University of Strathclyde

Strathclyde Institute of Pharmacy & Biomedical Sciences (SIPBS)

Leishmania mexicana phosphoproteomics

by

Heidi Rosenqvist

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

2011

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree

Signed: ***Date:***

PREFACE

When different disciplines of science join forces, great results can be obtained. This is the essence of the current study. I would like to acknowledge Dr. Martin Wiese, SIPBS, University of Strathclyde, Glasgow, and Professor Ole Nørregaard Jensen, BMB, University of Southern Denmark, Odense for providing me the opportunity to explore *Leishmania* biology while simultaneously being allowed to greatly improve my proteomics and mass spectrometry skills. Martin additionally took care of funding for the study through a grant received from the German Research Council (Deutsche Forschungsgemeinschaft, grant number WI 2044/5-1).

The better part of the proteomics and mass spectrometry analyses were carried out in the Protein Research (PR) Group at University of Southern Denmark. A number of people have been involved directly and indirectly. Lene Jacobsen and Steffen Bak deserve special acknowledgements for teaching me how to operate the Easy-LCs, Orbitraps and the TSQ, respectively, as well as for answering numerous questions along the way. Tine Thingholm and Karin Hjernø are acknowledged for countless discussions regarding phosphopeptide enrichment strategies, quantitative analyses, and data analytical challenges. Richard Sprenger and Veit Schämmle supplied the study with MS^E results and bioinformatics achievements, for which I am utmost grateful. The people of the PR Group in Odense as well as those of the Wiese-, Alexander-, and Roberts-laboratories at SIPBS in Glasgow are acknowledged for providing excellent working environments and after hours socialising.

Last, but not least, I would like to acknowledge my family and my partner in life, Jimmi, for support and assistance when needed.

Heidi Rosenqvist

Glasgow February 7th 2011

TABLE OF CONTENTS

TABLE OF FIGURES AND TABLES

FIGURES:

TABLES:

TABLE OF APPENDICES

Due to the form of appendices for this thesis (primarily Excel sheets with numerous columns and lines), all appendices are available for download from my Dropbox. At the first mention of every appendix in the thesis, a URL is given to specify the location of the appendix in question.

Appendix A: Analytical strategies in mass spectrometry based phosphoproteomics. Review for Methods in Molecular Biology

Appendix B: Annotation of proteins in the *Leishmania mexicana* predicted protein list. These annotations are the result of great bioinformatics efforts by Veit Schämmle, Protein Research Group, University of Southern Denmark, based on the data obtained from the MS^E experiments.

Appendix C: Preliminary *Leishmania mexicana* protein library. This library is also referred to as the "all proteins identified"-list, with information about in which life stage the proteins were detected, as well as if they appeared to be phosphorylated.

Appendix D: Phosphopeptides and phosphorylation sites validated by Ascore. This appendix contains separate lists of validated phosphopeptides and phosphorylation sites from 11 whole cell lysate datasets and 2 cytosol fraction datasets, respectively. The lists are somewhat redundant.

Appendix E: Validated phosphopeptides and phosphorylation sites from *Leishmania mexicana* protein kinases and protein phosphatases

Appendix F: SRM output for selected protein kinases. Protein name, *L. major* orthologue, and *L.mexicana* contig reference are listed in column A; Column B shows

the peptide sequence with small case letters indicating modified amino acid residues (m: methionine oxidation; c: carbamidomethylated cysteine; s, t, y: phosphorylated serine, threonine or tyrosine); Column C shows the peptide score – green is perfectly acceptable, yellow is acceptable with some reservation (e.g. high CV or peak intensity mismatch compared to library spectrum), and white is disregarded); The ratios of the different sample are shown in column D, where WTpro is always 1.0; Column E indicates whether the peptide has been included for quantification of the protein; Columns F, H, J, L, N, P, R display the total peak areas of the transitions included in quantification, and columns G, I, K, M, O, Q, S show the coefficience of variance for the total peak area detections in the three replicate for each of the different samples. For the values listed in the total area columns, the E signifies times 10, e.g. 3.5E+04 equals 35,000.

Appendix G1: SRM output protein phosphatases - WT, ΔPK4, ΔMKK promastigotes and amastigotes. This output has a make-up similar to that of Appendix F.

Appendix G2: SRM output protein phosphatases – WT promastigotes, WT axenic amastigotes, and WT lesion-derived amastigotes. This output has a make-up similar to that of Appendix F.

Appendix H: MS^E results. The MS^E data evaluated with the predicted *Leishmania mexicana* protein database. Four sheets are in this appendix:

1. MSE table – MexORF database: Table of all the identified proteins in each of the different sample types, indicating their amounts in fmol and ng, as well as displaying sequence coverage and the number of peptides identifying each protein. The column headers are: L.maj-Pro, *Leishmania major* promastigotes; L.mex-Pro, *Leishmania mexicana* promastigotes; L.mex-Ama, *Leishmania mexicana* amastigotes (axenic); CV, coefficient of variance; and n, the number of

sample runs in which the given protein was identified by the applied criteria (see section **2.2.12.1**).

2. Calculated amounts: Protein amounts calculated as pmol or µg pr 10^9 cells, as well as relative amount (mmol/mol and µg/mg). Column headers are similar to those mentioned above.
3. Standard ratio + sorted: The measured amounts of the different proteins in the different sample types related to each other. The ratios are based on the calculated amounts in the previous sheet. The standard ratios have also been sorted for proteins not detected in amastigotes (columns J-L), more abundant in amastigotes (columns M-O), less abundant in amastigotes (columns P-R), and only detected in amastigotes (columns S-U).
4. Simulated ratio + sorted: As detection of different proteins and peptides also depends on the dynamic range of the instrument conducting the analyses, this sheet contains extrapolations of the different ratios and fold changes to give approximate fold changes for those proteins not detected in the analyses.

Appendix I: Comparison of identified *Leishmania mexicana* phosphopeptide sequences to resembling sequence areas in other *Leishmania* species. The list encompasses phosphopeptides identified from wild type samples only, and compares their sequences to those found in *Leishmania major*, *Leishmania infantum*, and *Leishmania braziliensis*. In the table, * indicates the N- or C-terminus of the protein, amino acid residues in red indicate the replacement of the residue that is phosphorylated in *Leishmania mexicana* by another, typically un-phosphorylatable, residue in the related species. Underlined residues are those that have been identified as phosphorylated in *Leishmania mexicana* and may show similar modifications in the related species. The sequences highlighted in green are identical between at least two of the species. These latter phosphopeptides would thus be identifiable in *Leishmania mexicana* even if another *Leishmania* database was applied.

ABSTRACT

Leishmania, a protozoan parasite of which there are around 30 different species, causes infections (Leishmaniasis) in millions of humans and animals around the world [1]. The leishmaniases are among the most common “neglected tropical diseases” (NTDs), threatening 350 million people and being fatal to 51,000 annually [2, 3]. Treatment of leishmaniasis primarily relies on chemotherapy, but is far from unproblematic [4-8]. The search for new treatment options has turned attention towards proteins and peptides, either as drug targets [9] or as the medical reagents [5]. The function of proteins and peptides may be greatly modulated by chemical modifications. Protein phosphorylation(s), can lead to biological responses being “turned on/off” or attenuated, making studies of phosphoproteins and protein phosphorylation patterns extremely interesting in the context of identification of potential drug targets as well as signalling pathways [10]. The current study exploited phospho-specific peptide enrichment [11, 12] and a variety of quantitative approaches, to establish a pipeline for analysis of *Leishmania mexicana* proteins and their phosphorylation patterns. The pipeline led to the generation of the first preliminary library of *L. mexicana* proteins and potential phosphoproteins, containing more than 2,000 entries. 5,127 different peptides with various potential phosphorylation sites and patterns were detected, and almost 2,000 of these validated along with their more than 2,300 phosphorylation sites (slightly redundant). Additionally, a validated list of 424 non-redundant phosphorylation sites in 107 protein kinases and 36 protein phosphatases have been constructed. Quantitative analyses were carried out at both the general and phosphoproteomics level, resulting in identification of significant life-stage as well as wild type versus kinase knock-out mutant differences. Most significant was the decreased level of protein phosphorylations in MAP kinase kinase (MKK) knock-out mutants, as well as the decreased abundance or even absence of ribosomal proteins in the amastigote life stage.

LIST OF ABBREVIATIONS

AAA: Amino acid analysis

ABC: Ammonium bicarbonate, NH_4HCO_3

BLAST: Basic Local Alignment Search Tool

CID: Collision-induced dissociation

2DE: Two-dimensional gel electrophoresis

ΔMKKamast: MAP kinase kinase (MKK) knock-out amastigotes

ΔMKKpro: MKK knock-out promastigotes

ΔPK4amast: MAP kinase kinase 4 (PK4) knock-out amastigotes

ΔPK4pro: MAP kinase kinase 4 (PK4) knock-out promastigotes

DTT: Dithiothreitol

ECD: Electron capture dissociation

ePK: Eukaryotic protein kinase

ESI: Electrospray ionisation

ETD: Electron transfer dissociation

EtOH: Ethanol

FA: Formic acid

GO: Gene Ontology

HCD: High collision dissociation

HILIC: Hydrophilic interaction liquid chromatography

HPLC: High performance liquid chromatography

iFCS: Heat-inactivated foetal calf serum

IMAC: Immobilised metal affinity chromatography

iTRAQ: Isobaric tag for relative and absolute quantitation

LC-MS/MS: Liquid chromatography tandem mass spectrometry

LTQ: Linear trap quadrupole

MAP, alternatively MPK: Mitogen-activated kinase

MeCN: Acetonitrile

MeOH: Methanol

MS: Mass spectrometry

MS^2: Tandem mass spectrometry (MS/MS)

MS^3: Additional MS analysis of MS/MS products (MS/MS/MS)

MSA: Multistage activation

MS^E: Exact mass MS acquisition in a data independent manner using both high and low collision energy (the E represents the collision energy)

NTD: Neclected tropical disease(s)

PBS: Phosphate-buffered saline

PMSF: Poly-methyl sulphonyl fluoride

PR Group: Protein Research Group

PTM: Post-translational modification

RF: Radio frequency

Q-TOF: Quadrupole Time-of-Flight

SAX: Strong anion exchange

SCX: Strong cation exchange

SDU: University of Southern Denmark

s: Phospho-serine, occasionally iTRAQ-labelled serine

S: Serine

SIMAC: Sequential immobilised metal affinity chromatography (IMAC)

SRM: Selected reaction monitoring

STAGE: STop-And-Go-Extraction

t: Phospho-threonine, occasionally iTRAQ-labelled threonine

T: Threonine

TFA: Trifluoro-acetic acid

TSQ: Triple stage quadrupole

WTamast: Wild type amastigotes

WT_{LDA}: Wild type lesion-derived amastigotes

WTpro: Wild type promastigotes

y: Phospho-tyrosine, occasionally iTRAQ-labelled tyrosine

Y: Tyrosine

CHAPTER 1: INTRODUCTION

In this chapter, the biological system (Leishmania parasites) will be introduced. The link between the biological system and phosphoproteomics is described in the section on signalling pathways. The phosphoproteomics section contains relevant excerpts of sections that I have written for an invited review. The entire review can be found in ***Appendix A*** (*http://dl.dropbox.com/u/3011619/Appendix%20A.pdf*). *A description of the technical basis of the study is provided in the mass spectrometry section, followed by a description of the relevant approaches for mass spectrometry-based quantitative analyses and bioinformatics analyses. The chapter is concluded by perspectives, aim and hypotheses of the study.*

1.1 Leishmania and leishmaniasis

1.1.1 Leishmania

Leishmania parasites pass through several morphologically different life stages during their stays in the phlebotomine sand fly vector and the mammalian host, respectively [13, 14]. In the posterior part of the sand fly's abdominal midgut, the small spherical and non-motile amastigotes from infected macrophages of a bloodmeal differentiate into small, sluggish procyclic promastigotes with short flagella [14-16]. After 24-48 hours, the procyclic promastigotes differentiate into large and slender nectomonads that will migrate towards the anterior thoracic midgut of the sand fly. Around day 4 in the sand fly midgut, leptomonads develop, and as the parasites reach the boundary of the foregut, haptomonads and metacyclic cells are observed. The haptomonads are non-motile, leaf-like parasites with short flagella, specialised to plug the stomodeal valve between the thoracic midgut and the foregut of the sand fly, whereas the infective metacyclic cells are small, rapidly swimming forms with elongated flagella, adapted for successful transmission to the mammalian host. Proliferation of the parasites within the sand fly

occurs at the procyclic promastigote as well as the leptomonad stages [14]. When transmitted to the mammalian host, the metacyclic cells take up residence in the phagolysosomes of macrophages, where the increased temperature (as compared to the sand fly gut) and lower pH cause the cells to differentiate into amastigotes and proliferate [16]. Through their life cycle, the parasites alternate between the proliferative forms of procyclic promastigotes and amastigotes, and the cell cycle arrested form of highly infectious metacyclic promastigotes [17]. **Figure 1.1** illustrates the life cycle of *Leishmania* in the sand fly vector, whereas **Figure 1.2** illustrates the combined life cycle of *Leishmania* in both the sand fly vector and the mammalian host.

