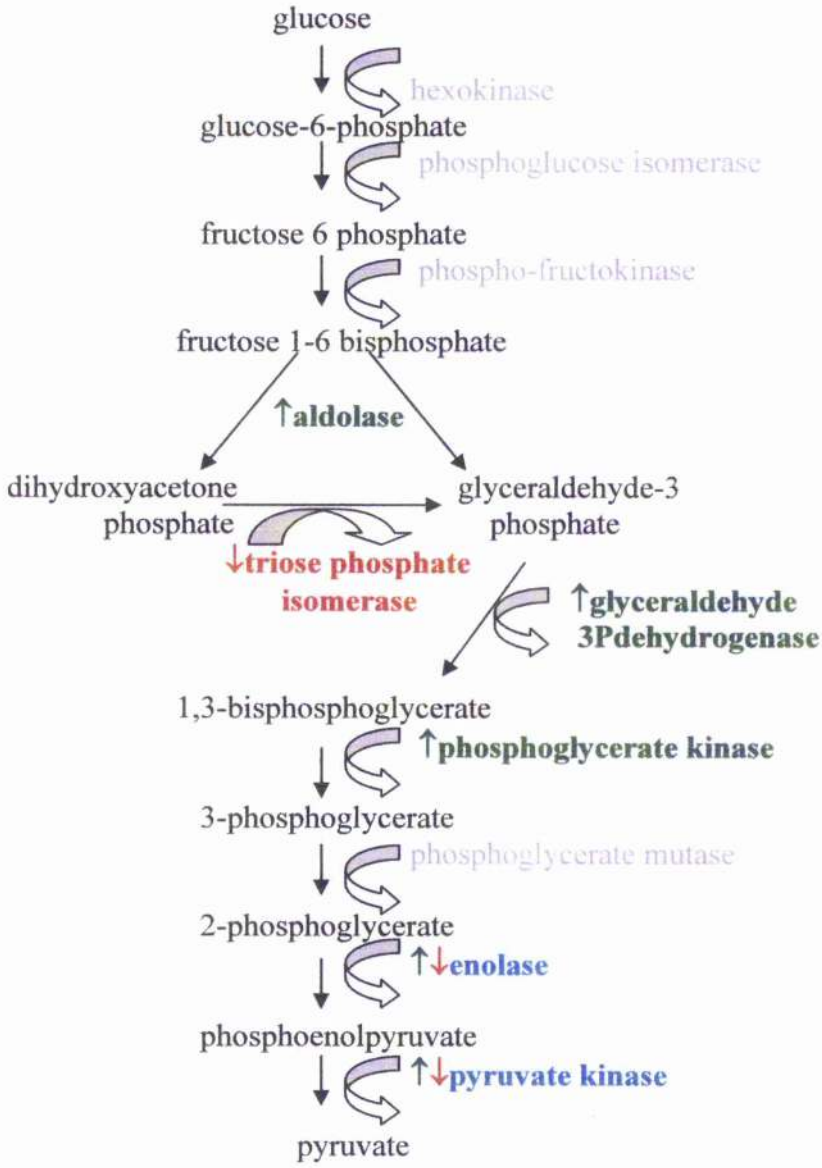


Figure 8-4 Proteins modulated in the glycolytic pathway. ↑ indicates up regulation in protein expression in the pathogen infected cell. ↓ indicates down regulation in protein expression in the pathogen infected cell. Six of the ten enzymes involved in the glycolytic pathway are modulated in the parasite infected cell.

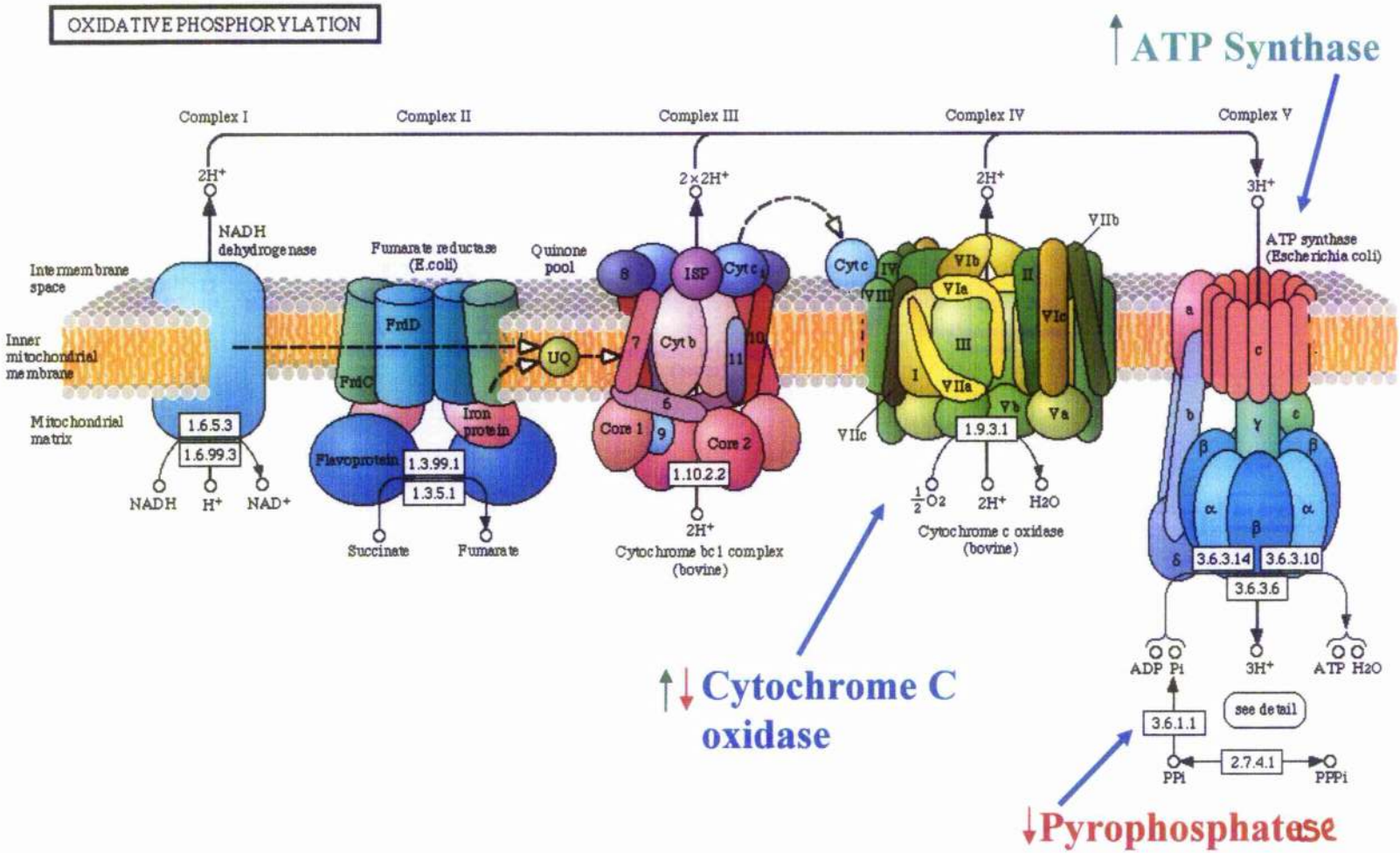


Figure 8-5 Proteins modulated in the oxidative phosphorylation pathway (pathway taken from www.kegg.com). Three of the final steps of the oxidative phosphorylation pathway are modulated in the *Toxoplasma* infected cell.

8.4.3 Cell cycle

Ten of the modulated proteins function in mitosis and cell proliferation. Apoptosis and the cell cycle are intimately linked and both processes are controlled by many of the same proteins (Vermeulen *et al.*, 2003). Apoptosis is known to be inhibited in the *Toxoplasma gondii* infected cell (Blader *et al.*, 2001) therefore the cell cycle may be modulated accordingly. Perhaps inhibition of the apoptotic proteins also results in the inhibition of host cell cycle. We hypothesise that the host cell shuts down cell division upon infection. This may be due to (1) a bystander effect due to the inhibition of proteins common to both apoptosis and cell proliferation or (2) the parasite stimulates this inhibition to stop the host cell using up energy and cellular intermediates which the parasite wants for its own survival.

Some parasites are known to utilise host cell division to their advantage including the intracellular apicomplexan parasite *Theileria*. This parasite infects lymphocytes causing uncontrolled proliferation of these host cells and the parasite divides in synchrony with the host cell ensuring dissemination of the parasite throughout the lymphocyte cell population (Rocchi *et al.*, 2006). *Toxoplasma* cannot use cell division for this purpose as it is bound by the parasitophorous vacuole and therefore cannot be passed through the body by the mechanism used by *Theileria*.

The stage of host cell cycle determines whether *Toxoplasma gondii* can invade or not. Initial attachment of *Toxoplasma* tachyzoites to host cells is an important event in the life cycle of the parasite and hence critical in the pathogenesis of this infection. Parasite attachment increases as the synchronized host cells proceed from the G1 phase to the mid-S phase and then decreases as the cells enter the G2-M boundary. *Toxoplasma* tachyzoites bind specifically to a host cell receptor which is up regulated in the mid-S phase of the cell cycle (Grimwood *et al.*, 1996).

Several of the modulated cell cycle proteins were of interest in terms of the host-pathogen response. Dynactin 2 expression is modulated in the infected cell. This mitotic protein is involved in spindle organisation and chromosome alignment. Another of its functions is as a component of a large macromolecular complex required for the cytoplasmic dynein-driven movement of organelles along microtubules (NCBI). This protein may therefore be involved in the movement of the host mitochondria, ER and lysosomes in the infected cell. http://www.jcb.org/cgi/content/abstract/132/4/617.

Ubiquilin 2, which is involved in spindle pole body duplication, the cytokenetic protein NEDD5, lamin B2 which is involved in nuclear envelope assembly and disassembly during cell division and caldesmon and plays an essential role during cellular mitosis and receptor capping, are all modulated in the infected cell. Also modulated is the expression of cell proliferation proteins including the anti-oxidant protein peroxiredoxin 1, calgizzarin, a negative regulator of cell proliferation, and prohibitin. These results indicate that the cell cycle is indeed being modulated in the host-pathogen response. Whether this response is host or parasite driven remains unclear.

8.4.4 Transcription and translation

Due to the high degree of modulation seen in the host cell genome and proteome as detected by the MA and proteomic studies respectively, it would be expected that there will be many changes in the expression of proteins involved in the host cell transcription and translation functional categories.

Six transcriptional and three translational proteins were identified as modulated in expression in the infected cell with the general trend being up regulation. The transcriptional regulatory proteins brain abundant membrane attached signal protein (BASP1), presenelin binding protein, CLE7 (C14orf166), PIAS-Ny protein (PIAS2) and zinc finger protein were all up regulated. BASP1 is a transcriptional regulator of genes involved in cell division regulation (Carpenter *et al.*, 2005) and PIAS2 is a transcriptional co-regulator which is involved primarily in gene silencing. LMNB2 whose primary function is involved in cell cycle is also a regulator of transcription and is also up regulated in the infected cell. Heterogeneous nuclear ribonucleoprotein K isoform a (HNRPK) is down regulated in the infected cell. This transcriptional-inducer and translational-regulator protein is likely to play a role in the nuclear metabolism of hnRNAs. This protein is also a repressor of a key transcription factor, C/EBP, involved in the induction of genes during the acute phase of the immune response (Miau *et al.*, 1998). Therefore HNRPK expression represses induction of IR genes. HNRPK is down regulated in the infected cell therefore allows the induction of IR genes in a pro-host response manner.

The three proteins involved in translation in this study were all up regulated in expression in the parasite infected cell. This response could be host-driven, to increase protein synthesis to replace the intermediates scavenged by the parasite, or parasite driven to increase the available cellular intermediates. WBSCR1 is up regulated in the infected cell and its function is to stimulate protein translation by stimulating the activities of several

eukaryotic initiation factors during the steps involving mRNA binding and utilization in the initiation of protein biosynthesis. Elongation factor EEFIA2 is involved in protein biosynthesis and is up regulated in the infected cell. Ubiquitin A-52 is up regulated in the infected cell and is involved in the ATP-dependent selective degradation of cellular proteins, the maintenance of chromatin structure, the regulation of gene expression, the stress response and ribosome biogenesis. HSPA6, which is firstly classified into protein fate, is modulated in the infected cell and is involved in regulation of translational initiation.