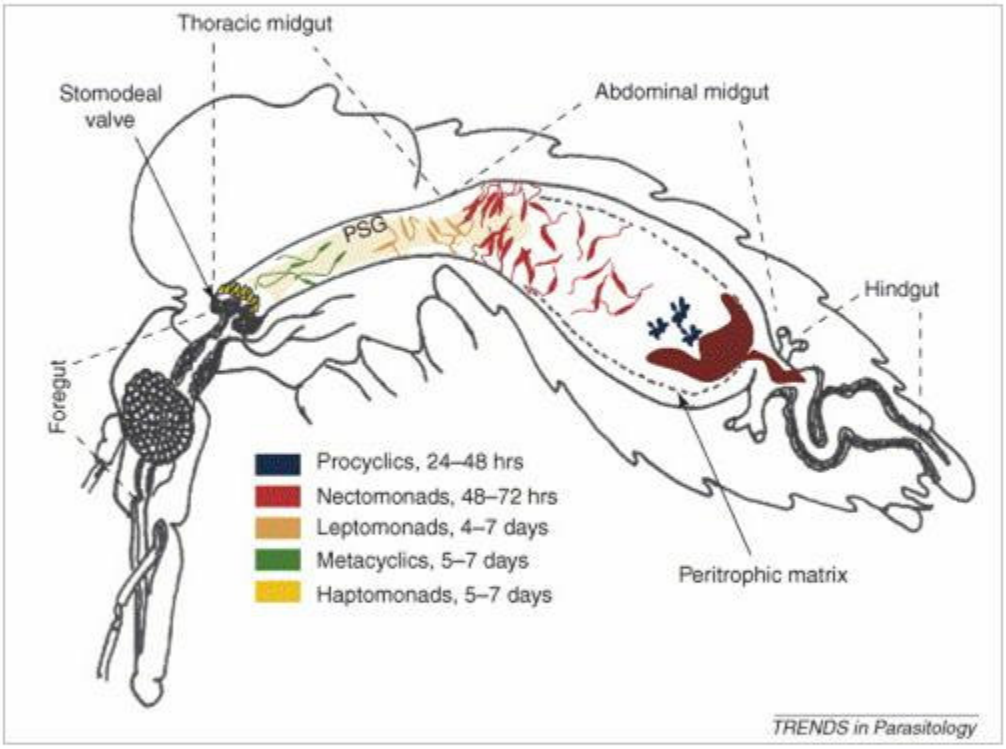


Figure 1.1: The life cycle of Leishmania in the sand fly vector with specification of the timewise appearance of the different morphological forms of promastigotes [14]. The amastigotes derived from infected macrophages of the bloodmeal migrate from the posterior abdominal midgut to the stomodeal valve of the sand fly while passing through 5 different developmental stages of distinct morphology.

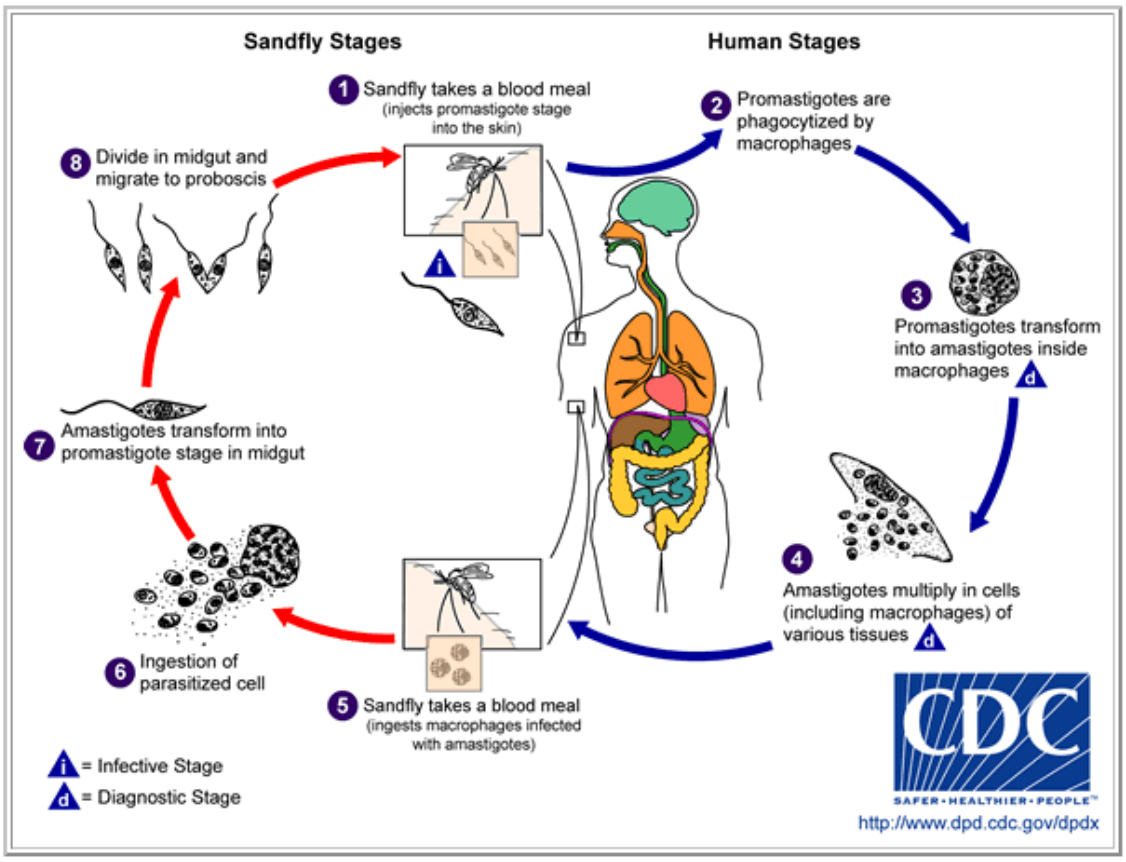


Figure 1.2: Life cycle of Leishmania. Female sandflies inject the infective metacyclic promastigotes into their hosts during blood meals (1). In the wound of the host, the promastigotes are phagocytosed by macrophages (2), in which they differentiate into amastigotes (3). The amastigotes proliferate in the infected cells, and affect various tissues, depending on which Leishmania species is the cause of infection (4). At this point, the host clinically presents with leishmaniasis. Sand flies feeding on blood from infected individuals or animals will also get infected as they ingest amastigote-filled infected macrophages (5) (6). In the midgut of the sand fly, the amastigotes will differentiate into promastigotes (7), proliferate and migrate to the proboscis (8) for the cycle to start all over again [18].

1.1.2 Leishmaniasis

Infection with *Leishmania* is termed leishmaniasis. Leishmaniasis is registered in 88 countries around the world, and is endemic in more than 60 countries [3, 19]. **Figure 1.3** shows the areas of the world in which leishmaniases are encountered, along with the approximate number of people infected or at risk of infection in 2003.

Globally, there is a yearly incidence of 1-1.5 million cases of cutaneous and 500 000 cases of visceral leishmaniasis [20]. Visceral leishmaniasis is a serious concern, and if untreated lethality may be as high as 100 % within 2 years of onset [3]. It is characterised by irregular bouts of fever, weight loss, hepatosplenomegaly (the spleen may become much larger than the liver), anaemia, and possible skin pigmentation ("kala azar" – black disease) [3, 19]. Death usually occurs due to secondary bacterial infections in the advanced stages of the disease [3, 19]. Cutaneous leishmaniasis presents as ulcers at the site of a sand fly bite, but can generate a large number of lesions [3, 19]. Mucocutaneous leishmaniasis may occur by itself or as a long-term complication of cutaneous leishmaniasis. It affects the mucous membranes and cartilage of the nose, oral cavity and pharynx, causing disabling facial changes that may negatively affect the abilities to breathe and eat. As with visceral leishmaniasis, mucocutaneous leishmaniasis carries an increased risk of secondary infection causing significant mortality [3, 19].

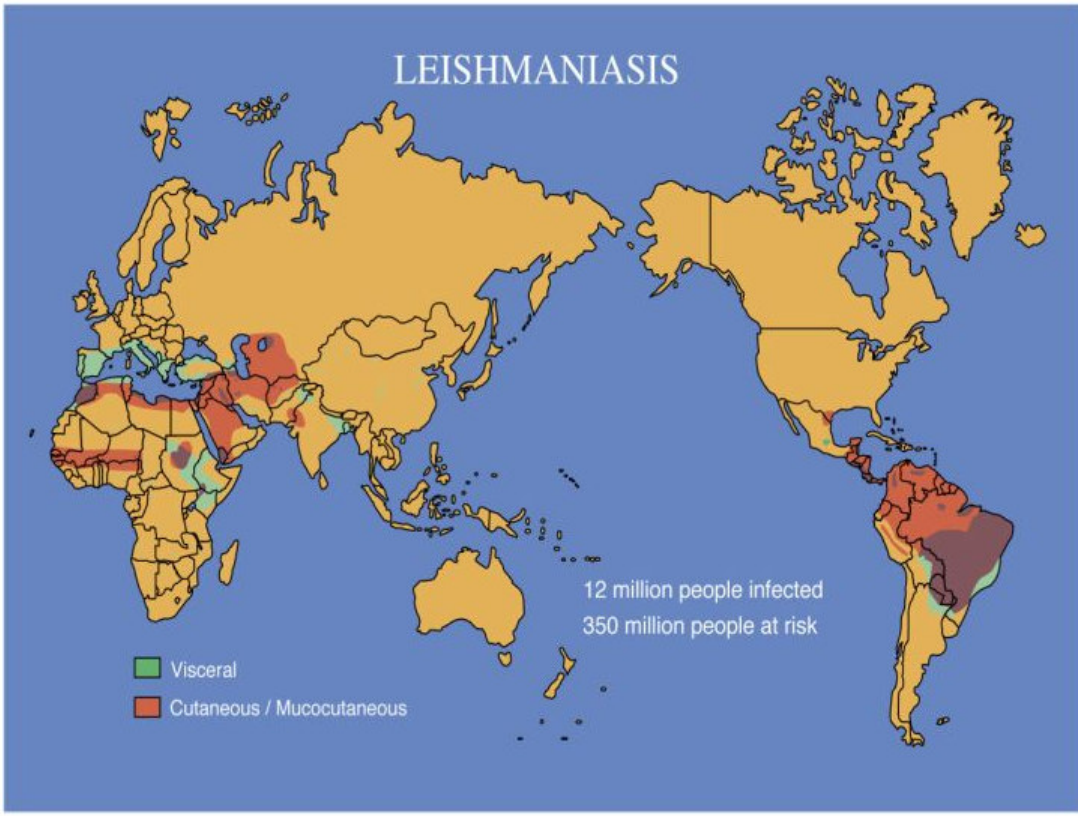


Figure 1.3: Mapping the distribution of leishmaniasis. The colour code refers to the different forms of leishmaniasis with green representing the visceral forms and red the cutaneous and mucocutaneous forms. Areas where both visceral and cutaneous/mucocutaneous forms exist are purple [21].

Because the many different species of *Leishmania* give rise to various clinical presentations, treatment of leishmaniasis is challenging. There is no absolutely safe, simple, and effective treatment of leishmaniasis [22]. When leishmaniasis is cured, this is always caused by host macrophages that have been activated by the cellular immune response to eliminate the parasites. The humoral immune response, though strongly induced by *Leishmania*, does not appear to provide any protection; in fact antibodies are associated with non-healing forms of leishmaniasis. However, based on the facts that only a minor part of the population develop active disease in endemic areas, and successfully cured patients rarely become re-infected, vaccination could be a possible means of fighting leishmaniasis [20]. Along with looking for potential vaccines, research is looking to elucidate the areas of new possible drugs/drug targets and resistance mechanisms [6, 7, 23, 24].

1.2 Signal transduction

Signal transduction is essential in almost every aspect of a cell's survival and function. Different signalling pathways regulate gene expression, protein synthesis and modifications, as well as the cell's response and adaptation to environmental changes. Disruption or changes of signal transduction can greatly compromise the function and survival of the cell. This is of course important when wanting to find new ways to deflect or prevent health issues caused by pathogenic organisms such as *Leishmania*. Parsons and Ruben have reviewed the different pathways and molecules that are known to function in Trypanosomatids, displaying some significant differences from other eukaryotes [25].

1.2.1 Signalling networks: Trypanosomatids vs. other eukaryotes

Based on the genomic data available, trypanosomatids seem to miss some of the otherwise essential components involved in eukaryotic signal transduction. In contrast to other eukaryotes, signalling pathways culminating in activation of transcription factors have not been identified in trypanosomatids, and several known components of signal transduction are missing, including receptor protein kinases and phosphatases, serpentine receptors and heterotrimeric G proteins [25, 26].

The vast majority of organisms, from *E. coli* to humans, rely on regulation of transcription initiation to adapt to environmental changes. Unlike most other organisms, however, *Leishmania* and other kinetoplastida in general do not regulate differential protein expression at the level of transcription initiation [27]. In contrast to the majority of organisms, regulation of protein expression seems to be achieved in the absence of regulation of RNA polymerase II activity [28]. Instead, it is thought to be achieved by post-transcriptional regulation of mRNA processing and stability (e.g. trans-splicing, translation efficiency, and RNA- and protein stability) [25, 27-29].

The significant morphological and metabolic differences between the two main life stages of *Leishmania*, must entail differential regulation of signal-mediated processes. Since the environmental factors (e.g. pH and temperature) are so different between the insect vector and mammalian host, some kind of sensory, transporter and/or receptor-type of molecules would be likely to be involved in initiation of the signalling cascades directing differentiation. However, such molecules remain to be identified [30].

1.2.2 Post-translational protein modifications as "messengers"

Signal transduction is performed by chemical "messengers", e.g. small molecules, peptide hormones and chemical modifications to proteins alone or in concert. Intracellular signalling networks are often ruled by post-translational protein modifications (PTMs), such as phosphorylation, glycosylation, acetylation, and ubiquitination. Affecting an estimated 30 % of the protein population at any given time point, phosphorylation is one of the most abundant and widely studied PTMs [10, 31]. The presence or absence of phosphorylation(s) at specific amino acid(s) of a protein sequence may enhance or attenuate the activity of the protein and affects its interaction with other proteins.

1.2.3 Regulation of phosphorylations

Protein phosphorylation is regulated by kinases that add phosphate groups to specific amino acids, and phosphatases that remove the phosphate groups again by hydrolysis [32]. Protein kinases are one group of potential drug targets subjected to the interest of research [7], and given the regulatory function of these proteins they might well be involved, directly or indirectly, in different resistance mechanisms. Research focused on protein kinases and the protein kinase complement of an organism's genome were termed "kinomics" and "the kinome", respectively, back in 2002 [33, 34], while the corresponding phosphatase complement is termed "the phosphatome" [35, 36]. Studies of both the *Leishmania* kinome and phosphatome have recently been conducted as part of the triTryp-projects, but not least on the kinome level a lot of questions still remain to be answered [26, 37, 38].