8.4.5 Protein fate

The protein fate functional category encompasses many of the heat shock protein (HSPs), also called stress proteins, which are a group of proteins present in all cells. HSPs are present in cells under normal conditions and make up 5-10% of the total cellular protein content. Upon the cell being stressed (e.g.heat/cold/oxygen deprivation) this can increase to 15% (Pockley, 2001). HSPs act like 'chaperones', making sure that the cell's proteins are in the right shape and in the right place at the right time. HSPs shuttle proteins from one compartment to another inside the cell and facilitate the degradation of selected proteins by the ubiquitin/proteasome system. In cells, damaged or excess proteins are tagged with molecules called ubiquitin. Ubiquitin serves as a signal to move these unwanted proteins into specialized structures known as proteasomes, where the proteins are degraded. The ubiquitin-proteasome system acts as the cell's quality control system by disposing of damaged, misshapen, and excess proteins (Garrido *et al*, 2003). HSPs are also believed to play a role in the presentation of peptides on the cell surface to help the immune system recognise diseased cells (http://www.antigenics.com/products/tech/hsp/). Therefore it is expected that there will be modulation of proteins in this category in the stressed infected cell as; (1) these proteins play a role in the immune response, therefore a pro-host response, (2) the presence of the parasite and the scavenging of host cell intermediates by the parasite will undoubtedly stress the cell and *Toxoplasma* is known to induce hypoxia in the host cell (Spear *et al.*, 2006), (3) increased protein synthesis will result in an increase in proteins needing folded and transported within the cell and (4) increased waste products will need degraded.

Seven HSP proteins categorised into protein fate were modulated in the parasite infected cell with a general trend of up regulation. These were; BiP protein (HSP70), Chaperonin (CCT2), Chaperonin HSP60 (HSPD1), MTHSP75 (HSPA9B), HSPC054, HSPA6 and HSPA8 (another isoform of HSP70).

The up regulated HSP70 is an ATP-dependant molecular chaperone that assists the folding of newly synthesised polypeptides, the assembly of multi-protein complexes and the transport of proteins across cellular membranes. Perhaps expression of this protein is increased due to the increased protein transport in the infected cell due to parasite scavenging (Garrido *et al*, 2003).

CCT2 and HSPD1 are modulated in the infected cell and are implicated in mitochondrial protein import and macromolecular assembly. These proteins may facilitate the correct folding of imported proteins and are generated under stress conditions in the mitochondrial matrix. The encoded mitochondrial protein may function as a signalling molecule in the innate immune system, assisting the folding of proteins upon ATP hydrolysis, folding of actin and tubulin and may play a role in mitotic spindle formation (Pockley, 2001). All of these functions have been discussed as of importance in the parasite infected cell, in particular, tubulin modifications (discussed in the structural category) and ATP hydrolysis.

Chaperones are proteins whose function is to assist other proteins in achieving proper folding and the chaperones cyclophilin A and tumour rejection antigen 1 are both down regulated in the infected cell. Cyclophilin A may play a role in cyclosporin A-mediated immunosuppression (harvester). Therefore down regulation of this protein would result in an inhibition of immunosupression and therefore could be a pro-host response. Immunosupression in the host results in uncontrolled proliferation of the parasite leading to host death.

PDI related protein 5, CDC37 cell division cycle 37 homolog and calumenin are up regulated in the infected cell. These proteins all function in protein folding and sorting.

8.4.6 Protein binding

Twelve proteins involved in protein binding were modulated in the parasite infected cell. Of these twelve, nine were down regulated in expression. The proteins included in this category are proteins with binding function or cofactor requirement (structural or catalytic) whose main function is the binding of specific ligands. Of interest are transgelin 1 and 2 which are down regulated in the infected cell. These actin cross-linking proteins are also involved in calcium interactions and contractile properties of the cell that may contribute to replicative senescence. Tropomodulin (TMOD3) is down regulated in the infected cell and is involved in actin and tropomyosin binding and this may link in with the changes seen in the structural category. UMPCMPK is down regulated in the infected cell and furnishes the

cell with pyrimidines and binds ATP. The parasite can synthesis its own pyrimidines de novo therefore there is no need for increased host cell pyrimidine biosynthesis; this protein must have a secondary function which is of importance in the infected cell.

8.4.7 Cellular transport

MIPS defines this category as the directed movement of proteins in a cell, including the movement of proteins between specific compartments or structures, such as organelles of a eukaryotic cell. As previously mentioned, host cell mitochondria and ER are recruited to around the PV in the infected cell and lysosomes are directed to the PV. The movement of these organelles is microtubule driven. Tubulin alpha (TUBA1) is modulated in the infected cell and is a structural constituent of microtubules and involved microtubule based movement. TUBA1 may therefore be involved in the movement of host cell mitochondria, ER and lysosomes.

The transport of specific molecules across lipid membranes is an essential function of all living organisms and a large number of specific transporters have evolved to carry out this function. The largest transporter gene family is the ATP-binding cassette transporter superfamily. These proteins translocate a wide variety of substrates including sugars, amino acids, metal ions, peptides, and proteins, and a large number of hydrophobic compounds and metabolites across extra- and intracellular membranes (Dean, 2001). ATP binding cassette transporter A13 was up regulated in the infected cell. This anion transporter is a key gatekeeper influencing intracellular cholesterol transport. This protein may be up regulated in the pathogen infected cell as part of the mechanism used to scavenge host cell lipids by the parasite.

Two isoforms of the chloride intracellular ion channel were down regulated in the parasite infected cell; CLIC1 and CLIC4. Chloride channels are a diverse group of proteins that regulate fundamental cellular processes including stabilization of cell membrane potential, transepithelial transport, maintenance of intracellular pH, and regulation of cell volume. Why these processes are decreased in the infected cell is unclear.

8.4.8 Signal transduction

We hypothesised that the infected host cell would show many changes in signal transduction (involved in controlling the cells behaviour). This was indeed the case and nine proteins involved in signal transduction were found to be modulated in the parasite

infected cell. Of interest was up regulation of annexin 2 which regulates cellular growth and signal transduction pathways. Ran-binding protein and reticulocalbin 1 precursor also involved in signal transduction were seen to be down regulated in the infected cell. Cardiotrophin like cytokine was modulated in the infected cell and is involved in cytokine and chemokine mediated signalling. Phospholipase C alpha was up regulated in the infected cell and is an activator of secondary messenger pathways. Zyxin is down regulated in the infected cell and was identified many times. This protein is thought to be a component of a signal transduction pathway that mediates adhesion-stimulated changes in gene expression.

8.4.9 Immune and stress response

Super oxide dismutase (SOD2) is modulated in the infected cell. This protein responds to oxidative stress and protects against oxidative damage by reactive oxygen species. HSP27 is modulated in the infected cell and has been shown to increase the anti-oxidant defence of cells by decreasing reactive oxygen species cell content, increasing the level of reduced glutathione (GSH) and neutralising the toxic effects of oxidised proteins (Garrido *et al*, 2003). PDI related protein 5 is up regulated in the infected cell. This protein has oxidoreductase activity and catalyzes the rearrangement of disulfide bonds in proteins.

Six isoforms of the anti-oxidant protein peroxiredoxin (PRDX) were identified as modulated in the infected host cell – PRDX1, PRDX2, PRDX3, PRDX4, PRDX5 and PRDX6. The peroxiredoxins are a family of antioxidant enzymes, which reduce hydrogen peroxide and alkyl hydroperoxides, are a major cellular response to oxidative stress and are clearly important in either the pro parasite or pro host response in this dynamic system. PRDXs are widely distributed throughout the cell as different isoforms indicating that the peroxiredoxin enzyme plays an essential role in the anti-oxidant defence of the cell. Little is known at the protein level in mammalian cells about the response to challenges by oxidative stress because most studies are carried out at the RNA level.

PRDX1 is down regulated in the host cell upon infection and has previously been mentioned in the cell cycle category as it is involved in cell proliferation, however this protein also plays an anti-oxidant role. The encoded protein of PRDX1 may play an antioxidant protective role in cells, and may contribute to the antiviral activity of CD8(+) T-cells.

Peroxiredoxin 2 and 3 are also down regulated in the infected cell. This protein is involved in redox regulation of the cell and may play an important role in eliminating peroxides generated during metabolism which is extensively modified in the infected cell. PRDX2 enhances natural killer (NK) cells activity.

Thioredoxin peroxidase (PRDX4) is up regulated in the infected cell. This is an antioxidant enzyme that plays a regulatory role in the activation of the transcription factor NF-κB therefore is linked to apoptosis (Jin *et al*, 1997). Thioredoxin peroxidise is a novel inhibitor of apoptosis and is involved in protecting cells from apoptosis in response to reactive oxygen species (Zhang *et al*., 1997).

Peroxiredoxin 5 is modulated in the infected cell. The encoded protein may play an antioxidant protective role in different tissues under normal conditions and during inflammatory processes. Peroxiredoxin 6 (PRDX6) is also modulated in the infected cell and may play a role in the regulation of phospholipid turnover as well as in protection against oxidative injury. PRDX6 may be modulated due to the scavenging of host cell lipids by the parasite.

Stress response protein (STIP1) is modulated in the infected cell. UV excision repair protein RAD23 homologB (RAD23B) is up regulated in the infected cell and is involved in DNA excision repair and may play a part in DNA damage recognition.

An immune response protein up regulated in the infected cell is Annexin 1 which is thought to have anti-inflammatory activity, therefore up regulation of this protein would fall into the pro-host category. Annexin 5 is also modulated in the infected cell. Annexin 5 is a phospholipase A2 and protein kinase C inhibitory protein with calcium channel activity and a potential role in cellular signal transduction, inflammation, growth and differentiation (Alfonso *et al*, 2005).

8.4.9.1 Response to pest/pathogen/parasite

An interesting gene ontology category that lies within the MIPS category of cell rescue/defence/stress response is response to pest/pathogen/parasite. Proteins assigned to this category have previously been shown to be directly involved in the host response to pest/pathogen or parasite.

The following six proteins were linked to the gene ontology category GO:0009613 : response to pest, pathogen or parasite; ANXA1, CHIT1, CLC, PRDX5, SERPINA and SOD2. The functions of these proteins and their hypothetical role in the *T.gondii* infected cell are discussed elsewhere in this chapter.

8.4.10 Cell motility

Only one protein involved in cell motility was identified as modulated in expression in the infected cell; myristoylated alanine-rich C-kinase substrate (MARCKS). This protein is up regulated in the infected cell. The protein encoded by this gene is a substrate for protein kinase C and is an actin filament cross linking protein which anchors the actin cytoskeleton to the PM (Larsson, 2006). The protein is thought to be involved in cell motility, phagocytosis, membrane trafficking and mitogenesis. The corresponding gene was identified as up regulated 10-fold in the *T.gondii* infected cells in the transcriptional study by Gail *et al* (2001).

8.4.11 Structural

The second largest number of differentially expressed proteins had structural function. As previously mentioned, infection with *Toxoplasma gondii* causes extensive remodelling of the host cell, including recruitment of host mitochondria and ER to around the PV therefore it would be expected that there will be many differences in protein expression in this category due to infection. The distribution of both mitochondria and the ER is governed by the microtubule cytoskeleton. Early post infection with *T.gondii*, nearly 75% of the PVM is associated with organelles. It is likely that the primary function for associated mitochondria and ER is to provide essential nutrients, lipids and products of intermediary metabolism to the parasite (Sinai *et al.*, 1997).

Three types of filaments make up the cytoskeleton; (1) tubulins (e.g. β-tubulin), (2) intermediate filaments (vimentin) and (3) microfilaments (vinculin, filamin, β-actin) (Predic *et al.*, 2002) and cytoskeletal interactions are known to be important in a wide variety of cellular processes including signalling systems, cell-cell interaction and mitosis.

Microtubules (MTs) are cytoskeletal fibers that are formed by polymerization of α,β-tubulin monomers and exhibit structural and functional polarity. They are important components of cilia, flagella, the mitotic spindle, and other cellular structures. Microtubules have been visualized extending from the host cell into the vacuolar space,

inducing the delivery of host lysosomes containing cholesterol and other nutrients into the *Toxoplasma* vacuole. The host microtubule organising centre has been shown to move from its usual position on the nuclear envelope to the PV. MTs were specifically concentrated around the PV in contrast to the MTs in uninfected cells which formed a framework around the nucleus. Host MT remodelling is specific to *Toxoplasma gondii* as mammalian cells harbouring other apicomplexan parasites (e.g. Neospora caninum) did not show any alterations in the MT cytoskeleton organisation under parasite assault. The host MT bundles clearly co-distributed with vimentin intermediate filaments around the PV while the distribution of host actin microfilaments remained unchanged (Coppens *et al.* 2006). The parasite essentially hijacks the microtubule organizing centre of the cell, which controls cell architecture, movement and division. The parasite uses the microtubule organizing centre to create pathways into the cell, inducing a unique process of microtubule-based invaginations from the host cell cytoplasm into the vacuolar space, in order to attract lysosomes closer to its vacuole. The parasite then secretes a protein that cuts off the lysosomes pathway into the cell, trapping the lysosome inside the vacuole. Lipids, amino acids, sugars and metals, which are food for the parasite, then diffuse through small pores in the lysosomal membrane and nourish the parasite (Coppens, 2006). Tubulin beta-1 chain is modulated in the infected cell and is a structural constituent of microtubules and involved in microtubule based movement (Predic *et al.*, 2002). Tubulin epsilon 5, a structural constituent of the cytoplasm and microtubule associated protein 1B is down regulated in the infected cell.

Intermediate filaments are cytoskeletal fibers that are formed by polymerization of monomeric globular actin; also called *actin filaments*. Microfilaments are cytoskeletal fibers formed by polymerization of several classes of cell-specific subunit proteins including keratins, lamins, and vimentin. They function as important structural components of cells. Microfilaments play an important role in muscle contraction, cytokinesis, cell movement, and other cellular functions and structures.

Desmin are class-iii intermediate filaments and were up regulated in the infected cell. Three lamin proteins were modulated in the infected cell - lamin A/C, lamin B1 and lamin B2. Lamin proteins are type V intermediate proteins and are thought to be involved in nuclear stability, chromatin structure and gene expression. Lamins are components of the nuclear lamina, a fibrous layer on the nucleoplasmic side of the inner nuclear membrane, which is thought to provide a framework for the nuclear envelope and may also interact with chromatin. These are considered to be nucleoskeletal IF proteins that are determinants of nuclear size, shape, mechanical integrity, and positioning of nuclear pores. They also

play important roles in DNA replication, transcription and the disassembly of the nucleus in mitosis and the reassembly of the nucleus in daughter cells. Vimentin is a class III intermediate filament component of the cytoplasm and is heavily modulated in the infected cell. Vimentin is phosphorylated and present in several isoforms (Alfonso *et al*, 2005). In a normal non-infected cell, vimentin is distributed throughout. However, in *T.gondii* infected cells, vimentin is reorganised to around the parasitophorous vacuole where it is thought it acts to anchor the PV to the host cell nuclear envelope (by the network of intermediate filaments), and may also help to bring metabolites and nutrients to the parasites (Coppens and Joiner, 2001(Mauel, 1996). Vinculin (VCL) is modulated in the infected cell. This protein is involved in cell adhesion and may be involved in the attachment of the actin-based microfilaments to the plasma membrane.

Actins are highly conserved proteins that are involved in cell motility, structure and integrity. Four actin proteins were down regulated in the infected cell; actin beta, actin gamma 1 propeptide, actinin-1 alpha and capping protein. Actinin-1 alpha is an F-actin cross-linking protein which is thought to anchor actin to a variety of intracellular structures. Capping protein is a member of the F-actin capping protein alpha subunit family. The protein regulates growth of the actin filament by capping the barbed end of growing actin filaments. Lasp-1 protein functions as an actin-binding protein and possibly in cytoskeletal organization. Filamin is modulated in the infected cell and promotes orthogonal branching of actin filaments and links actin filaments to membrane glycoproteins. Filamin anchors various transmembrane proteins to the actin cytoskeleton and serves as a scaffold for a wide range of cytoplasmic signalling proteins. Moesin is up regulated in the infected cell. Moesin appears to function as a cross-linker between plasma membranes and actin-based cytoskeletons. Moesin is localized to filopodia and other membranous protrusions that are important for cell-cell recognition and signalling and for cell movement. Myosin alkali light chain isoform1 is down regulated in the infected cell. Myosin is a hexameric ATPase cellular motor protein. Myosin, non muscle, heavy chain is up regulated in the infected cell. Cellular myosin appears to play a role in cytokinesis, cell shape, and specialized functions such as secretion and capping. Tropomyosin 2 is modulated in the infected cell. This protein can differentially alter actin filament organization, cell size, and shape.

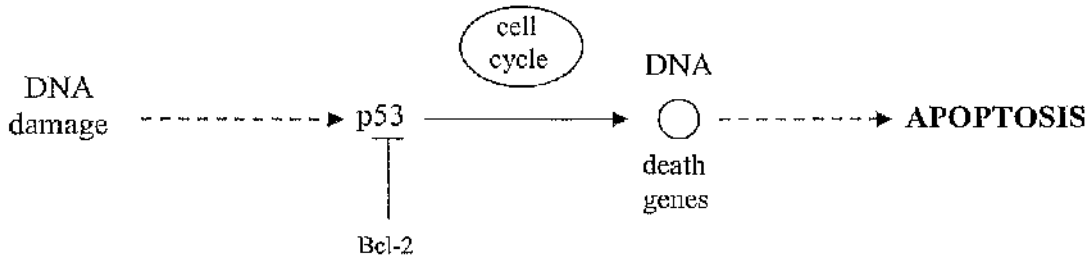
Cofilin 1 (CFL1) and collagen type 7 (COL7A1) are both down regulated in the infected cell and involved in cell growth and/or maintenance. Cytoskeleton associated protein 4 (CKAP4) is up regulated in the infected cell. Vilin 2 (VIL2) is up regulated in the infected

cell and it thought to be involved in connections of major cytoskeletal structures to the plasma membrane.

8.4.12 Cell fate and apoptosis

Programmed cell death is involved in development, aging and homeostasis and plays an essential role as a defence mechanism against pathogens (Hasnain *et al.*, 2003). Parasites have evolved diverse strategies to modulate host cell apoptosis thereby interfering with the host IR, facilitating parasite survival and promoting dissemination within the host (Dlugonska, 2004). Two apoptotic pathways exist which both involve the activation of caspases, the central effectors of apoptosis. The first pathway is termed extrinsic and is triggered by an extracellular stimulus through plasma membrane proteins of the TNF receptor family known as death receptors and leads to the direct activation of caspases, starting with caspase-8. The second is the intrinsic pathway which involves intracellular stress signals that elicit the production or activation of pro-apoptotic molecules (caspase 9), which converge on the mitochondria to trigger their permeabilsation (Concannon *et al.*, 2003).

Intrinsic pathway



The mitochondrial pathway of apoptosis is of special interest with regard to *T.gondii* because of the intimate association of the host mitochondria with the PV (Dlugonska, 2004). *Toxoplasma* has previously been shown to manipulate host cell apoptosis (Dlugonska, 2004; Goebel *et al.*, 2001; Hasnain *et al.*, 2003; Luder *et al.*, 2001; Molestina *et al.*, 2003; Sinai *et al.*, 2004). Blocking host cell apoptosis may give several advantages to the intracellular parasite. (1) Inhibition of programmed cell death ensures a safe environment for the parasite to reside and replicate within, (2) *T.gondii* is auxotrophic for purines, several amino acids and cholesterol, blocking host cell apoptosis ensures a ready supply of these essential metabolites and (3) it is also thought that inhibition of apoptosis is important for stage conversion of tachyzoites to bradyzoites and therefore establishment of chronic infection (Dlugonska, 2004).

Mitochondria act as control centres regulating apoptosis. A pore called the mitochondrial permeability transition pore (mtPTP) forms in damaged mitochondria and this pore consists of VDAC (the adenine nucleotide translocator) and several other mitochondrial proteins. VDAC plays the central role in mitochondrial mediated apoptosis (Berg *et al*, 5th Edition, p 491). Our studies showed that the cell fate protein VDAC is up regulated after infection thus indicating mitochondrial damage in the host cell. VDAC works in concert with the adenine nucleotide translocator (ANT), and as previously discussed, the parasite is auxotrophic for adenine and has to scavenge host cell adenine intermediates. It is possible to hypothesis that up regulation of this protein may result in an increase in adenine translocation in the host cell, or, the host cell may be trying to defend itself by up regulating expression of this protein and therefore activating apoptosis and the parasite is somehow inhibiting apoptosis downstream.

The metabolic protein LRP1 plays a role in modulating mitogenic signalling, cell adhesion, cell proliferation and apoptosis (Benes *et al.*, 2003). It is down regulated in the infected cell as is LMNA, a specific substrate of caspase-6 and therefore is involved in the apoptotic response (extrinsic pathway) (http://www.nature.com/emboj/journal/v21/n8/pdf/7594416a.pdf) again fitting into apoptosis inhibition.

The cell fate protein YWHAH (14-3-3 protein) is down regulated in the infected cell. The primary function of 14-3-3 proteins is to inhibit apoptosis. It is a critical anti-apoptosis factor (Xing *et al.*, 2000). 14-3-3 proteins also participate in phosphorylation dependant protein-protein interactions that control progression through the cell cycle, initiation and maintenance of DNA damage checkpoints, activation of MAPK, and co-ordination of integrin signalling and cytoskeletal dynamics (Wilker and Yaffe, 2004). Down regulation in expression of this protein therefore would not lead to inhibition of apoptosis.

The metabolic protein adenylate kinase 2 (AK2) is localized in the mitochondrial intermembrane space and may play a role in apoptosis. A study by Kohler *et al* showed that adenylate kinase 2 is released from the mitochondrial intermembrane space during apoptosis (Kohler *et al.*, 1999). AK is a ubiquitous enzyme that contributes to the regulation of the homeostasis of the cellular adenine and guanine nucleotide pools. The function of adenylate kinases in normal cells is restricted to the phosphorylation of AMP to ADP, which is further phosphorylated to ATP either by glycolytic enzymes or through oxidative phosphorylation. It is speculated that AK2, once released from mitochondria,

may affect the levels of ADP and ATP and thereby contribute to the apoptotic process (Kohler *et al*, 1999).

The electron transport protein TXNDC5 encodes a protein-disulfide isomerise and is up regulated in the infected cell. Its expression is induced by hypoxia and one role may be to protect hypoxic cells from apoptosis. Oxygen consumption by the parasite may lead to a localised hypoxic response in the host cell (Spear *et al.*, 2006) therefore this protein functions to protect the host cell and the parasite at the same time.

Three of the peroxiredoxins modulated post infection are implicated in apoptosis. Several transfection experiments have shown that the over expression of various peroxiredoxins is able to counteract several pro-apoptotic signals, thereby also indicating the importance of the oxidative status of cells in the onset of apoptosis (Rabilloud *et al.*, 2002). Peroxiredoxin 1 protein expression decreases after infection which agrees with decreased apoptosis in the infected cell. Peroxiredoxin 2 is also down regulated and may play an active role in resistance or susceptibility to tumour necrosis factor α-induced apoptosis. Thioredoxin peroxidise is up regulated and plays a regulatory role in the activation of the transcription factor NF-κB, a transcription factor which induces anti-apoptotic activity (Jin *et al.*, 1997).

Prohibitin (PHB) is modulated in the infected cell and has been implicated in apoptosis. In this study PHB was classified into the cell cycle category. The expression of PHB changes with senescence suggesting that the decline in prohibitin might be associated with the accumulation of damage from mitochondrial oxygen radicals. Prohibitin induces the transcriptional activity of p53 (component of the intrinsic apoptotic pathway) and is thought to disrupt mitochondrial function resulting in apoptosis (Dlugonska, 2004).

PIAS2 is up regulated post invasion in the infected cell and plays a crucial role as a transcriptional co-regulator of the p53 pathway, a component of the intrinsic apoptotic pathway. Calgizzarin has been implicated in apoptosis and serum albumin precursor is a specific inhibitor of apoptosis (Zoellner *et al.*, 1996).

HSP27 and HSP70 are both modulated in the infected host cell and inhibit key effectors of the apoptotic machinery including the apoptosome, the caspase activation complex, and apoptosis inducing factor (HSP70 only). Both play a role in the proteasome-mediated degradation of apoptosis-regulatory proteins (Garrido *et al*, 2003). HSP27 can prevent the formation of the apoptosome and the subsequent activation of caspases and inhibits

cytochrome c-mediated activation of caspases in the cytosol. HSP27 binds to cytochrome c released from the mitochondria to the cytosol and prevents cytochrome-c-mediated interaction of Apaf-1 with procaspase-9, this in turn inhibits formation of apoptosome. Thus, HSP27 interferes specifically with the mitochondrial pathway of caspase-dependent cell death (Bruey *et al.*, 2000). HSP70 protects cells from stress-induced caspase dependent apoptosis, both upstream and downstream of the death-associated mitochondrial events. HSP70 may also protect cells from energy deprivation and/or ATP depletion associated with cell death (Garrido *et al.*, 2003). The ubiquitin-proteosome system itself plays a major role in the modulation of apoptosis. HSPs could indirectly affect apoptosis by facilitating the degradation of death regulatory proteins. One of these regulatory proteins is NF-κB, whose several steps of activation involve the ubiquitin-proteosome system including maturation of the transcription factor subunits, activation of IKKs (IKB kinases) and degradation of IKalpha. HSP27 stimulates the degradation of IKalpha and thereby increases the intracellular content of NF-κB (Garrido *et al*, 2003).

Cargo selection protein TIP47 (M6PBRP1) is down regulated in the infected cell and a study by Tardy *et al* showed that mannose 6-phosphorylated proteins are required for tumour necrosis factor-induced apoptosis (extrinsic pathway) (Tardy *et al.*, 2004).