Mapping of the *Leishmania major* genome [39] revealed 179 eukaryotic protein kinases (ePKs) and 17 atypical protein kinases that were not closely related to the ePKs in terms of sequence and presence of subdomains [38]. Among the different families of protein kinases, the cyclin-dependent kinases (CSKs) and the mitogen-activated protein (MAP)

kinases have caught most attention with regard to leishmaniasis [7]. Homologues and/or orthologues of these evolutionary conserved enzymes involved in important cell cycle regulation and passing on of extracellular signals to intracellular responses have been found in all trypanosomatids [26, 40, 41]. The interest in MAP kinase (MAPK) and MAP kinase kinase (MKK) homologues of *Leishmania* is understandable as studies in *Leishmania mexicana* have shown that many of the currently known MAPKs and MKKs are important or even essential for parasite survival [16, 17], differentiation [42], proliferation [43], and flagellar length [15, 44, 45]. However, knowledge about the signalling pathways is still scarce, as the *Leishmania* genome seemingly misses some of the typical players of MAP kinase signalling known from other eukaryotes, e.g. the Raf- and G-proteins acting upstream of MAP kinase activation [26, 38].

Characterisation of the triTryp kinome revealed a complete lack of typical eukaryotic protein tyrosine kinases and tyrosine kinase-like proteins in trypanosomatids [30]. Tyrosine phosphorylation is present in trypanosomatids [41, 46, 47], as is also documented in the current study, and trypanosomatid protein tyrosine phosphatases have long been known [48, 49]. Tyrosine phosphorylation in trypanosomatids is thought to be carried out by dual-specificity kinases and/or some of the atypical kinases not closely resembling other eukaryotic kinases [30].

The challenges of understanding signalling networks in *Leishmania* and other trypanosomatids are not solely caused by the lack of knowledge about which kinases conduct tyrosine phosphorylation and how they do it (auto- and/or substrate phosphorylation). Just as significant is the fact that very few protein kinase-substrate relations are known [44, 50]. In the search for new information in this field, proteomics and mass spectrometry can play important roles [10]. Characterisation of protein phosphorylations and changes of these under normal and perturbed conditions can lead to identification of kinase specific phosphorylation motifs to provide ideas about potential responsible kinases. In combination with other experimental approaches [51], quantitative phosphoproteomics may provide identification of kinase substrates [52].

1.3 Phosphoproteomics

Proteomics is the science of identity, function, expression, etc. of proteins in a cell, tissue or organism [53, 54], and thus phosphoproteomics deals with the facts concerning the phosphorylated complement of a proteome [55]. In order to understand what is going on in a cell or organism, knowledge about PTMs of proteins is just as important as knowing the identity of the proteins. Several papers and reviews describe the challenges of PTM research, and how the focus and techniques have changed just within the past decade [31, 56-61].

1.3.1 Analytical means for phosphoproteomics studies

A typical phosphoproteomics workflow consists of several different sample preparation and analytical levels (**Figure 1.4**) [62]. Mass spectrometry enables detection of changes in the molecular weight of a protein or peptide due to PTMs like phosphorylations. A bottom-up approach (i.e., protein analysis at the peptide level) is often chosen due to superior accuracy as well as the desire to obtain sequence specific information about the protein phosphorylation sites [31]. The chemical nature of the phosphorylated peptides as well as the stoichiometric relation between these and their un-phosphorylated peers, however, often calls for specific enrichment steps and/or strategies in order to obtain information about the position(s) of the phosphorylation(s). Several enrichment strategies exist to allow for specific separation of phosphorylated and non-phosphorylated peptides, mainly by chromatographic means [31, 63-67]. Only those strategies employed in the current study will be introduced in the following sections.

Figure 1.4 – Legend on p. 16

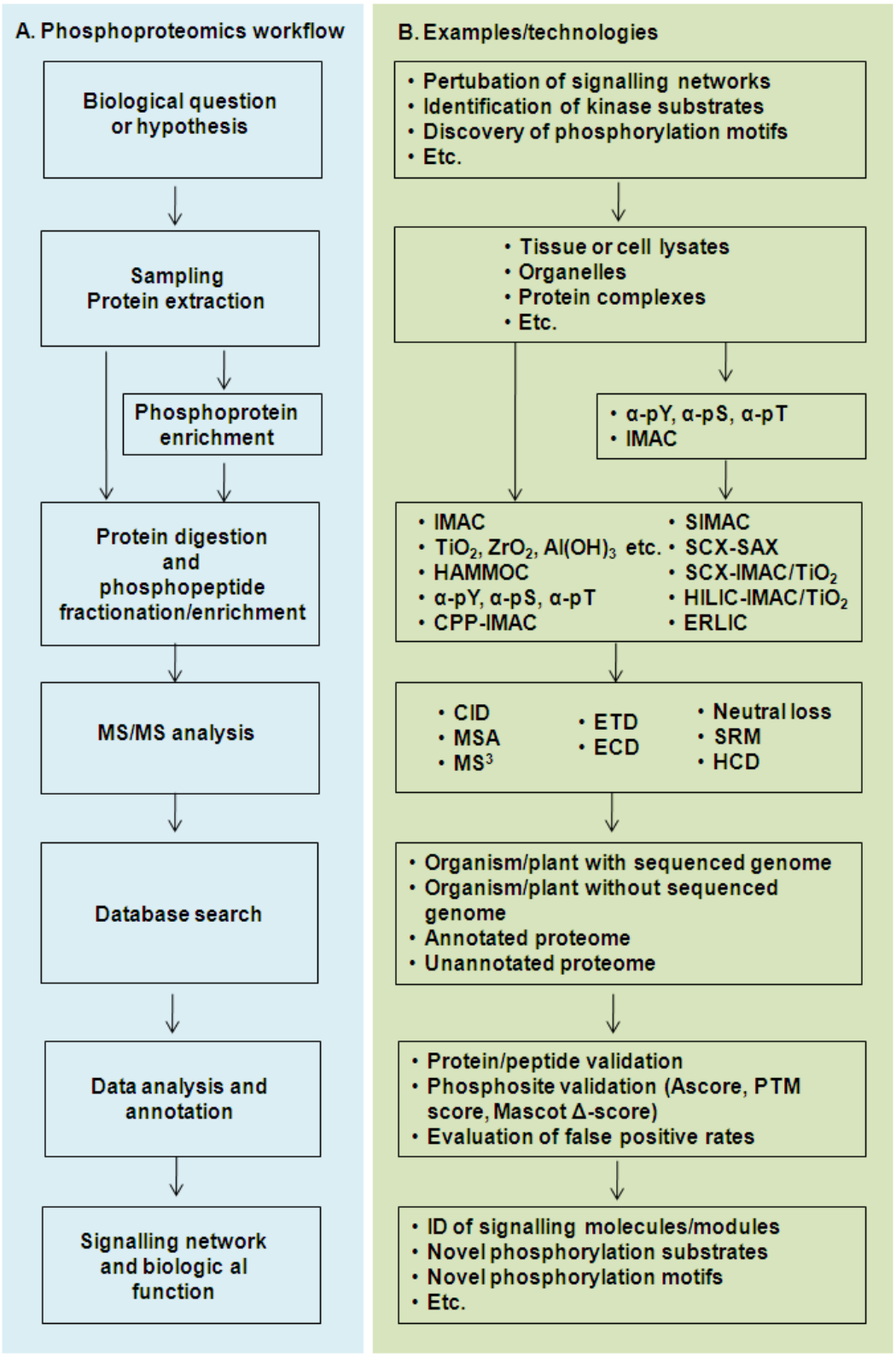


Figure 1.4: Example of a typical work flow in phosphoproteomics (A) with examples of available techniques (B). The type of sample as well as procedure for protein extraction is determined by the biological question(s) or hypothesis to investigate. Enrichment of phosphorylated entities can take place on protein and/or peptide level with different individual or combined procedures. At the mass spectrometry level, MS^2, MS^3 or pseudo-MS^3 is typically required for phosphopeptide analyses, and the subsequent sequence database search will depend on the type of sample being studied. Validation of phosphopeptides and phosphorylation sites as well as evaluation of false positive rates is part of the data analysis and annotation step. Finally, the identified phosphoproteins may be placed in specific pathways, and the specific phosphorylation sites may be grouped as known or new as may their phosphorylation site motifs [62]. Abbreviations: α-pY, phosphotyrosine antibody; α-pS, phosphoserine antibody, α-pT, phosphothreonine antibody; IMAC, immobilised metal affinity chromatography; HAMMOC, hydroxyl acid-modified metal oxide chromatography; CPP, calcium phosphate precipitation; SIMAC, sequential elution from IMAC; SCX, strong cation exchange; SAX, strong anion exchange; HILIC, hydrophilic interaction liquid chromatography; ERLIC, electrostatic repulsion-hydrophilic interaction chromatography; CID, collision-induced dissociation; MSA, multistage activation; ETD, electron transfer dissociation; ECD, electron capture dissociation; SRM, selected reaction monitoring; HCD, high energy collision dissociation; PTM, post-translational modification; Mascot Δ-score, Mascot delta-score; ID, identification.

1.3.1.1 Immobilised Metal Affinity Chromatography – IMAC

Immobilised metal affinity chromatography (IMAC) was one of the first methods to gain popularity in phosphoproteomics. IMAC exploits the affinity of positively charged metal ions (usually Fe^{3+} or Ga^{3+}, but Zr^{4+} has also been applied) to catch the negatively charged phosphopeptides on a chromatography column [64, 66-68]. While phosphopeptides are caught by the metal cations, the un-phosphorylated peptides that are mainly positively charged pass through the column and can be collected for separate analyses. The phosphopeptides can be eluted by an alkaline buffer. Both eluted phosphopeptides and the un-phosphorylated peptides of the column flow through can be analysed by mass spectrometry with different settings [69]. Unspecific binding to the IMAC material is recognised, e.g., binding of un-phosphorylated peptides containing multiple acidic (i.e., negatively charged) amino acid residues, and different means, like conversion of peptides to the corresponding peptide methyl esters and sample pH equilibration, have been used to circumvent it [70, 71].

1.3.1.2 TiO_2

Metal oxides/hydroxides have also proven their potential in phosphopeptides enrichment [72-77]. Of these, TiO_2 was the first metal oxide shown to efficiently retain organic phosphates under acidic conditions, while allowing their elution under alkaline conditions [78]. In 2004, Pinkse *et al.* exploited these characteristics for an LC-MS/MS based procedure, allowing for online phosphopeptides enrichment, separation and sequencing by TiO_2 [74]. The procedures for use of TiO_2 in phosphoproteomics have since been assessed, improved and modified by different groups [11, 75, 79, 80]. Similar to IMAC in pH dependency for phosphopeptides and elution, TiO_2 seems to be more specific in not catching as many un-phosphorylated peptides [11], although some claim that unspecific binding is still an issue that needs to be addressed [80]. The ion exchange properties of the metal dioxide are responsible for the phosphopeptide affinity

of TiO_2 [74], and may also affect the retention of singly vs. multiply phosphorylated peptides. IMAC appears to be more effective than TiO_2 in retaining multiphosphorylated peptides, whereas TiO_2 seem to favour singly phosphorylated peptides. However, TiO_2 will actually bind both types of phosphopeptides, only the conditions at which they can be eluted seem to differ [81]. Typical elution conditions (elution buffer pH 10-11.5) will favour release of singly phosphorylated peptides [11], but a change of pH and elution buffer may cause the release of multiphosphorylated peptides as well [81].

1.3.1.3 Sequential elution from IMAC – SIMAC

The different enrichment procedures for phosphopeptides, reviewed in **Appendix A**, all have strengths and limitations. Hence, combination of different enrichment procedures is often advantageous for phosphopeptides enrichment [65, 71, 82-84]. One such combination is between IMAC and TiO_2, known as sequential elution from IMAC or SIMAC [12]. SIMAC allows enrichment of singly and multiply phosphorylated peptides in separate fractions, which subsequently allows for optimised LC-MS/MS analyses of the different fractions. Multiply phosphorylated peptides suffer from poor ionisation and are suppressed in the presence of both singly and un-phosphorylated peptides, so the separation of these peptides in SIMAC improves recovery of heavily phosphorylated peptides. This also leads to better coverage of the phosphoproteome as demonstrated by a doubling in the number of identified phosphorylation sites by SIMAC compared to TiO_2 alone in Thingholm *et al.*'s test [12].

1.4 Mass spectrometry

Mass spectrometry (MS) enables analyses of small molecules, e.g. proteins, peptides, DNA or RNA, by their mass and charge, with a precision down to a few parts per million (ppm) [54]. The technique is based on experimental work dating back as far as 1886, the first mass spectrometer being constructed in 1912 by J. J. Thomson [85]. A mass spectrometric analysis comprises three main events: 1) ion production; 2) ion transmission; and 3) ion detection [86], and to enable these a mass spectrometer generally consists of an ion source, one or more mass analysers, and a detector (**Figure 1.5**). Being transformed from either solid or liquid phase to gas phase, the sample is ionised in the ion source. Pressure and temperature affect the kinetic energy of the generated ions, hence a vacuum system with a pressure of 10^5-10^8 Torr controls their motion during the transmission through the mass analyser(s) to the detector [86]. Travelling in vacuum makes it less likely for the ions to collide with the neutral background gaseous molecules, thus increasing their chance of reaching the detector. In the mass analyser, the generated ions are separated by their mass and charge, e.g. by their Time-of-Flight, TOF, through a vacuum tube. To conduct more advanced tandem MS or MS^n analysis, a collision cell as well as a second mass analyser (tandem MS "in space"), or a trap that can store ions for repeated fragmentation (MS "in time") are required between the initial mass analyser and the detector [85]. The detector collects the signals from the different ions and generates a mass spectrum where the x-axis displays the mass-to-charge (*m/z*) ratio of the ions, and the y-axis displays the intensity of the ion signals, correlating somewhat with the amount of the given ions present in the sample.

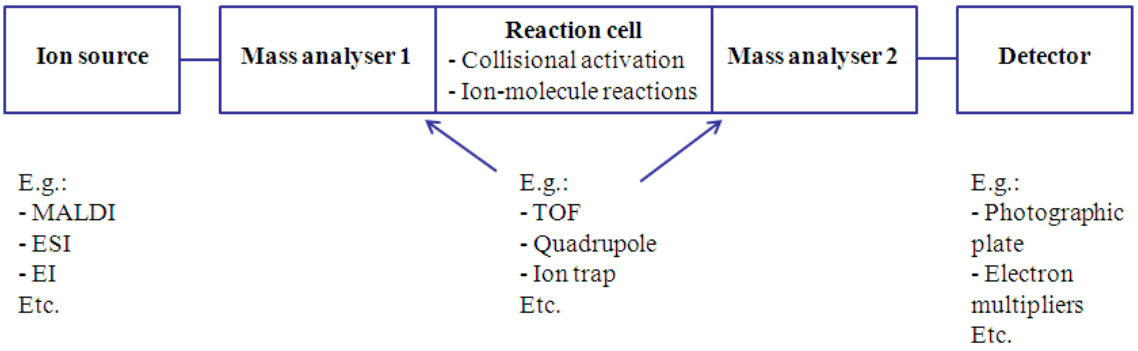


Figure 1.5: General mass spectrometer make-up. Each box represents a component that can be one of many different types, and only few are mentioned in this figure. Matrix-assisted laser desorption/ionisation (MALDI), electrospray ionisation (ESI), and electron ionisation (EI) are examples of ion source types. MALDI and ESI are especially popular for proteomics analyses. Time-of-Flight (TOF), quadrupoles, and ion traps are examples of mass analyser types. Mass spectrometers only capable of MS (peptide/protein scan) analyses will contain only a single mass analyser. For mass spectrometers capable of MS^n (where *n* is 2 or more) analyses, some kind of reaction cell as well as a second mass analyser are present. The first and second mass analyser can be of the same type (e.g. TOF-TOF set-up), but need not be. Quadrupole-TOF (Q-TOF) and ion trap-Fourier Transform Ion Cyclotron Resonance (ion trap-FTICR) are popular examples of mass analyser hybrid types of instruments.

With the introduction of the two major soft ionisation techniques, Electrospray Ionisation (ESI) and Matrix-assisted Laser Desorption/Ionisation (MALDI), in the late 1980s, analysis of proteins and peptides by mass spectrometry became possible [85, 87-89]. Since then, mass spectrometry has become increasingly popular, and as the technology has evolved [90], it is now an essential tool in proteomics [91], as it not only allows for identification analyses, but also for characterisation (e.g. post-translational modifications), quantification, structural analysis, and interaction studies [58, 86, 92-101]. Many different mass spectrometers exist, combining different types of ionisation and mass analysers. Hence, a full introduction to all these types is far beyond the scope of this thesis. The results presented here are primarily based on analyses performed on LC-coupled Orbitrap, triple quadrupole, and to a lesser extent Q-TOF and MALDI-TOF instruments. Therefore, only short and very general introductions to the components and functions of these types of instruments will be given here, with a separate section on the fragmentation strategies applied. For more thorough descriptions of instruments and applicability, the reader is referred to the references.

1.4.1 Ion source

Which ionisation technique and instrument to use depends somewhat on the type of sample on hand, and the analyses wished to be conducted. MALDI TOF instruments, with or without tandem MS capabilities, are ideal for protein identification and analysis of simple samples, e.g. purified proteins or spots from 1D or 2D gels. Because the sample is crystallised on the MALDI target plate, mass spectrometric analyses can be repeated on the same sample which is advantageous e.g. when working with tandem mass spectrometry as initial data analyses can then determine which peaks to select for fragmentation since the most abundant ions are not always the ones of special interest.

Mass spectrometry can be coupled directly [102] or indirectly [103] to a high performance liquid chromatography (HPLC or LC) system to enable fractionation of the

sample components prior to the MS analysis. This is especially advantageous when working with complex samples and therefore has many applications in modern proteomics [104]. Because the sample remains in solution in ESI, this instrument type commonly exploits in-line LC fractionation.

1.4.1.1 Matrix-assisted laser desorption/ionisation, MALDI

For MALDI MS, the analyte is mixed with matrix and allowed to crystallise on a metal plate target. The matrix is a chemical compound, typically an organic acid, which has an absorption band that closely coincides with the energy of the laser radiation, thus facilitating ionisation of the analyte. In the MALDI ion source, the matrix-analyte mixture is exposed to high-pulse laser shots, typically from a nitrogen laser, which leads to desorption and ionisation of both the analyte and matrix [86]. The principle of MALDI MS is demonstrated in **Figure 1.6**.

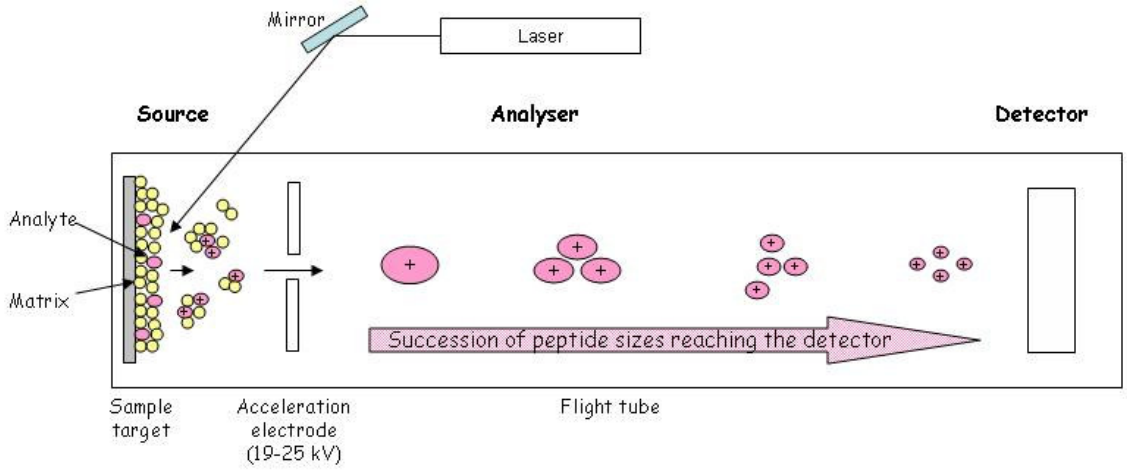


Figure 1.6: Simple illustration of the principle of matrix-assisted laser desorption/ionisation mass spectrometry. A mixture of analyte and matrix is allowed to crystallise on the target plate before insertion into the mass spectrometer. When a laser irradiates the analyte and matrix crystals, it induces ionisation of the matrix, desorption, and transfer of protons from the photo-excited matrix to the analyte to form a protonated molecule. The protonated molecules (ions) enter the mass analyser where they are separated by their mass-to-charge ratio (m/z). The ions' arrival at the detector depends on their mass, charge, and kinetic energy (inspired by [105])

1.4.1.2 Electrospray ionisation, ESI

ESI MS is conducted on analytes in solution, usually introduced into the mass spectrometer by an HPLC-system up front [94]. The analyte solution is pumped through a capillary needle on which a high positive or negative voltage is applied to generate an electrical field. The electrical field produces charge separation on the liquid surface, causing a fine mist of the analyte solution to be generated from the tip of the needle. The droplets of the mist carry an excess of positive charges, i.e., protons. Proton transfer can take place either in the analyte solution or in the droplets of the electrospray. Moving from the source towards the entrance of the mass analyser, the solvent evaporates. As this proceeds, droplet size decreases, and Coulombic repulsion between the positive charges creates gas phase ions [106]. In proteomics analyses, the generated ions will usually have multiple charges, how many depends, inter alia, on their molecular mass and structure. **Figure 1.7** illustrates the ESI process.

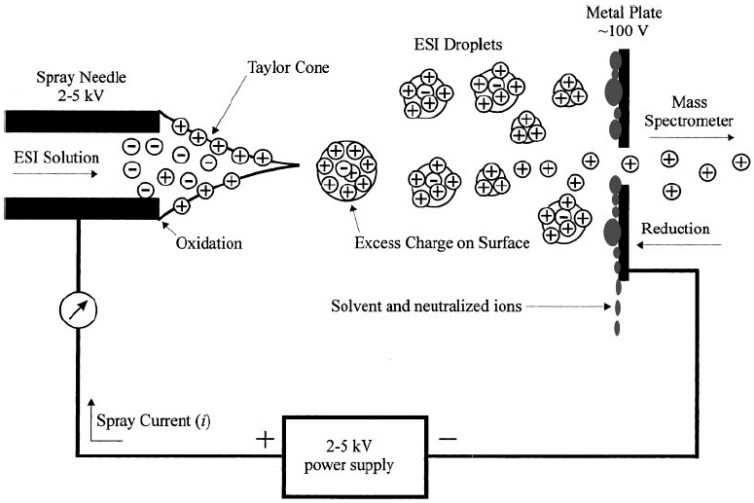


Figure 1.7: The electrospray ionisation process. The analyte solution is sprayed through a fine capillary needle to which a high voltage is applied (typically 4-5 kV). Between the needle and the counter electrode, an electric field gradient is generated, producing a Taylor cone with excess of positive charges on its surface. Charged droplets are formed at the tip of the Taylor cone, and as these droplets move towards the entrance of the mass analyser, the solvent evaporates, and Coulombic repulsion between the positive charges causes tiny droplets of free, charged analyte molecules to be formed for analysis [102, 106].

1.4.2 Mass analysers

The different types of mass analysers applied to proteomics analyses can be divided into two main groups: 1) Beam analysers, in which the ions leave the ion source in a beam, passing through the analytical field to the detector, and 2) trapping analysers, where ions may be generated within the analyser itself or be injected from an external ion source, and are trapped in the analysing field. TOF and quadrupolar analysers belong to the beam analyser category, while ion traps, FTICR and Orbitraps belong to the trapping analyser category [107].

Four types of mass analysers are commonly used for proteomics analyses:

1. Quadrupole (Q)
2. Ion trap (quadrupole ion trap, QIT; linear ion trap, LIT; or linear trap quadrupole (LTQ)
3. Time-of-flight (TOF), and
4. Fourier-transform ion cyclotron resonance (FTICR) mass analyser

They each apply different physical principles and show different analytical performance, but some hybrid types of instruments (eg. Q-TOF, TOF-TOF, and LTQ-FTICR) are designed to take advantage of the capabilities of different mass analysers [91].

1.4.2.1 Quadrupolar analysers

A quadrupole is a real mass-to-charge analyser, which does not depend on the kinetic energy of ions leaving the source. To separate ions by their mass-to-charge, a quadrupole analyser uses the stability of ion trajectories in oscillating electrical fields. The oscillating electrical fields are generated by four circular or hyperbolic rods arranged in parallel (see **Figure 1.8**). Two of the rods are affected by a positive

potential, and the other two rods by a negative potential. The potential of the individual rods changes consistently, making the ions travelling between them oscillate. Ions are accelerated along the z-axis between the rods, but also by forces induced by the electrical field along the x- and y-axes. The trajectory of the ions will remain stable as long as they never touch the rods [85].

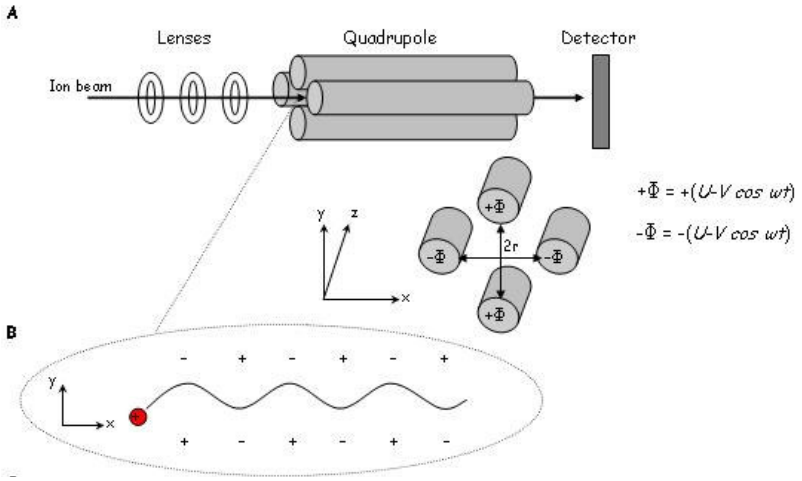


Figure 1.8: **Schematic illustration of a quadrupole mass analyser and the underlying principles**. (**A**) The ions pass a series of focusing lenses before entering the area between the rods in the quadrupole instrument. A variable potential, +Φ or –Φ, is applied to the rods, making the ions oscillate between them. In the equations for + Φ and –Φ, U is the direct potential, V the "zero to peak" amplitude of the radio frequency (RF) voltage, and ω the angular frequency. (**B**) A 2D illustration of the oscillating trajectory of a positive ion (**red dot**) in the area between the rods (Adapted from [85]).

Quadrupoles can be combined in series to allow MS/MS analyses. Triple quadrupole instruments are among the most wide-spread constructs, since their development in the late 1970s [108]. In these instruments, the first quadrupole provides full scan and ion isolation properties, the second quadrupole acts as a collision cell for ion fragmentation, and the final quadrupole provides full scan of the fragment ions [85]. This arrangement also allows for targeted quantitative analyses, e.g. of modified peptides [109-111]. For improved quantitative capabilities, the current generation of triple stage quadrupole (TSQ) instruments have increased sensitivity and signal-to-noise ratio, thus allowing detection of even lower amounts of the desired analyte [112, 113]. **Figure 1.9** below illustrates the build-up of a modern TSQ instrument.