Galectin isoforms 1 and 3 are members of the β-galactosidase-binding gene family and are implicated in a variety of biological functions including cell growth, cell adhesion and cell differentiation and were modulated in the infected cell. Galectin 3 is known to bind IgE therefore is also involved in the IR. Galectins exhibit an anti-apoptotic activity in various cell types in response to a wide range of apoptotic insults. High levels of cellular expression of galectin are associated with resistance to an apoptotic insult. It is though that Galectin is strongly related to the anti-apoptotic signalling pathway of NF- κβ, a TF that induces anti-apoptotic activity. Therefore galectin-3 can suppress apoptosis and contribute to cell survival. Conversely it has also been shown that galectin-3 secreted by tumour cells, induces apoptosis, thus aiding host survival (Nakahara *et al.*, 2005).

The metabolic proteins cathepsins B and D are modulated during infection. Cathepsins have long been implicated in apoptosis and cathepsin B contributes to TNF-alpha mediated apoptosis by promoting mitochondrial release of cytochrome c (Guicciardi *et al.*, 2000). Several studies have investigated the role of cathepsins in programmed cell death and concluded that (1) agents that disrupt lysosomes and cause redistribution of cathepsins to the cytoplasm leads to induction of apoptosis, (2) apoptosis can be blocked by inhibiting cathepsin, (3) cathepsin depletion by RNA interference results in the suppression of

apoptotic signalling pathways and (4) cells from cathepsin deficient mice were found to be more resistant to stress-induced apoptosis than their normal counterparts (Tardy *et al*, 2004).

Toxoplasma gondii has also been implicated in inducing apoptosis of macrophages and cells of the immune system (Dlugonska, 2004). The ability of *Toxoplasma* to inhibit or promote host cell apoptosis is a mechanism employed by the parasite to ensure its survival. Inhibition of apoptosis ensures a safe environment for replication, along with a ready supply of essential metabolites, induction of apoptosis of certain immune cells down regulates the parasite-specific IR resulting in increased parasite survival and prevention of immunopathological processes.

8.4.13 Calcium

Calcium ions play a key role in many cellular events. Intracellular calcium is an important secondary messenger and plays a role in the control of many intracellular processes including those involved in regulation of cytoskeleton and those associated with protein synthesis and secretion. Bonhomme *et al* (1999) showed that *Toxoplasma* invasion is dependant on calcium from the extracellular medium (Bonhomme *et al*., 1999). Just prior to invasion, signalling molecules secreted by the parasite are thought to affect calcium influx into the host cell. Upon invasion signalling events are thought to lead to the release of calcium from internal pools therefore increased calcium concentration in the cell. Host cytosolic calcium varies, increased at the beginning on infection and decreased 48hpi (Coppens and Joiner, 2001). The proteomic studies revealed four calcium binding proteins were modulated in expression in the parasite infected cell; calgizzarin, clathrin light polypeptide A and reticulocalbin 1 precursor were all down regulated and calumenin was up regulated. Other modulated proteins involved in host cell calcium interactions were transgelin 1 and 2 and annexin 5, a phospholipase A2 and protein kinase C inhibitory protein with calcium channel activity and a potential role in cellular signal transduction, inflammation, growth and differentiation (Alfonso *et al*, 2005). As was noted in the studies of Coppens and Joiner (2001) that host cell calcium decreases 48hpi perhaps the decrease seen in these calcium binding proteins is linked to this.

8.4.14 Mitochondrial Proteins

Toxoplasma gondii has been shown to recruit host cell mitochondria to around the PV during infection. One member of the family of *Toxoplasma* rhoptry proteins, ROP2, has

been implicated in recruitment of host cell mitochondria to the exterior face of the parasitophorous vacuole membrane (PVM). This is apparently accomplished by the protein having its processed N terminus exposed to the host cell cytosol where it is 'mistakenly' recognised as a mitochondrial import signal. Host mitochondria recognise this sequence and in trying to import the ROP2 protein, which is firmly anchored in the PVM, are brought into close proximity to the PVM (Bradley *et al*, 2005). This indicates that the host cell mitochondria are of great importance to the replicating parasite, as is the close proximity of the host organelle to the PV.

Mitochondria are eukaryotic organelles that play a crucial role in several cellular processes, including energy production, β-oxidation of fatty acids, the urea cycle, and programmed cell death. The mitochondrial genome is about 16,500 bp long. The low complexity of the mitochondrial genome indicates that the vast majority of the mitochondrial proteins (estimated to be 1,000) are encoded by the mitochondrial genome (Da *et al*., 2003). Mitochondria are self replicating and contain their own double-stranded circular DNA. Mitochondria have a central role in energy metabolism but are also involved in many cellular processes. Only 13 proteins involved in oxidative phosphorylation are encoded by the mitochondrial genome; most of the estimated 1500 human mitochondrial proteins involved in mitochondrial function are nuclear encoded, synthesised in the cytosol and targeted to mitochondria (Cotter *et al*., 2004).

MitoProteome is a mitochondrial protein sequence database generated from experimental evidence and public databases, and contains both mitochondrial and nuclear encoded entries (http://www.mitoproteome.org/). The database contains 847 human mitochondrial proteins, 615 of which were experimentally determined by mass spectrometry. Each protein is extensively annotated with data extracted from external databases, including: gene data extracted from LocusLink and Ensembl, disease data extracted from OMIM, interaction data extracted from MINT and DIP, protein family data extracted from PFAM, domain and Motif data extracted from InterPro, fingerprint data extracted from PRINTS and BLAST hits extracted from NCBI (Cotter *et al*., 2004). All of the gene names from the proteomic experiments were searched against MitoProteome. The following table (Table 8-20) lists the 52 proteins which were modulated in expression in *Toxoplasma gondii* infected cells that are present in the MitoProteome database. This is almost one third of all modulated proteins identified in the proteomics study. This highlights just how essential the mitochondrial functions are in the infected cell, both for the parasite (increased glycolysis and inhibition of apoptosis) but also in protecting the host (immune and stress response).

Gene name	Protein	FunCat	FunCat
NDUFs5	NDUFS5	1	metabolism
AKR1B1	Aldose Reductase	1	metabolism
ALDH1A3	Aldehyde dehydrogenase 1A3	1	metabolism
CPOX	Coproporphyrinogen oxidase	1	metabolism
CTSD	Cathepsin D preprotein	1	metabolism
ECHS1	Enoyl-coenzyme A hydratase 1	1	metabolism
MTHFD2	MTHFD2 protein	1	metabolism
AK2	Adenylate kinase 2 isoform a	1	metabolism
ALDOA	Aldolase A	2	glycolysis
ATP5D	ATP synthase beta subunit	2	glycolysis
COX6B	Cytochrome C oxidase subunit Vib	2	glycolysis
GAPD	Glyceraldehyde 3 phosphate dehydrogenase	2	glycolysis
P4HB	Protein disulfide isomerise	2	glycolysis
PGK1	Phosphoglycerate kinase 1	2	glycolysis
PKM2	Pyruvate kinase M2 isozyme	2	glycolysis
TPI1	Triose phosphate isomerise	2	glycolysis
PHB	Prohibitin	3	mitosis
PRDX1	Peroxiredoxin 1	3	mitosis
EEFIA2	Elongation factor 1 alpha	5	translation
WBSCR1	WBSCR1	5	translation
HSPA5	BiP protein	6	protein fate
HSPA8	HS70kd pro 42kd ATPase	6	protein fate
HSPA9B	HSP -MTHSP75	6	protein fate
HSPB1	Heat shock protein 27	6	protein fate
HSPD1	Chaperonin HSP60	6	protein fate
PPIA	Peptidyl prolyl isomerase A (cyclophilin A)	6	protein fate
PBP	Neuropolypeptide h3	7	protein binding
TAGLN2	Transgelin 2	7	protein binding
SERPINA1	Visceral adipose-specific SERPIN	8	protein activity regulation
CLIC1	Nuclear chloride ion channel protein	9	cellular transport
ANXA2	Annexin 2	10	signal transduction
PRDX2	Peroxiredoxin 2	11	stress response
PRDX3	Peroxiredoxin 3 isoform a precursor	11	stress response
PRDX4	Thioredoxin peroxidase	11	stress response
PRDX5	Peroxiredoxin 5	11	stress response
SOD2	Superoxide dismutase	11	stress response
STIP1	STIP1	11	stress response
ANXA1	Annexin 1	13	immune response
ALB	Serum albumin precursor	14	cell fate
VDAC1	Porin 31HM	14	cell fate
MYH6	Myosin alpha heavy chain	15	cellular environment
ACTG1	Actin gamma 1 propeptide	16	structural
ACTN1	Actinin-1 alpha	16	structural
CKAP4	Cytoskeleton associated protein 4	16	structural
DES	Desmin	16	structural
LMNA	Lamin A/C, isoform 2, 70kDa lamin	16	structural
MSN	Moesin	16	structural
VIL2	Vilin 2	16	structural
VIM	Vimentin	16	structural
ANXA5	Annexin 5	17	interaction with environment
CHCHD3	Coiled-coil-helix-coiled-coil-helix domain	19	unclassified
GRP58	Protein disulfide-isomerase A3 precursor	19	unclassified

Table 8-20 Mitochondrial proteins modulated in the *Toxoplasma gondii* infected cell. The table lists the gene name, protein name and functional category of the mitochondrial linked proteins.

8.5 Conclusion

From inside the vacuole, *Toxoplasma gondii* stimulates redistribution of host cell organelles and cytoskeleton, modulates host cell transcription, and prevents host cells from undergoing apoptosis. Identification of these host cell pathways is vital to the discovery of novel drug targets to treat and prevent toxoplasmosis (Spear *et al.*, 2006). Mitochondria and ER are the most abundant organelles in mammalian cells, particularly in the perinuclear region where the *Toxoplasma* vacuoles develop. Studies looking at other intracellular pathogens (*Leishmania amazonensis* and *Coxiella burnetti*) which grow in vacuoles found that host organelles were never observed to form a close interaction with these pathogen vacuoles (Sinai *et al.*, 1997). This is therefore a host cell response specific to this intracellular pathogen and therefore many of the mitochondrial linked proteins identified as modulated in expression in this study are likely to be a *Toxoplasma* specific response.

The intracellular pathogen *Toxoplasma gondii* must induce a finely balanced host response. The parasite must avoid elimination by effectors of the host immune response but must also ensure its own replication is controlled by the host to prevent host death (Dlugonska, 2004). The proteins modulated in the parasite infected cell can be classed into three groups (1) pro-host, (2) pro-parasite and (3) bystander. Previously it has been documented that elucidation of pro-parasite responses is difficult in the context of the pathogen infected cell (Spear *et al.*,, 2006). This study has identified many pro-host proteins. A few examples include the up regulation of the metabolic protein AKR1B1 which protects cells from toxic products, down regulation of HNRPK which allows for the induction of immune response genes and the up regulation of cyclophilin A which inhibits immunosupression. Pro-parasite response proteins include the anti-apoptotic proteins modulated in the host cell and also the up regulation of the glycolytic enzymes. It is not possible within the scope of this study to classify the proteins modulated due to a bystander effect. Grouping the modulated host proteins by function and examining the pathways they are involved in provides an insight into the host-pathogen interaction. How the parasite causes the modulations seen in the cell is unknown but likely involves the parasite co-opting host pathways (Spear *et al.*, 2006).

In some cases it was difficult to determine whether a protein's expression had been up or down regulated due to parasitisation as the protein was often identified from both the infected and non-infected cells. These same proteins are most likely to be different

isoforms of the same protein, modulated by PTM. The type of MS used in these analyses could not detect these PTMs therefore the isoforms were often identified as the same protein. Study of PTMs remains very difficult and with the development of new technologies, the study of these modifications will hopefully become easier.

Chapter IX Summary and Forward Perspectives

This chapter serves to summarise the most significant findings of this study and highlight the areas which have been recognised as worthy of future research.

The main aim of this project was to investigate the host response to infection with an intracellular parasite. This aim was pursued by subtractive analysis of the whole proteome from *Toxoplasma gondii* infected and non infected cells. *Toxoplasma gondii* is an intracellular apicomplexan parasite which is widespread throughout the world and is a major of cause of disease and mortality in both humans and livestock. This parasite can survive within an immunocompetent host for the duration of the host's life and no drugs are currently available to treat the chronic stage of infection.

The subtractive analysis of the global expression profile of infected and non infected cells enabled the detection of one hundred and fifty seven proteins whose expression status differed during infection. Only a few comparative proteomic analyses of the host response to infection have been published to date, and these have focused on bacterial and viral pathogens. No published studies exist which investigate the host proteomic response to an apicomplexan parasite.

There have been vast advances in proteomic technologies over the past decade. The technique of 2-DE combined with MS was used in this work. Chapter II dealt with optimisation of the available proteomic techniques for use with our samples of interest. Optimisation is a paramount step in proteomics in order to profile as much of the proteome as possible. A robust, reliable and reproducible technique of sample preparation was developed and used for all following work in the thesis.

Chapter III investigated the differential protein expression seen in human foetal foreskin fibroblasts (HFF) after infection with *Toxoplasma gondii* using a conventional gel-to-gel 2-DE approach. This technique allowed for the identification of 85 individual proteins modulated in expression due to the presence of the parasite in the cell 24 hours post infection. Several infection time points were investigated and many replicate gels were run in order to gain an accurate representation of the host cell proteome over time post infection.

DIGE was used in Chapter IV to try to overcome the limitations associated with the gel-to-gel comparisons, including gel to gel variation, and allows for accurate statistical tests to

be carried out on the data (Unlu *et al.*, 1997). DIGE allowed for the identification of 96 differentially expressed proteins in the *Toxoplasma gondii* infected cell at 24hpi. The volume changes seen in these pathogen-infected cells were calculated using the DeCyder software and a Student's t-test carried out to determine statistically significant spot volume changes. DIGE proved a more reliable method to analyse differentially expressed proteins in the infected cell than the conventional gel to gel approach. The different techniques results in two data sets with some overlapping data showing that the approaches are complementary.

Chapter V dealt with the gene to protein correlation in *Toxoplasma gondii* infected cells and also further validation of some of the proteomic results. The modulated proteins were compared with the differentially expressed genes in *T.gondii* infected cells as determined in the microarray study by Blader *et al* (2001). Using several statistical tests the level of gene to protein correlation was deemed weakly positive. This type of gene to protein comparison is difficult as the data sets are very different. 2-DE gels can profile several thousand spots per gel and there are potentially 90,000 protein species in the cell (Jenkins and Pennington, 2001). Therefore we only examined a small percentage of the entire proteome. This is in contrast to the transcriptional studies where 22,000 genes were profiled (Blader *et al.*, 2001). Therefore although the MA analysis can determine which genes remained unchanged upon infection, the proteomic studies cannot determine this, leading to a bias in the data set. Proteomics is therefore a complementary approach to MA analysis. The data from the 24h proteomic and MA studies were compared. Ten percent of the genes and proteins profiled in both experiments did not correlate in expression status. A further 29% of the modulated proteins matched to genes on the array that did not alter in expression during infection. This subset of proteins includes many not previously implicated in the host response in infection, including the glycolytic protein aldolase b; proteins implicated in apoptosis (cathepsin D, peroxiredoxin 1 and 2, thioredoxin peroxidise and Irp1); the chloride and nuclear ion channel proteins and proteins involved in microtubule reorganisation (dynactin 2 and microtubule associated proteins 1B).

To confirm further some of the changes seen in the proteome of the infected cell, Western blotting was carried out. Three proteins of interest were investigated using this technique with the results of (1) GAPD WB confirming the changes seen in the MA and proteomic experiments, (2) cathepsin B, where the WB results contradicted the MA and proteomic results and (3) caldesmon, where it was not possible to determine any results due to a very unspecific antibody. The WB results for caldesmon and cathepsin differed from the expected result and this could have been due to suboptimal concentration of antibody, poor

antibody specificity or difficulty in differentiating between isoforms of interest. Further confirmatory studies are therefore required.

An extremely important biological question that arises from this work is: does protein expression change due to the presence of an individual parasite or is there common protein expression modulation regardless of what parasite cells are exposed to? In chapter VI to address whether the proteins modulated during infection could be a non-specific response of HFFs to intracellular parasitism (i.e. independent of the specific pathogen), we investigated the response of HFFs to another intracellular parasite, *Leishmania major*. Replicate experiments and the same data analysis methods as used for the *T.gondii* experiments were carried out. The results indicate that the proteins modulated do not represent a general HFF response to intracellular parasites. This can also be compared to the data from the Blader study which compared the gene expression of HFFs infected with *T.gondii* for 24h with HFFs infected with *T.cruzi* and found very little correlation of gene expression between the two datasets. Due to technical difficulties it was not possible to identify many of the differentially expressed proteins in the *Leishmania* infected cells, however this work did prove that the host cell does not respond similarly to all pathogens as a great deal fewer proteins changed in volume in the *Leishmania* infected cells as compared to the *Toxoplasma* infected cells.

One of the most common PTM in the cell is phosphorylation. Phosphorylation controls activation and deactivation of protein function and the phosphoproteome shows which proteins are active in the cell. MA cannot be used to investigate the phosphorylation status of proteins therefore the importance of this PTM in *T.gondii* infected cells has not previously been investigated. Proteomics can be used to investigate PTM and Chapter VII dealt with the phosphoprotoeme of *T.gondii* infected cells. Using three different techniques, the phosphorylated proteins from infected and non-infected cells were compared. Purification of the phosphoproteome using the QIAGEN PhosphoProtein purification kit allowed for identification of nine differentially expressed proteins in the infected cell. This technique did not prove very compatible with 2-DE. Diamond Pro-Q® stain selectively stains phosphoproteins and was used to stain whole cell extracts and phosphorylated extracts from infected and non-infected cells. This method did not provide good results. The final technique involved development of a novel approach to study the phosphoproteome which used alkaline phosphatase to dephosphorylate proteins in concert with DIGE to compare the phosphorylated proteome from infected and non-infected cells. No published studies have used this combination of techniques and the method proved highly successful. A modulation of expression was seen in hundreds of spots with the

identification of twenty four differentially expressed proteins. Of interest was the identification of phosphorylation of vimentin during infection. Vimentin is extensively reorganised in the pathogen infected cell and phosphorylation of vimentin obviously plays a major role in its activity.

Chapter VIII combined all of the results from the *Toxoplasma gondii* proteomic studies in this thesis and categorised all modulated proteins by function in order to gain a better understanding into the modulated proteins and pathways in the infected cell. One hundred and fifty seven proteins were identified as modulated during infection and grouping by function revealed extensive modulation of many host cell processes. Host cell metabolism, glycolysis and structural proteins were heavily modulated post infection. It has previously been shown that apoptosis is inhibited in the infected cell and to this end we identified fourteen proteins modulated in the *Toxoplasma* infected cell which are linked to apoptosis. Infection causes extensive remodelling of the host cell including a redistribution of mitochondria to around the PV indicating the importance of these organelles to the parasite. Over one third of all the identified modulated proteins in this study were linked to the mitochondria.

The differentially expressed proteins can be classed into three categories (1) pro-host, (2) pro-parasite and (3) bystander. Previous work indicated that identification of pro-parasite molecules was difficult in the context of the parasite infected cell (Spear *et al.*, 2006). This work identified many pro-host proteins including up regulation of the metabolic protein AKR1B1 which protects cells from toxic products, down regulation of HNRPK allows for the induction of immune response genes and the up regulation of cyclophilin A which inhibits immunosupression and is involved in mitochondrial biogenesis. Potential pro-parasite response proteins identified included the large number of anti-apoptotic proteins modulated in the host cell and also the up regulation of the glycolytic enzymes. The up regulation of the glycolytic pathway fits in with the existing MA data which also showed this result. The proteomic studies also identified up regulation of a novel glycolytic protein not previously associated with the host response to infection. How the differentially expressed proteins identified by the proteomic approaches fit into these three groups and what parasite factors are involved cannot be fully deduced from the work presented here. However, integration of the proteomics data with the microarray data and biochemical and cell biological analysis of mutant host cells and parasite strains will allow for such assignments. An in depth understanding of the host cell proteins and pathways that are modulated due to the parasite will help to identify parasite auxotrophies or host cell pathways essential for parasite survival. This in turn may help elucidate novel drug targets.

As is clear from the results in this thesis, the host response to *Toxoplasma gondii*, or indeed any intracellular parasite, is not straight forward. Many thousands of proteins are expressed in typical eukaryotic cells, and this is reflected in the complexity of 2D PAGE patterns typically observed for the total protein extracts from such cells. Despite the success of the conventional 2-DE technique in investigating the host-response to infection, complete proteome analysis by gel-to-gel electrophoresis suffers from a number of technical limitations. The process is labour intensive and requires significant technical expertise if reproducible gels and quantitative data are to be generated. A parallel approach used fluorescent difference gel electrophoresis (DIGE) which allows for samples labelled with different dyes to be co-separated simultaneously on the same gel, thereby eliminating gel-to-gel variation. All 2-DE techniques suffer from the following problems; different proteins can migrate to the same position on the gel making quantitation difficult, proteins with masses greater than 100,000 and those with pI values of less than 4 and greater than 9 are often not detectable on the gel and 2-DE often does not have sufficient resolution to separate multiple species originating from a single protein with PTM, such as those with carbohydrates (Wu *et al.*, 2002). Due to these limitations it is probable that a number of differentially expressed proteins between the infected and non-infected samples were not identified.

Further proteomic and functional studies are required to follow up the data presented here. Several approaches including fractionating cells into their various organelles before running consecutive gels, sequentially running 'zoom' gels, which separate proteins according to a series of narrow isoelectric ranges or using narrow range pH IPG strips would allow for increased protein profiling. Even after this, around 20% of the proteome may still be missing (Abbott, 1999). Non-gel based techniques that could be used to investigate further this host pathogen interaction include metabolic isotopic labelling for quantitative proteomics, stable isotope labelling with amino acids in cell culture allowing for in vivo incorporation of a label into proteins for MS based quantitative proteomics, or chemical labelling using ICAT (Guerrera and Kleiner, 2005). Pulsed [^{35}S] methionine labelling of cells *in vitro* allows for detection and identification of newly synthesised proteins rather than only the steady state proteome.

Small peptides are notoriously difficult to analyse in conventional 2-DE based proteomic approaches (Jenkins and Pennington, 2001). Surface enhanced laser desorption/ionisation time-of flight mass spectrometry (SELDI-TOF-MS) provides a methodology to extend examination of the pathogen infected proteome and to include smaller peptides which may be vital in the host response. Unlike MALDI-MS the surface in SELDI-MS plays an active

role in the extraction, presentation and enrichment of the sample, eliminating the need for meticulous sample preparation, including desalting, that is necessary prior to MALDI-MS (Jenkins and Pennington, 2001).

To confirm further the changes seen in the proteomic experiments, growth or inhibition studies would be very informative. Inhibition or over expression of a modulated host cell protein whose function appears to be essential for parasite survival could then be analysed in the infection system by monitoring parasite growth. Localisation studies using green fluorescent protein tagged to the proteins of interest could be used to determine where in the cell the protein is functioning and whether this differs during infection as compared to the non-infected cell. Continuation of the Western blotting to confirm the changes seen at the protein level would also aid validation of the results in the thesis and reverse genetics and phenotypic studies would help to elucidate whether the modulated protein or pathway is essential for parasite or host cell survival.

Follow up work to this study is necessary to determine the protein interactions and PTMs in the infected cell. Protein interaction can be investigated using protein MAs where proteins (e.g. antibodies or receptors) are immobilised onto an array and probed with the sample of interest to determine the interactions, immunoprecipitation can be used to identify members proteins of a protein complex and the yeast-2 hybrid system can also be used to investigate protein-protein interactions (Wu *et al.*, 2002). Further work investigating the phosphoproteome is required and alongside optimisation of the techniques discussed above, WB using antibodies specific for phosphorylation or immunoprecipitation with an anti-tyrosine antibody in combination with MALDI and MS/MS (Wu *et al.*, 2002) could be tried.

As the first interaction between the invading parasite and host cell is the cell membrane interface, one very interesting aspect to investigate using proteomics is membrane proteins. Many of these proteins are assumed to play important roles in cell adhesion, cell recognition and cell differentiation. Investigation of the changes in expression of the host cell membrane proteins during invasion of an intracellular pathogen may elucidate novel information surrounding invasion and survival of the pathogen. Due to the nature of the parasite *Toxoplasma gondii,* which actively invades host cells, the interactions between this parasite and the host cell membrane proteins are likely to be of utmost importance to this pathogens success. Many limitations surround the study of membrane proteins mainly due to their hydrophobic nature. Techniques are currently being developed for use in analysing membrane proteins.

Proteomic methods are 'hypothesis generating'. This study has demonstrated the potential of proteomics to provide unexpected insights into molecular mechanisms by virtue of its ability to study protein expression in its entirety without pre-supposing a specific mechanism. The differential expression of the phosphoproteome represents a change in the level of post-translational modification of an existing gene product which could not have been identified by the application of purely nucleic-acid based technologies.

As was discussed in Chapter V, several reasons exist to explain the poor gene to protein correlation. These include; proteomic bias, where the limitations associated with 2-DE (as discussed previously) affect the degree to which the entire proteome is profiled on a 2-DE gel; mRNA translation rates; protein abundance differs due to differing rates of protein synthesis and degradation and lastly, the error and noise in both protein and mRNA experiments that limit the ability to get a clear understanding of the gene to protein correlation (Greenbaum *et al.*, 2003; Pratt *et al.*, 2002). Protein turnover is a significant missing dimension in proteomic experiments and should therefore be considered when assessing protein abundance data and comparing it to the relative abundance of similar mRNA species (Pratt *et al.*, 2002).