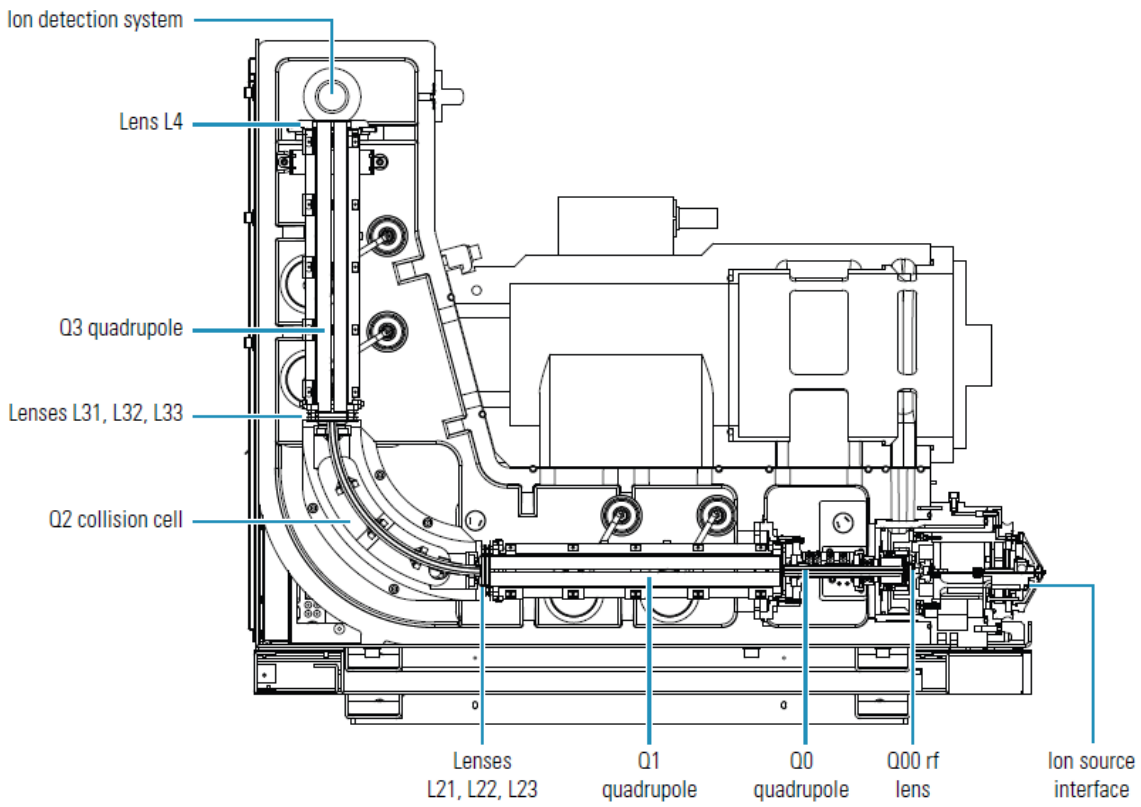


Figure 1.9: Illustration of a triple stage quadrupole (TSQ) instrument. Quadrupoles 1 and 3 (Q1 and Q3, respectively) are mass analysers, while the second quadrupole, Q2, is a collision cell. The Q0 quadrupole acts as an ion transmission device, where the electric field generated by RF voltage applied to the rods guides the ions along the quadrupole axis, increasing their translational kinetic energy. The lenses allocated throughout the instrument serve to focus the ions and regulate their translational kinetic energy as they pass through the different segments [114].

1.4.2.2 Time-of-flight analysers

The mass-to-charge separation of ions by time-of-flight (TOF) analysers is based on measurements of the ions' flight time through a vacuum flight tube. The mass of the ion influences the flight time, as heavier ions will move more slowly through the flight tube. TOF instruments have a high transmission efficiency leading to very high sensitivity. The resolution of TOF instruments depends on whether they are operated in linear or reflectron mode (**Figure 1.10 A** and **B**). Mass resolution is proportional to the flight time so the length of the flight tube as well as the acceleration of the ions play important roles for the achievement of high resolution [85].

TOF analysers are directly compatible with pulsed ionisation techniques like MALDI, but with some adaptation continuous ionisation techniques, such as ESI, can also take advantage of these powerful analysers. Instruments with ESI-sources are generally fitted with an orthogonal acceleration (oa) TOF (**Figure 1.10 C**). This allows conversion of the continuous ion flow into a pulsed one as the ions initially fill the space between the plate and first grid (G1), and when an injection pulsed voltage is applied to the plate, ions are accelerated orthogonally to their original trajectories [115].

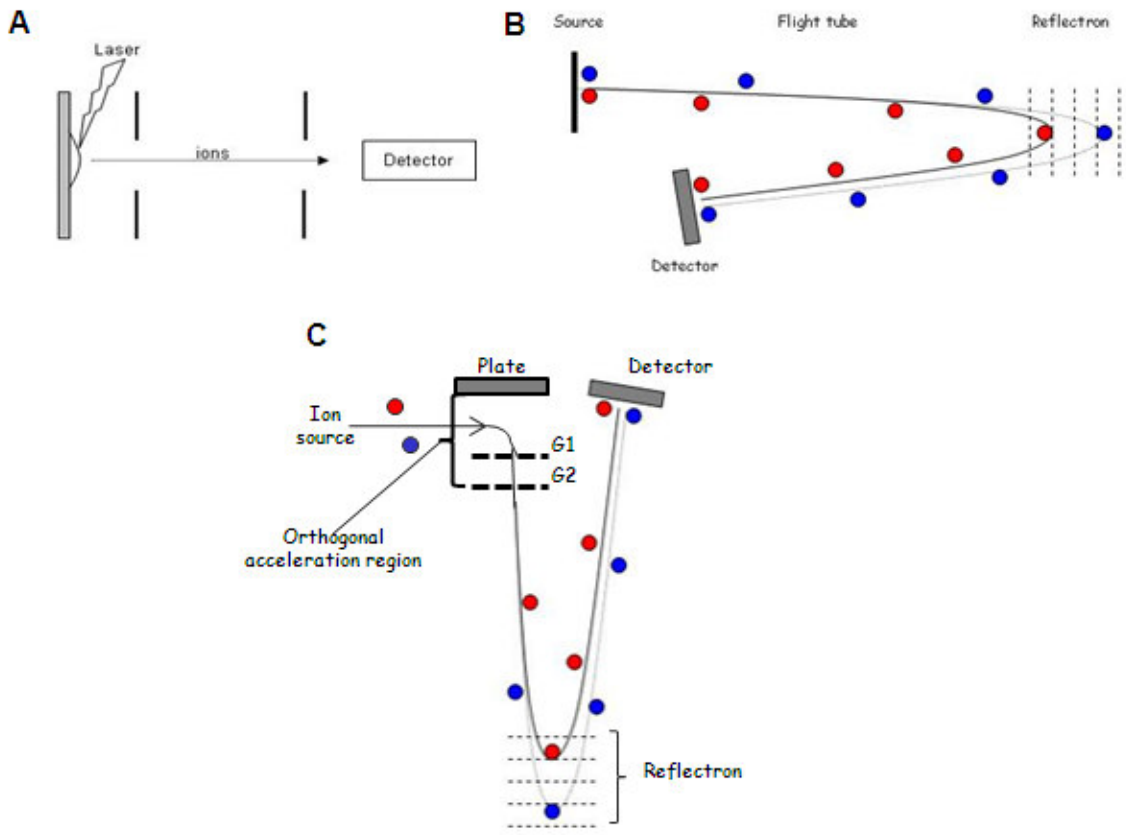


Figure 1.10: Linear (A) and reflector (B) detection modes for ions in a MALDI TOF mass spectrometer, and orthogonal TOF for ESI instruments (C). (A) For linear detection in a MALDI TOF instrument, ions follow a straight path through the TOF analyser. (B) Reflector detection is achieved by the addition of an electrostatic ion mirror, a reflector, in the mass analyser. The kinetic energy of the ions determines how deep into the electrostatic mirror they will travel before being deflected. This allows for correction of the kinetic energy dispersion between ions of identical molecular mass. This is illustrated by the **red** and **blue dots**, representing ions of identical mass but different initial kinetic energy. Because the **red ion** possesses less kinetic energy it will not travel as deep into the electrostatic mirror as the **blue ion**, hence being able to catch up, so that both ions will arrive at the detector simultaneously. In an orthogonal set-up (**C**), commonly used for ESI instruments, a pulsed injection voltage is applied to the plate, for ions to be accelerated orthogonally to their original trajectory (inspired by [85] and [115]).

1.4.2.3 Linear ion traps

A linear ion trap consists of two end-cap electrodes with a ring electrode in between. A radio frequency (RF) voltage is applied to the ring electrode for generation of a three-dimensional quadrupolar electric field that will trap the ions. Generation of unstable ion trajectories in a mass-selective manner (by increasing the RF voltage applied) enables mass analysis. Towards the centre of the iontrap, ion trajectories can be dampened by collision with helium ions. Thus, ion traps are capable of both scanning and tandem or even MS^n analyses [107, 116, 117]. **Figure 1.11** illustrates the build-up of a quadrupole ion trap.

Ion traps resemble quadrupoles in terms of *m/z* range. However, ion traps are superior to quadrupoles in terms of sensitivity at higher *m/z* (sensitivity does not drop). The resolution provided by ion traps is moderate, compared to the relatively low one of quadrupoles, and can be significantly improved by changing the RF voltage scan rates [107].

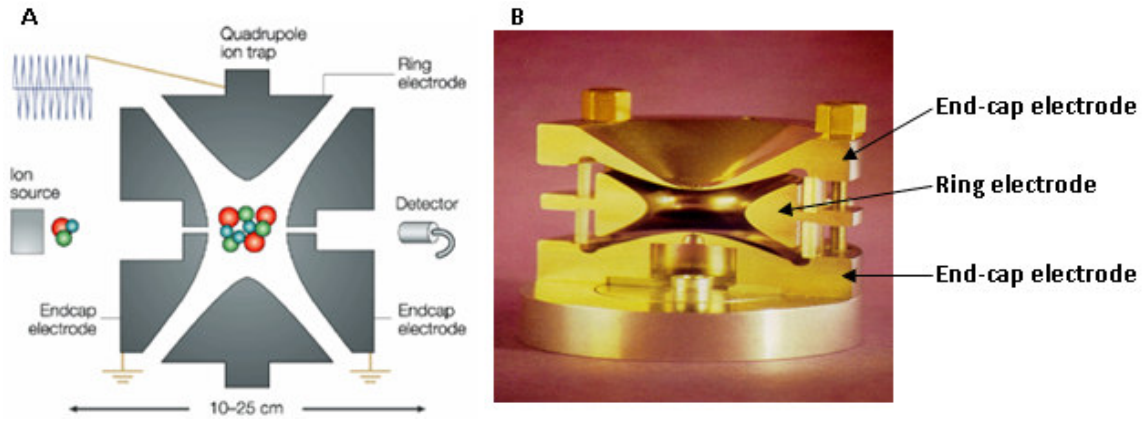


Figure 1.11: Ion trap mass analyser. A: Schematic overview of the quadrupolar setup of an ion trap. An RF voltage applied to the ring electrode generates an electric field where ions can be trapped. **B:** Picture of a cross-section of an ion trap, displaying the asymmetric trapping field between the quadrupolar setup of the electrodes. Modified from [107, 116].

1.4.2.4 Orbitrap analysers

Orbitraps are ion traps without RF or magnets to hold the ions inside [118]. In an Orbitrap, the incoming ions are trapped in an electrostatic field [119, 120]. The electrostatic field that attracts the ions towards the central electrode is compensated by a centrifugal force originating from the initial tangential velocity of the ions, causing them to move in complex spiral patterns similar to that of satellites in orbit [118]. The Orbitrap mass analyser consists of a spindle-like central electrode and a barrel-like outer electrode [121] as depicted in **Figure 1.12**. It is an analyser characterised by large dynamic range, high mass resolution, and consequently high mass accuracy (0.2-5 ppm, depending on signal-to-noise and whether calibration is performed internally or externally) [118, 121, 122].

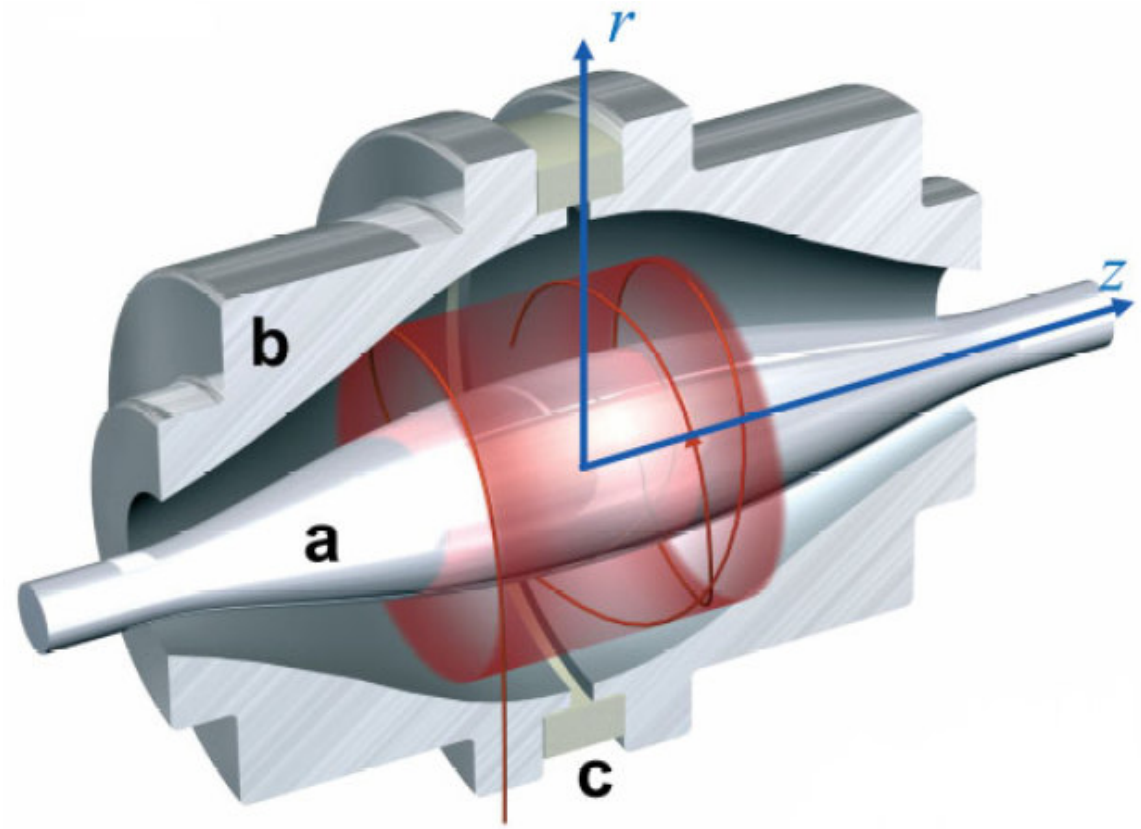


Figure 1.12: Cross section of an Orbitrap mass analyser. Ions are injected into the Orbitrap offset from its equator (z=0), perpendicular to the z-axis, and will move in spirals (red arrow line) around the central electrodes (**a**) without further excitation needed. The outer electrode (**b**) is split in half by an insulating ceramic ring (**c**). Ions of different mass and charge will have different oscillation frequencies, which can be determined by Fourier transformation to produce accurate *m/z* readings for the ions [118, 121].