The gene to protein correlation in the dynamic cell system remains unclear. Although gene expression measurements may indicate changes in protein levels, it is difficult to infer protein expression from gene expression. The steady state proteome in essentially any organism is somewhere between 30 and 80% of the possible gene products. Many of these proteins are expressed at relatively low levels (10^1-10^2/cell) although some are expressed at much higher levels (10^4-10^{10}/cell) (Leibler, 2002). Most proteins exist in multiple post-translationally modified forms. This presents a great challenge for proteomics as it is necessary to detect a large number of distinct molecular species, most of which are present at relatively low levels and many of which exist in multiple modified forms, in order to profile the entire proteome. A disadvantage of the proteomic technique is that protcomics looks at the levels of protein expression. If the expression of a protein is greatly increased in a certain situation, one might assume it is involved in the specific phenotype. However if an enzyme in a pathway is down regulated in expression or inhibited, this in turn will lead to an accumulation of the upstream component in the pathway. The sensitivity and reproducibility of 2-DE does not match that of microarray technology and quantitative analysis is difficult because the same protein can be found in many different spots on the 2-D gel (Palzkill, 2002).

The proteomic characterisation of host-pathogen interactions has a significant role in combating infectious diseases. Combining host response and pathogen characterisation at the genomic, proteomic and phenotypic levels will allow for a systems biology view. Challenges lie in comparing the genomic and proteomic data sets and the development of analytical and bioinformatic tools and databases to handle this vast amount of data. Modelling the host-pathogen interaction in this way will allow for a thorough understanding of the disease progression and elucidation of therapeutic targets.

The results from the investigation into the host-pathogen response to infection using proteomics can be integrated into the wider plan of systems biology. Systems biology is a novel approach to studying, analysing and ultimately controlling biological systems through the complex interaction of all levels of biological information (Forst, 2006). The goal of host-pathogen systems biology is the discovery and understanding of the interaction leading ultimately to drug development and therapeutic intervention. The host-pathogen interaction is finely balanced between the host immune response and pathogen virulence. For example, pathogens can hijack host cells and use host cell capabilities to the pathogens advantage, or they can evolve so rapidly that sheer diversity overwhelms the immune system, as in the case of HIV infections. The work in thesis fits in with a bottom-up approach to modelling the host-pathogen response. This approach involves a full understanding of the host-pathogen system requiring knowledge of all its components. The bottom-up approach can be used to investigate which genes, proteins or phosphorylation states of proteins are expressed or up regulated in an infection process, leading to the testable hypotheses that the regulated species are important to disease induction or progression. By integrating of genomic, proteomic and metabolomic data, models can be developed that describe intra and inter cellular processes (e.g. during drug response or disease progression) (Forst, 2006).

Complementary to the genomic and proteomic approaches is metabolomics, the study of the metabolome. The metabolome represents the collection of all metabolites in a biological organism, which are the end products of its gene expression. Thus, while mRNA gene expression data and proteomic analyses do not tell the whole story of what might be happening in the infected cell, metabolic profiling can give an instantaneous profile of the physiology of the cell (Dunn *et al.*, 2005). Investigation of the *Toxoplasma gondii* infected host cell metabolome would help increase our understanding of the dynamic host-pathogen response.

The work in this thesis has proven that 2-DE represents a powerful technique to investigate the host response to infection with an intracellular parasite. Several hundred proteins were identified as modulated due to the presence of *Toxoplasma gondii* in the host cell. Categorising these proteins by function revealed extensive modulation of host cell protein expression and pathways. Novel proteins not previously implicated in the host response to infection with this parasite have been identified and the role they play in the pathogen infected cell hypothesised. Further validative and functional studies will help to reveal the host proteins essential in both parasite and host cell survival and to gain a more in depth understanding of this intimate host pathogen relationship.

Appendices

Appendix I 24h gel-to-gel results

Appendix II 16h gel-to-gel results

Appenidix III 24h DIGE results

Appendix IV 12h DIGE results

Appendices 1-IV can be found on the enclosed CD

References

Abbott,A. (1999). A post-genomic challenge: learning to read patterns of protein synthesis. Nature *402*, 715-720.

Ajioka,J.W., Fitzpatrick,J.M., and Reitter,C.P. (2001). Toxoplasma gondii genomics: shedding light on pathogenesis and chemotherapy. Expert. Rev. Mol. Med. *3*, 1-19.

Alfonso,P., Nunez,A., Madoz-Gurpide,J., Lombardia,L., Sanchez,L., and Casal,J.I. (2005). Proteomic expression analysis of colorectal cancer by two-dimensional differential gel electrophoresis. Proteomics *5*, 2602-2611.

Banks,R.E., Dunn,M.J., Hochstrasser,D.F., Sanchez,J.C., Blackstock,W., Pappin,D.J., and Selby,P.J. (2000). Proteomics: new perspectives, new biomedical opportunities. Lancet *356*, 1749-1756.

Berg,J.M., Tymoczko,J.L. and Stryer,L. (2001). Biochemistry 5th Edition. W.H. Freeman and Company, New York.

Belcher,C.E., Drenkow,J., Kehoe,B., Gingeras,T.R., McNamara,N., Lemjabbar,H., Basbaum,C., and Relman,D.A. (2000). The transcriptional responses of respiratory epithelial cells to Bordetella pertussis reveal host defensive and pathogen counter-defensive strategies. Proc. Natl. Acad. Sci. U. S. A *97*, 13847-13852.

Belli,S.I., Walker,R.A., and Flowers,S.A. (2005). Global protein expression analysis in apicomplexan parasites: current status. Proteomics. *5*, 918-924.

Benes,P., Jurajda,M., Zaloudik,J., Izakovicova-Holla,L., and Vacha,J. (2003). C766T low-density lipoprotein receptor-related protein 1 (LRP1) gene polymorphism and susceptibility to breast cancer. Breast Cancer Res. *5*, R77-R81.

Bevan,M., Bancroft,I., Bent,E., Love,K., Goodman,H., Dean,C., Bergkamp,R., Dirkse,W., Van,S.M., Stiekema,W., Drost,L., Ridley,P., Hudson,S.A., Patel,K., Murphy,G., Piffanelli,P., Wedler,H., Wedler,E., Wambutt,R., Weitzenegger,T., Pohl,T.M., Terryn,N., Gielen,J., Villarroel,R., De,C.R., Van,M.M., Lecharny,A., Auborg,S., Gy,I., Kreis,M., Lao,N., Kavanagh,T., Hempel,S., Kotter,P., Entian,K.D., Rieger,M., Schaeffer,M., Funk,B., Mueller-Auer,S., Silvey,M., James,R., Montfort,A., Pons,A., Puigdomenech,P., Douka,A., Voukelatou,E., Milioni,D., Hatzopoulos,P., Piravandi,E., Obermaier,B., Hilbert,H., Dusterhoft,A., Moores,T., Jones,J.D., Eneva,T., Palme,K., Benes,V., Rechman,S., Ansorge,W., Cooke,R., Berger,C., Delseny,M., Voet,M., Volckaert,G., Mewes,H.W., Klosterman,S., Schueller,C., and Chalwatzis,N. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana. Nature *391*, 485-488.

Black,M.W. and Boothroyd,J.C. (2000). Lytic cycle of Toxoplasma gondii. Microbiol. Mol. Biol. Rev. *64*, 607-623.

Blader,I.J., Manger,I.D., and Boothroyd,J.C. (2001). Microarray analysis reveals previously unknown changes in Toxoplasma gondii-infected human cells. J. Biol. Chem. *276*, 24223-24231.

Bogdan,C. and Rollinghoff,M. (1999). How do protozoan parasites survive inside macrophages? Parasitol. Today *15*, 22-28.

Bogdan,C., Donhauser,N., Doring,R., Rollinghoff,M., Diefenbach,A., and Rittig,M.G. (2000). Fibroblasts as host cells in latent leishmaniosis. J. Exp. Med. *191*, 2121-2130.

Bonhomme,A., Bouchot,A., Pezzella,N., Gomez,J., Le Moal,H., and Pinon,J.M. (1999). Signaling during the invasion of host cells by Toxoplasma gondii. FEMS Microbiol. Rev. *23*, 551-561.

Bradford,M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248-254.

Bradley,P.J., Ward,C., Cheng,S.J., Alexander,D.L., Coller,S., Coombs,G.H., Dunn,J.D., Ferguson,D.J., Sanderson,S.J., Wastling,J.M., and Boothroyd,J.C. (2005). Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in Toxoplasma gondii. J. Biol. Chem. *280*, 34245-34258.

Bruey,J.M., Ducasse,C., Bonniaud,P., Ravagnan,L., Susin,S.A., Diaz-Latoud,C., Gurbuxani,S., Arrigo,A.P., Kroemer,G., Solary,E., and Garrido,C. (2000). Hsp27 negatively regulates cell death by interacting with cytochrome c. Nat. Cell Biol. *2*, 645-652.

Buates,S. and Matlashewski,G. (2001). General suppression of macrophage gene expression during Leishmania donovani infection. J. Immunol. *166*, 3416-3422.

Buxton,D. (1993). Toxoplasmosis: the first commercial vaccine. Parasitol. Today. *9*, 335-337.

Carpenter,B., Mackay,C., Alnabulsi,A., Mackay,M., Telfer,C., Melvin,W.T., and Murray,G.I. (2005). The roles of heterogeneous nuclear ribonucleoproteins in tumour development and progression. Biochim. Biophys. Acta. *1765(2)*, 85-100

Carruthers,V.B. (2002). Host cell invasion by the opportunistic pathogen Toxoplasma gondii. Acta Trop. *81*, 111-122.

Cash,P. (2002). Proteomics: the protein revolution. Biologist (London) *49*, 58-62.

Chambers,G., Lawrie,L., Cash,P., and Murray,G.I. (2000). Proteomics: a new approach to the study of disease. J. Pathol. *192*, 280-288.

Chang,K.P. (1978). Leishmania infection of human skin fibroblasts in vitro: absence of phagolysosomal fusion after induced phagocytosis of promastigotes, and their intracellular transformation. Am. J. Trop. Med. Hyg. *27*, 1084-1096.

Chang,Y.S., Yaccino,J.A., Lakshminarayanan,S., Frangos,J.A., and Tarbell,J.M. (2000). Shear-induced increase in hydraulic conductivity in endothelial cells is mediated by a nitric oxide-dependent mechanism. Arterioscler. Thromb. Vasc. Biol. *20*, 35-42.

Cohen,A.M., Rumpel,K., Coombs,G.H., and Wastling,J.M. (2002). Characterisation of global protein expression by two-dimensional electrophoresis and mass spectrometry: proteomics of Toxoplasma gondii. Int. J. Parasitol. *32*, 39-51.

Concannon,C.G., Gorman,A.M., and Samali,A. (2003). On the role of Hsp27 in regulating apoptosis. Apoptosis. *8*, 61-70.

Coppens,I. and Joiner,K.A. (2001). Parasitehost cell interactions in toxoplasmosis: new avenues for intervention? Expert. Rev. Mol. Med. 15 January, http://www.expertreviews.org/01002277h.htm

Coppens,I., Dunn,J.D., Romano,J.D., Pypaert,M., Zhang,H., Boothroyd,J.C., and Joiner,K.A. (2006). Toxoplasma gondii sequesters lysosomes from mammalian hosts in the vacuolar space. Cell *125*, 261-274.

Coppens,I. (2006). Contribution of host lipids to Toxoplasma pathogenesis. Cell Microbiol. *8*, 1-9.

Cotter,D., Guda,P., Fahy,E., and Subramaniam,S. (2004). MitoProteome: mitochondrial protein sequence database and annotation system. Nucleic Acids Res. *32*, D463-D467.

Cox,B., Kislinger,T., and Emili,A. (2005). Integrating gene and protein expression data: pattern analysis and profile mining. Methods *35*, 303-314.

Cuesta,R., Laroia,G., and Schneider,R.J. (2000). Chaperone hsp27 inhibits translation during heat shock by binding eIF4G and facilitating dissociation of cap-initiation complexes. Genes Dev. *14*, 1460-1470.

Cummings,C.A. and Relman,D.A. (2000). Using DNA microarrays to study host-microbe interactions. Emerg. Infect. Dis. *6*, 513-525.

Czupalla,C., Mansukoski,H., Pursche,T., Krause,E., and Hoflack,B. (2005). Comparative study of protein and mRNA expression during osteoclastogenesis. Proteomics *5*, 3868-3875.

Da,C.S., Xenarios,I., Langridge,J., Vilbois,F., Parone,P.A., and Martinou,J.C. (2003). Proteomic analysis of the mouse liver mitochondrial inner membrane. J. Biol. Chem. *278*, 41566-41571.

de Avalos,S.V., Blader,I.J., Fisher,M., Boothroyd,J.C., and Burleigh,B.A. (2002). Immediate/early response to Trypanosoma cruzi infection involves minimal modulation of host cell transcription. J. Biol. Chem. *277*, 639-644.

Dean,M., Hamon,Y; Chimini, G. (2001) The human ATP-binding cassette (ABC) transporter superfamily The Journal of Lipid Research. *42*, 1007-1017.

Denney,C.F., Eckmann,L., and Reed,S.L. (1999). Chemokine secretion of human cells in response to Toxoplasma gondii infection. Infect. Immun. *67*, 1547-1552.

Diehn,M. and Relman,D.A. (2001). Comparing functional genomic datasets: lessons from DNA microarray analyses of host-pathogen interactions. Curr. Opin. Microbiol. *4*, 95-101.

Dlugonska,H. (2004). Molecular modifications of host cells by Toxoplasma gondii. Pol. J. Microbiol. *53 Suppl*, 45-54.

Dunn,W.B., Bailey,N.J., and Johnson,H.E. (2005). Measuring the metabolome: current analytical technologies. Analyst *130*, 606-625.

Dzierszinski,F., Nishi,M., Ouko,L., and Roos,D.S. (2004). Dynamics of Toxoplasma gondii differentiation. Eukaryot. Cell *3*, 992-1003.

Eckmann,L., Smith,J.R., Housley,M.P., Dwinell,M.B., and Kagnoff,M.F. (2000). Analysis by high density cDNA arrays of altered gene expression in human intestinal epithelial cells in response to infection with the invasive enteric bacteria Salmonella. J. Biol. Chem. *275*, 14084-14094.

Ferguson,R.E., Carroll,H.P., Harris,A., Maher,E.R., Selby,P.J., and Banks,R.E. (2005). Housekeeping proteins: a preliminary study illustrating some limitations as useful references in protein expression studies. Proteomics. *5*, 566-571.

Florens,L., Liu,X., Wang,Y., Yang,S., Schwartz,O., Peglar,M., Carucci,D.J., Yates,J.R., III, and Wub,Y. (2004). Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. Mol. Biochem. Parasitol. *135*, 1-11.

Forst,C.V. (2006). Host-pathogen systems biology. Drug Discov. Today *11*, 220-227.

Gail,M., Gross,U., and Bohne,W. (2001). Transcriptional profile of Toxoplasma gondii-infected human fibroblasts as revealed by gene-array hybridization. Mol. Genet. Genomics *265*, 905-912.

Gail,M., Gross,U., and Bohne,W. (2004). Transferrin receptor induction in Toxoplasma gondii-infected HFF is associated with increased iron-responsive protein 1 activity and is mediated by secreted factors. Parasitol. Res. *94*, 233-239.

Garrido,C., Schmitt,E., Cande,C., Vahsen,N., Parcellier,A., and Kroemer,G. (2003). HSP27 and HSP70: potentially oncogenic apoptosis inhibitors. Cell Cycle *2*, 579-584.

Goebel,S., Gross,U., and Luder,C.G. (2001). Inhibition of host cell apoptosis by Toxoplasma gondii is accompanied by reduced activation of the caspase cascade and alterations of poly(ADP-ribose) polymerase expression. J. Cell Sci. *114*, 3495-3505.

Gorg,A., Obermaier,C., Boguth,G., Harder,A., Scheibe,B., Wildgruber,R., and Weiss,W. (2000). The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis *21*, 1037-1053.

Graveley,B.R. (2001). Alternative splicing: increasing diversity in the proteomic world. Trends Genet. *17*, 100-107.

Greenbaum,D., Luscombe,N.M., Jansen,R., Qian,J., and Gerstein,M. (2001). Interrelating different types of genomic data, from proteome to secretome: 'oming in on function. Genome Res. *11*, 1463-1468.

Greenbaum,D., Colangelo,C., Williams,K., and Gerstein,M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol. *4*, 117.

Grimwood,J., Mineo,J.R., and Kasper,L.H. (1996). Attachment of Toxoplasma gondii to host cells is host cell cycle dependent. Infect. Immun. *64*, 4099-4104.

Guerrera,I.C. and Kleiner,O. (2005). Application of mass spectrometry in proteomics. Biosci. Rep. *25*, 71-93.

Guicciardi,M.E., Deussing,J., Miyoshi,H., Bronk,S.F., Svingen,P.A., Peters,C., Kaufmann,S.H., and Gores,G.J. (2000). Cathepsin B contributes to TNF-alpha-mediated hepatocyte apoptosis by promoting mitochondrial release of cytochrome c. J. Clin. Invest *106*, 1127-1137.

Gygi,S.P., Rochon,Y., Franza,B.R., and Aebersold,R. (1999). Correlation between protein and mRNA abundance in yeast. Mol. Cell Biol. *19*, 1720-1730.

Hailu,A., Musa,A.M., Royce,C., and Wasunna,M. (2005). Visceral leishmaniasis: new health tools are needed. PLoS. Med. *2*, e211.

Hasnain,S.E., Begum,R., Ramaiah,K.V.A., Sahdev,S., Shajil,E.M., Taneja,T.K., Mohan,M., Athar,M., Sah,N.K., and Krishnaveni,M. (2003). Host-pathogen interactions during apoptosis. Journal of Biosciences *28* , 349-358.

Ikemoto,A., Bole,D.G., and Ueda,T. (2003). Glycolysis and glutamate accumulation into synaptic vesicles. Role of glyceraldehyde phosphate dehydrogenase and 3-phosphoglycerate kinase. J. Biol. Chem. *278*, 5929-5940.

Ivens,A.C., Peacock,C.S., Worthey,E.A., Murphy,L., Aggarwal,G., Berriman,M., Sisk,E., Rajandream,M.A., Adlem,E., Aert,R., Anupama,A., Apostolou,Z., Attipoe,P., Bason,N., Bauser,C., Beck,A., Beverley,S.M., Bianchettin,G., Borzym,K., Bothe,G., Bruschi,C.V., Collins,M., Cadag,E., Ciarloni,L., Clayton,C., Coulson,R.M., Cronin,A., Cruz,A.K., Davies,R.M., De Gaudenzi,J., Dobson,D.E., Duesterhoeft,A., Fazelina,G., Fosker,N., Frasch,A.C., Fraser,A., Fuchs,M., Gabel,C., Goble,A., Goffeau,A., Harris,D., Hertz-Fowler,C., Hilbert,H., Horn,D., Huang,Y., Klages,S., Knights,A., Kube,M., Larke,N., Litvin,L., Lord,A., Louie,T., Marra,M., Masuy,D., Matthews,K., Michaeli,S., Mottram,J.C., Muller-Auer,S., Munden,H., Nelson,S., Norbertczak,H., Oliver,K., O'neil,S., Pentony,M., Pohl,T.M., Price,C., Purnelle,B., Quail,M.A., Rabbinowitsch,E., Reinhardt,R., Rieger,M., Rinta,J., Robben,J., Robertson,L., Ruiz,J.C., Rutter,S., Saunders,D., Schafer,M., Schein,J., Schwartz,D.C., Seeger,K., Seyler,A., Sharp,S., Shin,H., Sivam,D., Squares,R., Squares,S., Tosato,V., Vogt,C., Volckaert,G., Wambutt,R., Warren,T., Wedler,H., Woodward,J., Zhou,S., Zimmermann,W., Smith,D.F., Blackwell,J.M., Stuart,K.D., Barrell,B., and Myler,P.J. (2005). The genome of the kinetoplastid parasite, Leishmania major. Science *309*, 436-442.

Jenkins,R.E. and Pennington,S.R. (2001). Arrays for protein expression profiling: towards a viable alternative to two-dimensional gel electrophoresis? Proteomics *1*, 13-29.

Jiang,X.S., Tang,L.Y., Dai,J., Zhou,H., Li,S.J., Xia,Q.C., Wu,J.R., and Zeng,R. (2005). Quantitative analysis of severe acute respiratory syndrome (SARS)-associated coronavirus-infected cells using proteomic approaches: implications for cellular responses to virus infection. Mol. Cell Proteomics. *4*, 902-913.

Jin,D.Y., Chae,H.Z., Rhee,S.G., and Jeang,K.T. (1997). Regulatory role for a novel human thioredoxin peroxidase in NF-kappaB activation. J. Biol. Chem. *272*, 30952-30961.

Jonas,B.A. and Privalsky,M.L. (2004). SMRT and N-CoR corepressors are regulated by distinct kinase signaling pathways. J. Biol. Chem. *279*, 54676-54686.

Jones A., Wastling J.M and Hunt E (2003) Proposal for a standard representation of two-dimensional gel electrophoresis data. Comparative and Functional Genomics 4, 492-501.

Jungblut,P.R., Zimny-Arndt,U., Zeindl-Eberhart,E., Stulik,J., Koupilova,K., Pleissner,K.P., Otto,A., Muller,E.C., Sokolowska-Kohler,W., Grabher,G., and Stoffler,G. (1999). Proteomics in human disease: cancer, heart and infectious diseases. Electrophoresis *20*, 2100-2110.

Kantengwa,S., Muller,I., Louis,J., and Polla,B.S. (1995). Infection of human and murine macrophages with Leishmania major is associated with early parasite heat shock protein synthesis but fails to induce a host cell stress response. Immunol. Cell Biol. *73*, 73-80.

Karp,N.A., Kreil,D.P., and Lilley,K.S. (2004). Determining a significant change in protein expression with DeCyder during a pair-wise comparison using two-dimensional difference gel electrophoresis. Proteomics *4*, 1421-1432.

Kato-Maeda,M., Gao,Q., and Small,P.M. (2001). Microarray analysis of pathogens and their interaction with hosts. Cell Microbiol. *3*, 713-719.

Kellam,P. (2000). Host-pathogen studies in the post-genomic era. Genome Biol. *1*, http://genomebiology.com/content/pdf/gb-2000-1-2-reviews1009.pdf

Kernec,F., Unlu,M., Labeikovsky,W., Minden,J.S., and Koretsky,A.P. (2001). Changes in the mitochondrial proteome from mouse hearts deficient in creatine kinase. Physiological Genomics *6*, 117-128.

Kim,K. and Weiss,L.M. (2004). Toxoplasma gondii: the model apicomplexan. Int. J. Parasitol. *34*, 423-432.

Kim,S.K. and Boothroyd,J.C. (2005). Stage-specific expression of surface antigens by Toxoplasma gondii as a mechanism to facilitate parasite persistence. J. Immunol. *174*, 8038-8048.

Kohler,C., Gahm,A., Noma,T., Nakazawa,A., Orrenius,S., and Zhivotovsky,B. (1999). Release of adenylate kinase 2 from the mitochondrial intermembrane space during apoptosis. FEBS Lett. *447*, 10-12.

Kramer,M.L. (2006). A new multiphasic buffer system for benzyldimethyl-n-hexadecylammonium chloride polyacrylamide gel electrophoresis of proteins providing efficient stacking. Electrophoresis *27*, 347-356.

Larsson,C. (2006). Protein kinase C and the regulation of the actin cytoskeleton. Cell Signal. *18*, 276-284.

Le Naour,F., Hohenkirk,L., Grolleau,A., Misek,D.E., Lescure,P., Geiger,J.D., Hanash,S., and Beretta,L. (2001). Profiling changes in gene expression during differentiation and maturation of monocyte-derived dendritic cells using both oligonucleotide microarrays and proteomics. J. Biol. Chem. *276*, 17920-17931.

Le Roch,K.G., Johnson,J.R., Florens,L., Zhou,Y., Santrosyan,A., Grainger,M., Yan,S.F., Williamson,K.C., Holder,A.A., Carucci,D.J., Yates,J.R., III, and Winzeler,E.A. (2004). Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle. Genome Res. *14*, 2308-2318.

Lefrancois-Martinez,A.M., Bertherat,J., Val,P., Tournaire,C., Gallo-Payet,N., Hyndman,D., Veyssiere,G., Bertagna,X., Jean,C., and Martinez,A. (2004). Decreased expression of cyclic adenosine monophosphate-regulated aldose reductase (AKR1B1) is associated with malignancy in human sporadic adrenocortical tumors. J. Clin. Endocrinol. Metab *89*, 3010-3019.

Lehr,S., Kotzka,J., Knebel,B., Schiller,M., Krone,W., and Muller-Wieland,D. (2002). Primary skin fibroblasts as human model system for proteome analysis. Proteomics. *2*, 280-287.

Leibler, D. C. (2002) Introduction to Proteomics: Tools for the New Biology. 1-95. Humana Press Inc, New Jersey.