1.4.3 Fragmentation

1.4.3.1 Tandem mass spectrometry

Fragmentation can take place in instruments with a single mass analyser via post-source decay, but this happens in a less controlled and selective manner, which can be challenging for subsequent data analysis [85, 102, 123-125]. To be able to fragment peptides, which is necessary to gain sequence and potential PTM information, the mass spectrometer needs to be fitted with two mass analysers. These are typically separated by a cell or compartment (trap) in which the actual fragmentation takes place. Such a setup enables activation of ions distinct from the ionisation step, and the precursor and product ions are characterised independently by their *m/z* ratios. The ion to be fragmented is selected in the first mass analyser, fragmented in the collision cell or trap, and its product ions separated by their *m/z* in the second mass analyser. The most common type of ion activation is collision-induced dissociation (CID). Here, collisions between the precursor ion and a neutral gas are accompanied by increased internal energy to induce fragmentation. This can be performed at both high and low energy in a wide range of instruments, including ESI Q-TOFs and LTQ-Orbitraps [91, 125]. **Figures 1.13** and **1.14** depict a Q-TOF type of instrument and an LTQ-Orbitrap, respectively, illustrating the position of the mass analysers and additional segments making up these instruments.

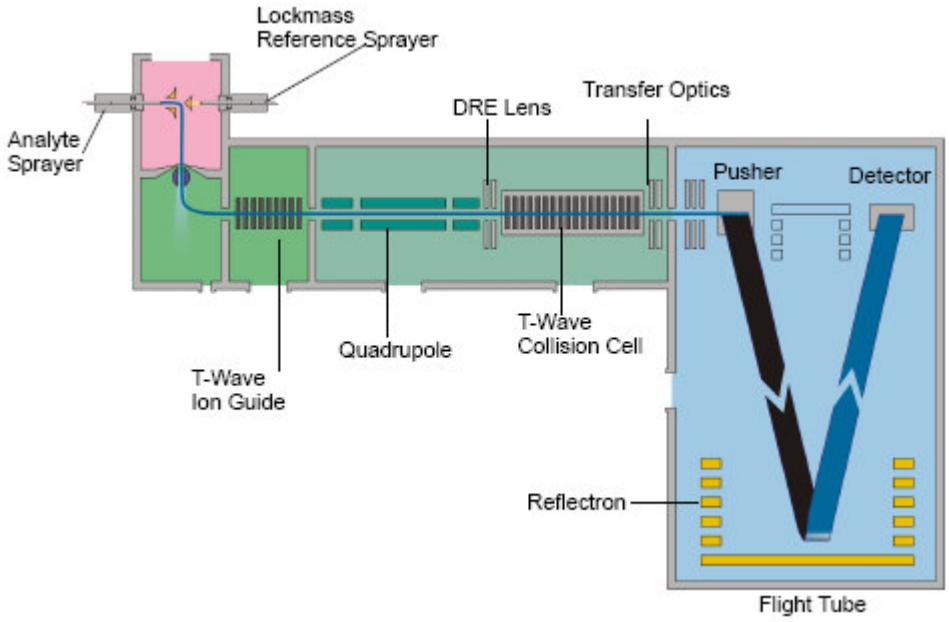


Figure 1.13: Illustration of the ESI Q-TOF Premier instrument. The analyte is introduced into the mass spectrometer via an electrospray source (pink segment) and guided to the first mass analyser, the quadrupole. In MS/MS mode, the quadrupole serves to isolate ions of interest and send them on to the collision cell (blue-green segment), where they can be fragmented. The fragment ions are then separated by their *m/z* in the orthogonal acceleration TOF (blue segment) [126].

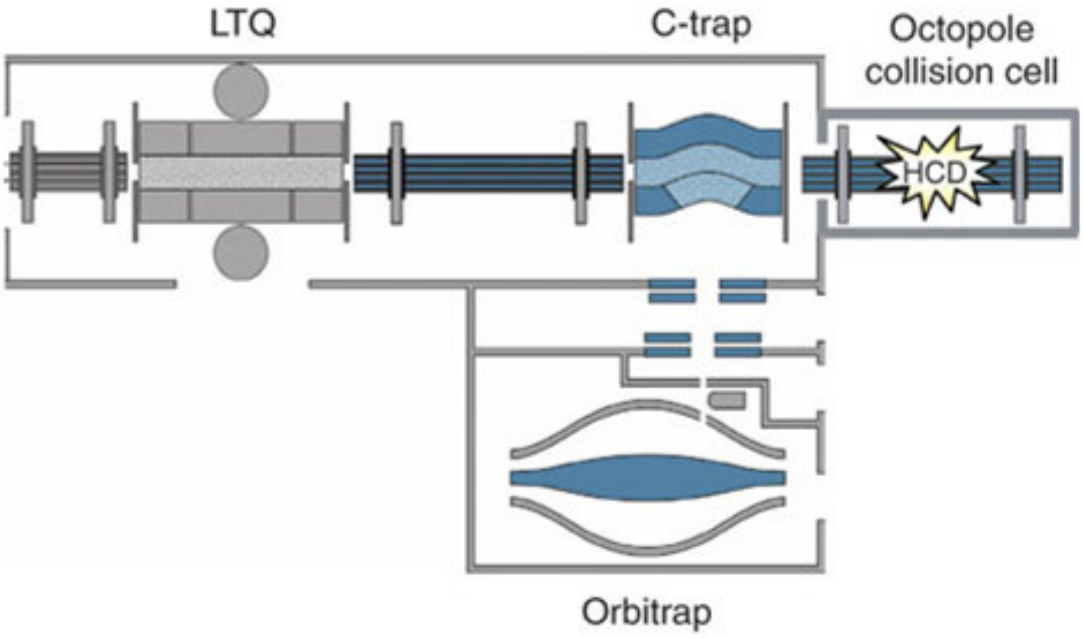


Figure 1.14: Illustration of the LTQ-Orbitrap XL instrument [121, 127]. Ions are introduced to the mass spectrometer by a MALDI or ESI (most common) type of source (far left). The ions enter the linear trap quadrupole (LTQ), are axially ejected and trapped in the C-trap. The C-trap is a curved RF-only quadrupole ion trap capable of radial rather than axial ion ejection, thus increasing the performance of the down-stream Orbitrap [121]. If high energy collision (C-trap) dissociation (HCD) is needed (e.g., for iTRAQ experiments), ions may be sent into the octopole collision cell before they are squeezed into a small cloud and injected into the Orbitrap for analysis [127].

1.4.3.2 MS^n

Fragmentation by tandem mass spectrometry may not always give the desired information about the analyte. Post-translational modifications like phosphorylations may significantly impair the ability to get decent sequence information for peptides analysed by traditional collision-induced dissociation as the phosphorylation is easily lost, giving rise to an intense signal (neutral loss) that will easily mask many sequence-specific signals.

To circumvent these limitations of traditional CID MS/MS, some instruments exploit different kinds of fragmentations, such as electron transfer dissociation (ETD) and electron capture dissociation (ECD) [128, 129], or may allow additional fragmentation (MS^n), where new precursor ions can be selected from the initial fragmentation spectra. A thorough exposition of all these is beyond the scope of this thesis, hence only the techniques (multistage activation and HCD) used for the majority of the results presented will be elaborated here.

1.4.3.2.1 Multistage activation (MSA)

Multistage activation is a pseudo-MS^3 procedure. For traditional MS^3 of phosphopeptides, the neutral loss ion of the MS^2 experiment would typically be the one selected for additional fragmentation. While this would likely give significantly more sequence information, the partial sequence information contained in the initial MS^2-spectrum of the peptide would be lost. Multistage activation allows a combination of the spectra obtained by MS^2 and MS^3 of a given peptide (see **Figure 1.15**).

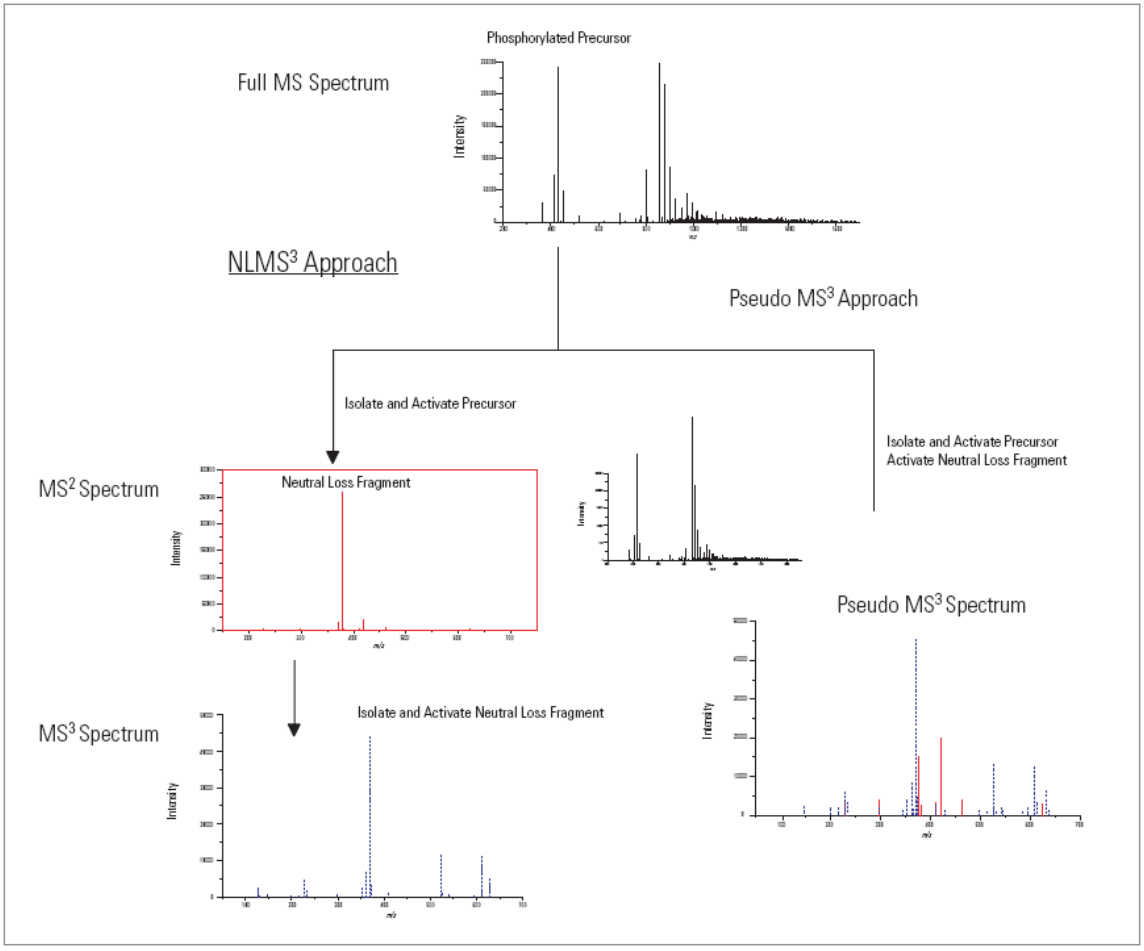


Figure 1.15: Multistage activation (pseudo MS^3) compared to traditional neutral loss MS^3. Where traditional neutral loss MS^3 is based on selection of the neutral loss ion from a separate MS^2 spectrum, multistage activation (MSA, pseudo-MS^3) allows simultaneous activation and fragmentation of neutral loss fragment ions detected in MS^2. This allows for generation of a combination of the traditional MS^2 and MS^3 spectra to increase sequence information in the resulting spectrum [130].

1.4.3.2.2 High energy collision dissociation (HCD)

The high energy collision dissociation fragmentation technique was originally introduced by Olsen *et al.* [127] as higher-energy C-trap dissociation because the C-trap of Orbitrap instruments plays a key role in it. In HCD, the C-trap is exploited as a collision cell by application of higher RF-voltage [127]. This enables detection of more fragment ions as compared to the outcome of typical linear ion trap fragmentation in an Orbitrap. HCD has proven advantageous in PTM analyses because it enables detection of PTM diagnostic ions [127, 131, 132]. It is also essential for certain quantitative analyses performed on Orbitrap instruments as detection of the low mass range iTRAQ reporter ions requires HCD [132].

1.5 Quantitative analyses

Differences at the protein and/or PTM level often impact the biological outcome and thus are of great interest research-wise. These differences need not just be the presence/absence of certain proteins or modifications, but are more often a matter of concentration differences (i.e., significant up- or down-regulation). Protein phosphorylation is a reversible, transient, and therefore highly dynamic PTM, that changes significantly over time, e.g. through the cell cycle [133] or the life stages of an organism [134], but also as a consequence of external perturbations [135] or disease [136]. Quantification can be either absolute [137, 138], where protein/peptide levels are calculated in the context of an internal standard of known concentration or relative, where changes in the protein/peptide levels in a sample cohort are calculated relative to the indexed levels in a reference sample [139]. Several factors influence the results obtained by mass spectrometry based proteomics and PTM analyses (e.g. ionisation of the analyte in question, varying detector responses, differences in pre-MS sample handling, column(s) and needle effects in LC-MS, etc.). This challenges quantitative MS analysis, yet different approaches are applicable, three of which were used in the current study. These will be introduced below, and additional information about quantitative mass spectrometry for proteomics and phosphoproteomics is available in the literature [140, 141].

1.5.1 Isotopic/isobaric labelling

Quantitative analyses by mass spectrometry are based on comparative analyses of peptide ion intensities from series of samples, either by a label-free approach (see sections **1.5.2.1** and **1.5.2.2**) or by stable isotopic/isobaric labelling approaches. Stable isotope labelling approaches allow multiplexed analysis of 2-8 samples. The label can be incorporated either metabolically (Stable Isotope Labelling by Amino acids in Cell culture, SILAC [142]), chemically (e.g., Isobaric Tags for Relative and Absolute Quantification, iTRAQ [143], or Tandem Mass Tags, TMT [144]), or enzymatically

(e.g. trypsin-catalysed ^{18}O-labelling [145]) [140]. Each type of labelling approach entails its own distinct advantages and limitations, but only those related to iTRAQ will be reviewed here.

1.5.1.1 iTRAQ

iTRAQ entered the quantitative mass spectrometry scene in 2004 [143]. It is based on amine-reactive isobaric tags [143], typically employed at the peptide level, thus entailing independence of the biological origin of the samples. Labelling is achieved by an *N*-hydroxysuccinimide (NHS) moiety that reacts with the free amine groups of the peptides, i.e., the N-terminus and lysine side chains [143, 146]. This implies that buffers used in all sample steps prior to iTRAQ labelling should be devoid of free amines, hence excluding ammonium bicarbonate as the typical buffer for protein digestion [146]. Apart from the NHS moiety, the iTRAQ label consists of a balancer and a reporter molecule [143]. The masses of the balancer and reporter molecules vary in such a way that the entire iTRAQ labelling complex will retain a constant mass (i.e., the higher the reporter molecule mass, the lesser the balancer molecule mass will be, and vice versa) [143] (see **Figure 1.16**). Originally employed in a 4-plex set-up, iTRAQ is now commercially available as 4-plex and 8-plex kits, enabling quantitative analyses of 2-8 samples. Since its introduction, iTRAQ has become an increasingly popular quantitative MS-approach, also for phosphoproteomics studies [147-149]. An advantage of iTRAQ in relation to phosphoproteomics studies is the fact that if enrichment of phosphopeptides by traditional approaches such as TiO_2, IMAC or SCX [148, 150, 151] is conducted after the labelling, it will simultaneously serve as a sample clean-up step prior to MS/MS analyses [148]. There are, however, also challenges and drawbacks to the use of iTRAQ. Occurrence of unrelated, near-isobaric peptides of low intensity co-eluting with precursor ions of interest skew reporter ion ratios [152-154] as reporter ion intensities in such cases represent more than a single peptide. Isotopic impurities may affect the dynamic range achievable by iTRAQ, as evaluated by Ow *et al.* [153], because the intensities of adjacent reporter ions affect each other, thus dampening any differences in

the levels of these. The most recent evidence of iTRAQ-related limitations is that isobarically labelled (iTRAQ and TMT) peptides and phosphopeptides undergo charge-enhancement, causing a significantly reduced identification potential [150]. Analyses of highly charged peptides require different fragmentation procedures, and the more effective procedure, ETD [155], may interfere with the number of samples that can be handled in an iTRAQ experiment [156].

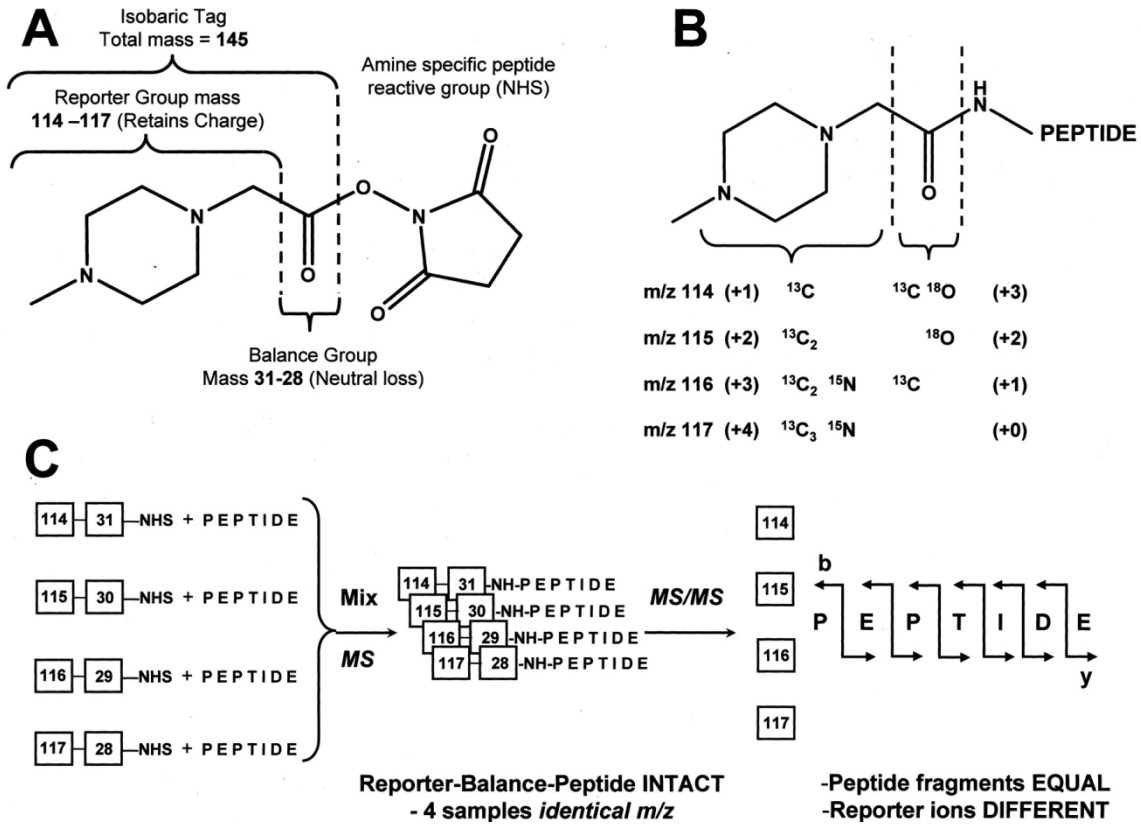


Figure 1.16: Illustration of iTRAQ labelling features [143]**.** The iTRAQ labels consist of three different groups, the amino specific peptide reactive group, a balancing group, and a reporter group, retaining the overall mass of the label by variations in the reporter and balance groups (**A**). The mass variations in these groups are achieved by differential isotopic atoms (^{13}C, ^{15}N, ^{18}O) enrichment (**B**). Identical peptides from 4 different situations labelled with different iTRAQ labels, can then be mixed and analysed by MS/MS in a single analysis (**C**). During MS/MS fragmentation, the reporter and balance groups are split from the peptide and each other to generate separate signals of *m/z* 114.1, 115.1, 116.1, and 117.1 along with peptide derived fragment ions in the spectrum. The intensities of the reporter ions reveal the relative levels of the analysed peptide in the four original situations.

1.5.2 Label-free approaches

Label-free quantification relies on reproducible, comparative LC-MS/MS analyses of samples. Therefore, it is applicable to all types of protein samples, and so is becoming more common in proteomics [157]. Yet, refinement of the label-free approaches is still needed, even though they enable a larger dynamic range of recorded protein/peptide changes than many of their labelling peers [158]. Label-free quantification is achieved by measuring changes in ion intensities (e.g., peptide peak areas or chromatography peak heights), or spectral counting (i.e., counting the number of spectra identifying a specific protein/peptide in the samples to quantify) [157]. While spectral counting is part of the Scaffold analyses conducted for evaluation of some of the data in this study (sections **3.1.1**, **3.1.2.3**, **3.2** and **3.5.4**), it was not used to extract quantitative information. For this, procedures based on ion intensities were employed, as will be described below.

1.5.2.1 Selected reaction monitoring, SRM

Selected reaction monitoring (SRM) is a targeted LC-MS/MS approach similar to selected ion monitoring (SIM), and multiple reaction monitoring (MRM), exploiting highly specific scans (precursor-to-product transitions). The analyses are carried out on triple quadrupole instruments (section **1.4.2.1**) where the peptide ion mass (precursor) is isolated in the first quadrupole mass analyser (Q1) for fragmentation in the second quadrupole. The third quadrupole mass analyser is then locked to only transmit specific fragment ions, transitions, of the given peptide. This way the Q1-Q3 transitions are dependent on efficient ionisation of the parent ion as well as subsequent efficient fragmentation of it, regardless how many transitions are assayed [110, 111]. In the resulting chromatogram this will display as an array of peaks (corresponding to the number of assayed transitions) of different intensities lying within one another (see **Figure 1.17**). While theoretical calculations may provide the required information, peptide/protein analyses by SRM typically benefit from preceding MS/MS analysis

(discovery analysis) to determine which fragment ions are the most intense and hence best suited as transitions. SRM has been used for highly sensitive identification of phosphorylation sites in one or more specific proteins [109, 159-161], but also provides the ability to perform quantitative analyses based on measurements on peptide or fragment ion peak areas [111, 162, 163]. The quantitative analyses, even at the phosphorylation level, may even require less starting material than traditional isotope-labelling LC-MS/MS experiments [163].

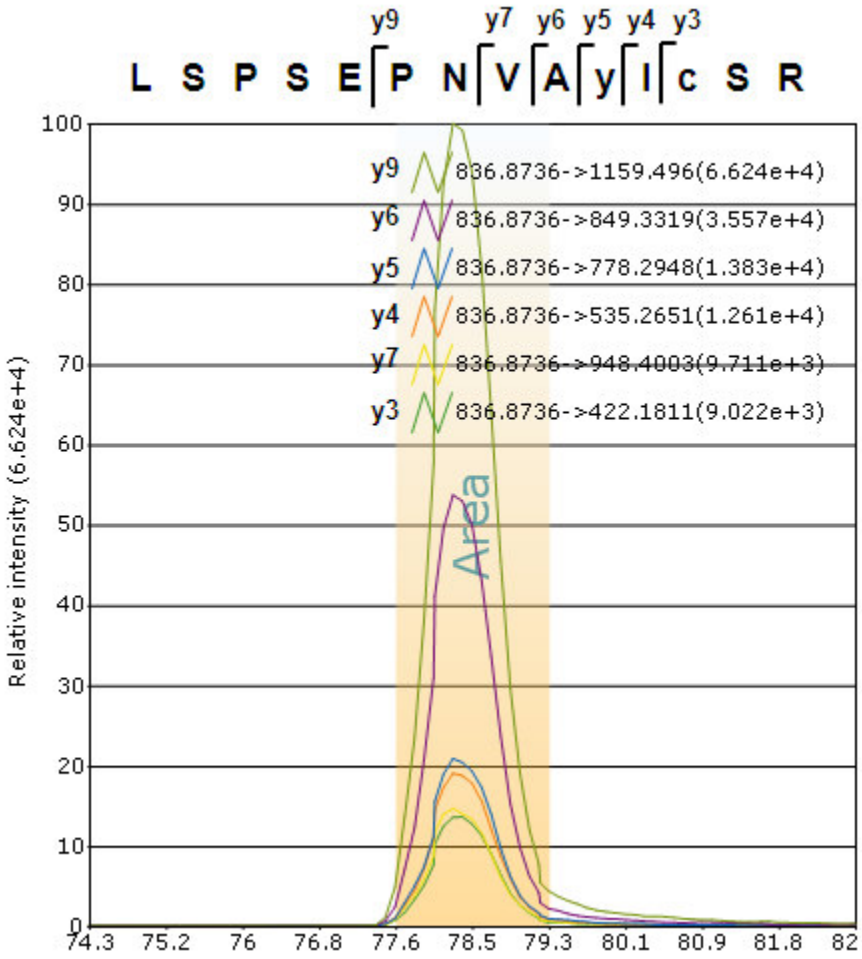


Figure 1.17: Array of SRM transitions in a chromatogram. SRM analysis of a phosphopeptide from GSK3β. Small case letters indicate modified amino acids (phosphorylation for the tyrosine residue, and carbamidomethylation for the cysteine). The fragment ions chosen as transitions are indicated in the sequence as well as in the chromatogram. Derived from the same precursor ion, all transitions have the same retention, thus generating the peak-within-peak display. The peak area, typically of the 2-3 most dominant transitions, can then be used to quantify the peptide.

1.5.2.2 MS^E

MS^E is a data independent MS approach enabling fragmentation of all ions in a survey scan and quantification of these by ion intensities of the intact peptides [164]. The method was first described in 2005 by Silva *et al.* [165]. Quantitative proteomics by MS^E is based on the finding that the MS signal intensity of the three most dominant tryptic peptides per mole of protein is constant within 10 % variation, which enables calculation of a universal signal response factor that can then be used to quantify every other sample constituent [138, 165]. Not only allowing identification and quantification of every peptide/protein within the dynamic range of the applied mass spectrometer, MS^E also holds other great potential applications for analyses of poorly characterised species. Based on the theory that the overall homology between related species is preserved at the protein level, MS^E offers the ability to improve protein identification/annotation in unsequenced organisms. This was elegantly demonstrated by Vissers *et al.* by comparative MS^E analyses of rabbit myocardium proteins and proteins from 5 other mammalian sample types (human heart, breast tissue, and plasma, rat liver and a mouse cell line) [166]. In contrast to the other included mammals, only few rabbit proteins had reached true protein database level, while the rest were still nested in a translated genome database. Similarities of curated and theoretical proteins/peptide by molecular weight, amino acid composition, and general abundance are translated into retention time and signal intensity similarities in the LC-MS analyses, enabling cross-identification and quantitation of proteins and peptides from related species [166].

1.6 Bioinformatics

Experimental design, construction of methods for mass spectrometry data acquisition as well as subsequent analysis and interpretation of the results, typically involves bioinformatics to some extent or another. In the current project, different software and web-tools were employed at varying extents. The most significant of these will briefly be described in this section.