Lian,Z., Kluger,Y., Greenbaum,D.S., Tuck,D., Gerstein,M., Berliner,N., Weissman,S.M., and Newburger,P.E. (2002). Genomic and proteomic analysis of the myeloid differentiation program: global analysis of gene expression during induced differentiation in the MPRO cell line. Blood *100*, 3209-3220.

Liebel,U., Kindler,B., and Pepperkok,R. (2004). 'Harvester': a fast meta search engine of human protein resources. Bioinformatics. *20*, 1962-1963.

Lilley,K.S. and Friedman,D.B. (2004). All about DIGE: quantification technology for differential-display 2D-gel proteomics. Expert. Rev. Proteomics. *1*, 401-409.

Lopez,M.F. (2000). Better approaches to finding the needle in a haystack: optimizing proteome analysis through automation. Electrophoresis *21*, 1082-1093.

Luder,C.G. and Seeber,F. (2001). Toxoplasma gondii and MHC-restricted antigen presentation: on degradation, transport and modulation. Int. J. Parasitol. *31*, 1355-1369.

Luder,C.G., Gross,U., and Lopes,M.F. (2001). Intracellular protozoan parasites and apoptosis: diverse strategies to modulate parasite-host interactions. Trends Parasitol. *17*, 480-486.

Lyons,R.E., McLeod,R., and Roberts,C.W. (2002). Toxoplasma gondii tachyzoite-bradyzoite interconversion. Trends Parasitol. *18*, 198-201.

Malmstrom,J., Larsen,K., Malmstrom,L., Tufvesson,E., Parker,K., Marchese,J., Williamson,B., Hattan,S., Patterson,D., Martin,S., Graber,A., Juhasz,H.P., Westergren-Thorsson,G., and Marko-Varga,G. (2004). Proteome annotations and identifications of the human pulmonary fibroblast. J. Proteome. Res. *3*, 525-537.

Mann,M., Ong,S.E., Gronborg,M., Steen,H., Jensen,O.N., and Pandey,A. (2002). Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol. *20*, 261-268.

Mann,M. and Jensen,O.N. (2003). Proteomic analysis of post-translational modifications. Nat. Biotechnol. *21*, 255-261.

Marte,B. (2003). Proteomics. Nature. *422*, 191.

Mauel,J. (1996). Intracellular survival of protozoan parasites with special reference to Leishmania spp., Toxoplasma gondii and Trypanosoma cruzi. Adv. Parasitol. *38*, 1-51.

McConnell,R.M., York,J.L., Frizzell,D., and Ezell,C. (1993). Inhibition studies of some serine and thiol proteinases by new leupeptin analogues. J. Med. Chem. *36*, 1084-1089.

Miau,L.H., Chang,C.J., Shen,B.J., Tsai,W.H., and Lee,S.C. (1998). Identification of heterogeneous nuclear ribonucleoprotein K (hnRNP K) as a repressor of C/EBPbeta-mediated gene activation. J. Biol. Chem. *273*, 10784-10791.

Molestina,R.E., Payne,T.M., Coppens,I., and Sinai,A.P. (2003). Activation of NF-kappaB by Toxoplasma gondii correlates with increased expression of antiapoptotic genes and localization of phosphorylated IkappaB to the parasitophorous vacuole membrane. J. Cell Sci. *116*, 4359-4371.

Molloy,M.P. (2000). Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients. Anal. Biochem. *280*, 1-10.

Mukherjee,S., Belbin,T.J., Spray,D.C., Iacobas,D.A., Weiss,L.M., Kitsis,R.N., Wittner,M., Jelicks,L.A., Scherer,P.E., Ding,A., and Tanowitz,H.B. (2003). Microarray analysis of changes in gene expression in a murine model of chronic chagasic cardiomyopathy. Parasitol. Res. *91*, 187-196.

Nakahara,S., Oka,N., and Raz,A. (2005). On the role of galectin-3 in cancer apoptosis. Apoptosis. *10*, 267-275.

O'Farrell,P.H. (1975). High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. *250*, 4007-4021.

Palzkill, T. (2002). Proteomics. Springer. Kluwer Academic Publishers.

Peri,S., Navarro,J.D., Kristiansen,T.Z., Amanchy,R., Surendranath,V., Muthusamy,B., Gandhi,T.K., Chandrika,K.N., Deshpande,N., Suresh,S., Rashmi,B.P., Shanker,K., Padma,N., Niranjan,V., Harsha,H.C., Talreja,N., Vrushabendra,B.M., Ramya,M.A., Yatish,A.J., Joy,M., Shivashankar,H.N., Kavitha,M.P., Menezes,M., Choudhury,D.R., Ghosh,N., Saravana,R., Chandran,S., Mohan,S., Jonnalagadda,C.K., Prasad,C.K., Kumar-Sinha,C., Deshpande,K.S., and Pandey,A. (2004). Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res. *32*, D497-D501.

Perkins,D.N., Pappin,D.J., Creasy,D.M., and Cottrell,J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551-3567.

Pockley,A.G. (2001). Heat shock proteins in health and disease: therapeutic targets or therapeutic agents? Expert. Rev. Mol. Med. *2001*, 1-21.

Pratt,J.M., Petty,J., Riba-Garcia,I., Robertson,D.H., Gaskell,S.J., Oliver,S.G., and Beynon,R.J. (2002). Dynamics of protein turnover, a missing dimension in proteomics. Mol. Cell Proteomics *1*, 579-591.

Predic,J., Soskic,V., Bradley,D., and Godovac-Zimmermann,J. (2002). Monitoring of gene expression by functional proteomics: response of human lung fibroblast cells to stimulation by endothelin-1. Biochemistry *41*, 1070-1078.

Raab, L. S. (1999) Cultural Revolution: Mycoplasma Testing Kits and Services. The Scientist. *13(20)*, 21.

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

Rabilloud,T., Blisnick,T., Heller,M., Luche,S., Aebersold,R., Lunardi,J., and Braun-Breton,C. (1999). Analysis of membrane proteins by two-dimensional electrophoresis: comparison of the proteins extracted from normal or Plasmodium falciparum-infected erythrocyte ghosts. Electrophoresis *20*, 3603-3610.

Rabilloud,T. (2002). Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics *2*, 3-10.

Rabilloud,T., Strub,J.M., Carte,N., Luche,S., Van Dorsselaer,A., Lunardi,J., Giege,R., and Florentz,C. (2002). Comparative proteomics as a new tool for exploring human mitochondrial tRNA disorders. Biochemistry *41*, 144-150.

Rabilloud,T., Heller,M., Gasnier,F., Luche,S., Rey,C., Aebersold,R., Benahmed,M., Louisot,P., and Lunardi,J. (2002). Proteomics analysis of cellular response to oxidative stress. Evidence for in vivo overoxidation of peroxiredoxins at their active site. J. Biol. Chem. *277*, 19396-19401.

Raggiaschi,R., Gotta,S., and Terstappen,G.C. (2005). Phosphoproteome analysis. Biosci. Rep. *25*, 33-44.

Renkema,G.H., Boot,R.G., Muijsers,A.O., Donker-Koopman,W.E., and Aerts,J.M. (1995). Purification and characterization of human chitotriosidase, a novel member of the chitinase family of proteins. J. Biol. Chem. *270*, 2198-2202.

Rocchi,M.S., Ballingall,K.T., Machugh,N.D., and McKeever,D.J. (2006). The kinetics of Theileria parva infection and lymphocyte transformation in vitro. Int. J. Parasitol. *36(7)*, 771-8

Roos,D.S., Donald,R.G., Morrissette,N.S., and Moulton,A.L. (1994). Molecular tools for genetic dissection of the protozoan parasite Toxoplasma gondii. Methods Cell Biol. *45*, 27-63.

Roos,D.S. (2001). Computational biology. Bioinformatics--trying to swim in a sea of data. Science *291*, 1260-1261.

Ruepp,A., Zollner,A., Maier,D., Albermann,K., Hani,J., Mokrejs,M., Tetko,I., Guldener,U., Mannhaupt,G., Munsterkotter,M., and Mewes,H.W. (2004). The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Res. *32*, 5539-5545.

Ruiz-Romero,C., Lopez-Armada,M.J., and Blanco,F.J. (2005). Proteomic characterization of human normal articular chondrocytes: A novel tool for the study of osteoarthritis and other rheumatic diseases. Proteomics *5*, 3048-3059.

Russell,R.B. (2002). Genomics, proteomics and bioinformatics: all in the same boat. Genome Biol. *3*, REPORTS4034.

Sharabiani,M.T., Siermala,M., Lehtinen,T.O., and Vihinen,M. (2005). Dynamic covariation between gene expression and proteome characteristics. BMC Bioinformatics. *6*, 215.

Sibley,L.D. (2002). No more free lunch. Nature *415*, 843-844.

Sinai,A.P., Webster,P., and Joiner,K.A. (1997). Association of host cell endoplasmic reticulum and mitochondria with the Toxoplasma gondii parasitophorous vacuole membrane: a high affinity interaction. J. Cell Sci. *110 (Pt 17)*, 2117-2128.

Sinai,A.P., Payne,T.M., Carmen,J.C., Hardi,L., Watson,S.J., and Molestina,R.E. (2004). Mechanisms underlying the manipulation of host apoptotic pathways by Toxoplasma gondii. Int. J. Parasitol. *34*, 381-391.

Spear,W., Chan,D., Coppens,I., Johnson,R.S., Giaccia,A., and Blader,I.J. (2006). The host cell transcription factor hypoxia-inducible factor 1 is required for Toxoplasma gondii growth and survival at physiological oxygen levels. Cell Microbiol. *8*, 339-352.

Stein,L.D. (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931-945.

Stensballe,A., Andersen,S., and Jensen,O.N. (2001). Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. Proteomics *1*, 207-222.

Tardy,C., Autefage,H., Garcia,V., Levade,T., and ndrieu-Abadie,N. (2004). Mannose 6-phosphorylated proteins are required for tumor necrosis factor-induced apoptosis: defective response in I-cell disease fibroblasts. J. Biol. Chem. *279*, 52914-52923.

Tyers,M. and Mann,M. (2003). From genomics to proteomics. Nature *422*, 193-197.

Unlu,M., Morgan,M.E., and Minden,J.S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis *18*, 2071-2077.

Van den Bergh,G. and Arckens,L. (2004). Fluorescent two-dimensional difference gel electrophoresis unveils the potential of gel-based proteomics. Curr. Opin. Biotechnol. *15*, 38-43.

Vermeulen,K., Berneman,Z.N., and Van Bockstaele,D.R. (2003). Cell cycle and apoptosis. Cell Prolif. *36*, 165-175.

Weiss,L.M. and Kim,K. (2000). The development and biology of bradyzoites of Toxoplasma gondii. Front Biosci. *5*, D391-D405.

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

Wilker,E. and Yaffe,M.B. (2004). 14-3-3 Proteins--a focus on cancer and human disease. J. Mol. Cell Cardiol. *37*, 633-642.

Wu,W., Hu,W., and Kavanagh,J.J. (2002). Proteomics in cancer research. Int. J. Gynecol. Cancer *12*, 409-423.

Xing,H., Zhang,S., Weinheimer,C., Kovacs,A., and Muslin,A.J. (2000). 14-3-3 proteins block apoptosis and differentially regulate MAPK cascades. EMBO J. *19*, 349-358.

Yanagida,M. (2002). Functional proteomics; current achievements. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. *771*, 89-106.

Yap,G.S. and Sher,A. (2002). The use of germ line-mutated mice in understanding host-pathogen interactions. Cell Microbiol. *4*, 627-634.

Yates,J.R., III, Speicher,S., Griffin,P.R., and Hunkapiller,T. (1993). Peptide mass maps: a highly informative approach to protein identification. Anal. Biochem. *214*, 397-408.

Zhang,C.G., Chromy,B.A., and Cutchen-Maloney,S.L. (2005). Host-pathogen interactions: a proteomic view. Expert. Rev. Proteomics. *2*, 187-202.

Zhang,P., Liu,B., Kang,S.W., Seo,M.S., Rhee,S.G., and Obeid,L.M. (1997). Thioredoxin peroxidase is a novel inhibitor of apoptosis with a mechanism distinct from that of Bcl-2. J. Biol. Chem. *272*, 30615-30618.

Zhu,H., Cong,J.P., Mamtora,G., Gingeras,T., and Shenk,T. (1998). Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. Proc. Natl. Acad. Sci. U. S. A *95*, 14470-14475.

Zhu,H., Bilgin,M., and Snyder,M. (2003). Proteomics. Annu. Rev. Biochem. *72*, 783-812.

Zoellner,H., Hofler,M., Beckmann,R., Hufnagl,P., Vanyek,E., Bielek,E., Wojta,J., Fabry,A., Lockie,S., and Binder,B.R. (1996). Serum albumin is a specific inhibitor of apoptosis in human endothelial cells. J. Cell Sci. *109*, 2571-2580.