1.6.1 GeneDB and TriTrypDB

The number and diversity of theoretically or experimentally characterised *Leishmania* proteins in the typical protein databases is relatively curtailed due to the low number of sequenced *Leishmania* genomes. GeneDB (http://www.genedb.org/Homepage) and TriTrypDB (http://tritrypdb.org/tritrypdb/) currently contain the most extensive records of *Leishmania* DNA and protein sequences, motifs and functional annotations. Both were employed in the processes of identifying *Leishmania mexicana* protein kinases and phosphatases from their *Leishmania major* orthologs, annotating *Leishmania mexicana* proteins to improve the database, and comparing protein kinase or phosphatase sequences between the available *Leishmania* species. These tasks were accomplished by the use of different Basic Local Alignment Search Tool (BLAST) analyses based on either nucleotide or amino acid sequences.

1.6.2 Proteome Discoverer

Processing and analysis of MS raw files from the iTRAQ experiments was performed in Proteome Discoverer (Thermo Scientific). This program enables generation of workflows for raw file processing alone or processing and Mascot database searching combined. The latter approach was used for the iTRAQ data to also let Proteome Discoverer normalise and calculate the iTRAQ ratios for the samples. As the iTRAQ raw files were acquired by a combination of MSA and HCD (see section **2.2.11.1**), the

workflow involved separate Mascot analyses of the spectra from each fragmentation procedure. The Proteome Discover output could subsequently be filtered by Mascot ion score, identification confidence etc. at either protein or peptide level for the final results report to only contain the most trustworthy identifications and quantifications.

1.6.3 Scaffold

Analyses of complex samples can challenge the ability to assess the data in the proper contexts, whether it be about characteristics of the identified proteins/peptides, reproducibility or differences between a number of samples. Different software solutions are available to assist in this process. In the current study, the Scaffold software (Proteome Software Inc., Portland, OR) was employed as it was the only software found to be capable of importing Mascot search results where the 6-frame translation library had been used as database. Scaffold was primarily used for comparative analyses (see sections **3.1.1**, **3.1.2.3**, and **3.2**). Mascot search results from the relevant samples were uploaded directly from the Mascot server, and by specifying which database had been used and how Scaffold should interpret the protein annotation information (see section **2.2.12.5**), the results would then be displayed in Scaffold for subsequent analyses. The display of the results can be customised in terms of the probability of correct identification and certain requirements for which identifications to display in the report (e.g. only proteins identified with at least 2 peptides and at least 1 phosphorylation site). Scaffold allows an easy overview of the identified peptides of every protein in a schematic overview accompanied by colour-coding of the protein sequence. It is also possible to evaluate the spectra of the individual peptides, and by use of Scaffold's spectral count feature get an overview of the abundance of proteins of interest in the different samples. Venn diagrams can be displayed to evaluate identified proteins, unique peptides or unique spectra of up to three samples at a time. If one of the typical proteomics databases (e.g. UniProt or NCBI) is used, Scaffold will also be able to display the distribution of GO annotations for the identified proteins.

Figure 1.18 – Legend on p. 54.

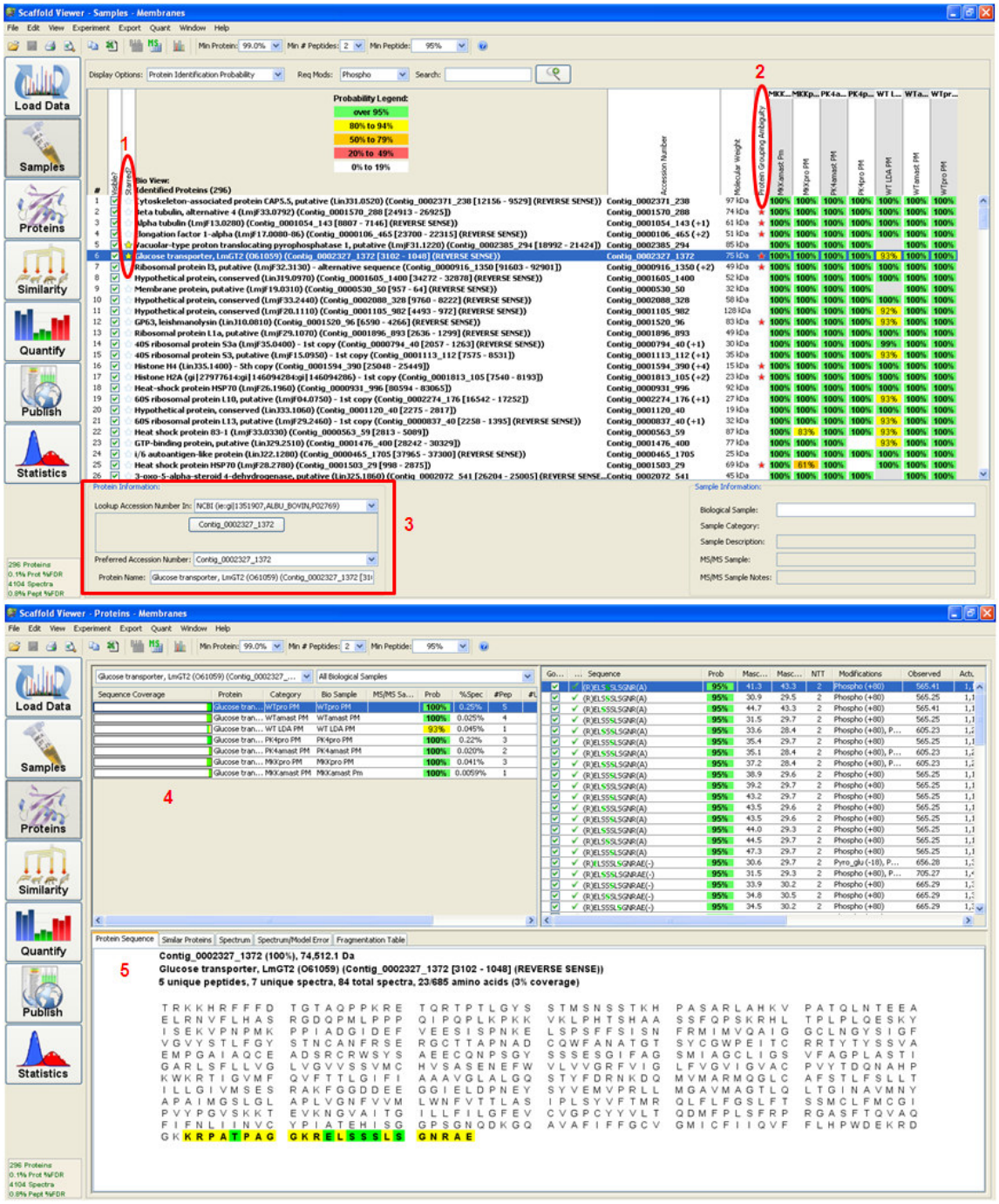


Figure 1.18: Proteomics search result display in Scaffold. The figure displays two illustrative screenshots from the Scaffold software. The top screenshot illustrates the labelling feature (1) of Scaffold, which enable customised filtering of the identified proteins. Scaffold also indicates whether there are other proteins with sequences similar to the identified proteins (2), and if so, these can be assessed individually. It is possible to edit the protein annotation as well as look up additional protein information in (3). Each identified protein can be assessed in the Proteins-menu (lower screenshot), and if multiple datasets are analysed, it is also possible to see a general comparison of the peptides and modifications identified for the protein in different datasets (4). Further sequence information, displaying the identified peptides and positions of modified amino acid residues, as well as other related information can then be found in the different tabs of result section (5).

1.6.4 Pinpoint

Selected reaction monitoring (SRM) analyses entail specification of which precursor and fragment ions the triple quadrupole instrument shall focus on. Different software solutions are applicable to construct methods for both sample and data analysis (e.g. MRMPilot™ (AB Sciex) [167], Skyline (MacCoss Lab Software) [168], and Pinpoint (Thermo Scientific) [169]. In the current study, Pinpoint was used to build transition lists containing the proteins of interest (peptide precursor ions and related fragment ions) to use for the mass spectral runs. Pinpoint enables transition list building based on either a theoretical digest approach or previous MS/MS analyses for a spectral library approach. The latter option obviously entails some advantages as to the choice of the most dominant transitions for the peptide(s) of interest; however the theoretical digest approach may be applicable for analyses of simple samples where the amount of material is limited and will not allow prior identification analyses. The SRM analyses conducted in the current study were all based on the spectral library approach. Proteome Discoverer reports of Mascot search results from initial LC-MS/MS analyses (discovery experiments) were uploaded into Pinpoint to serve as spectral libraries, facilitating inclusion of the relevant phosphopeptides and fragment ions. Because the data from the majority of the discovery experiments had been processed in DTASupercharge (see section **2.2.12.1**), which apply a slightly different procedure than Proteome Discoverer, discrepancies were found in the final results (e.g. certain peptides only identified from one of the processing procedures). These discrepancies triggered inclusion of phosphopeptides not formerly validated as well as manual inclusion of phosphopeptides that had been validated, but were not to be found in the Proteome Discoverer output. Once the transition lists have been generated they are exported in a CSV-format to allow instrument method set-up. The SRM analyses will then typically require a series of test analyses to optimise the instrument method. The test series involve modifications of the original Pinpoint file until it is suitable for the planned analyses. The acquired data can then be uploaded into the final Pinpoint method file in order to keep the peptide sequence and transition information relevant to the data when processing these. For protein and peptide quantification, a reference file is selected among the uploaded data

files, and Pinpoint will then process the data and return a results report displaying the protein and peptide ratios relative to the selected reference. The results report will also contain information about replicate variance in both a table format and a graphical view. Depending on the protein accessions, Pinpoint may also enable evaluation of the quantitative results in a biological context.

1.6.5 BioDesktop

The Protein Research group at SDU, Odense, is a motley collection of wet laboratory scientist, mass spectrometrists and bioinformaticians. The bioinformaticians assist, amongst other things, with tailored software to aid data analysis and interpretation. One such piece of tailored software is BioDesktop (in-house software developed by PhD-student Thomas Aarup Hansen, The Protein Research Group, SDU), which enables comparative analyses of Mascot search results by user-defined criteria, automatic calculation of Ascores for phosphopeptide validation as well as different MS fragmentation-specific analyses. The software utilise Mascot's customisable XML-file exports and applies specified algorithms to generate an output containing the results requested by the user. The results can then be exported to Excel for further analyses. In the current study, BioDesktop was mainly used for phosphopeptide Ascoring, where evaluation of the mass spectra in terms of the Ascored phosphorylation sites was an added benefit in cases where discrepancies between the Mascot- and Ascore-suggested phosphorylation site assignments were found.

1.7 Perspectives and aim of the study

At the beginning of the current study, few proteome analyses had been conducted in *Leishmania* [23, 100, 170-173]. The majority of these applied 2D gel electrophoresis (2DE) and subsequent mass spectrometry (MALDI or ESI LC-MS/MS) with relatively few specific identifications (29-75 proteins) [170, 171, 173]. Recent studies have increased these numbers significantly [174-176]. Yet, the *Leishmania* genome contains approximately 8,000 genes and likely not a drastically higher number of proteins due to the lack of gene introns and alternative mRNA splicing (other than *trans*-splicing for production of mature mRNA) [24, 25]. Comparing the expected number of proteins to the number of actual identifications, it is obvious that *Leishmania* proteomics is in majority unexplored land. Previous results of proteome analyses concurrently indicated significant inter-species differences in protein expression, emphasising the need for detailed species-specific analysis [170]. Differences in the protein phosphorylation profiles, kinase expression and activities in different stages of *Leishmania* development have been recognised [15, 16, 23, 44, 46, 100, 134, 175, 177-179], but studies have mainly concerned single specific proteins. While the outcome of reported proteomics studies in *Leishmania* has increased significantly just within the past couple of years, only very recently the same can be said about *Leishmania* phosphoproteomics [174].

The ambitious aim of the current study was to gain insight into the signalling pathways of *Leishmania mexicana* by comprehensive phosphoproteomics analyses in this species. It soon became apparent, though, that profound pathway mapping in *Leishmania mexicana* would not be the outcome of this study alone, thus requiring a redefinition of the aim. Hence, focus was placed on 1) a comprehensive phosphoproteomics analysis of protein kinases and phosphatases, both qualitatively and quantitatively; 2) generation of a general overview of proteins and potential phosphoproteins in *Leishmania mexicana*; and 3) in the context of the first two focus points to evaluate life stage-specific differences, as well as differences between wild type and specific MAP kinase and MAP kinase-kinase knock-out mutants. Because analyses of amastigotes involved both axenic

and lesion-derived parasites, important differences between the *in vitro* model and *in vivo* pathogen could also be assessed.

The phosphoproteomics analyses were carried out to evaluate the following hypotheses:

1. Protein phosphorylation in general, and of protein kinases and phosphatases in particular, differ between life stages. Differences will involve distinct phosphorylation patterns as well as phosphorylation abundance for the different phosphoproteins/-peptides.
2. The missing action of specific MAP kinase and MAP kinase kinases in viable knock-out mutant parasites will impair and change the protein phosphorylation pattern as compared to the wild type state. This will affect both the phosphorylation patterns and the overall amount of protein phosphorylation.
3. Differences between protein phosphorylation in axenic and lesion-derived amastigotes will be present, but should not be significant enough to render the model system useless for preliminary proteomics and phosphoproteomics analyses.

The proteomics analyses were conducted to provide some general protein background information in order to evaluate the qualitative and quantitative phosphoproteomics findings in a biological context.

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials

2.1.1 Solvents

For many of the experiments, stock solutions of the different reagents were prepared. Depending on the reagents, the solvents used for stock solutions would be either aqueous or organic. Aqueous solutions were prepared with either ddH_2O, HPLC water (Roth, Karlsruhe, Germany) or UHQ water from an ELGA Purelab ultra system (Holm & Halby A/S, Brøndby, Denmark). Organic solvents, acetonitrile (MeCN; HPLC grade) or ethanol (EtOH; analytical grade), were from Fisher Scientific (Loughborough, UK) and Rathburn (Walkerburn, Scotland), respectively.

2.1.2 Buffers

A host of different buffers were applied in the different harvest and lysis procedures. For convenience, the tables on the following pages list the contents of the different buffers, providing them with a short nickname that will be used in the subsequent descriptions of the procedures.

	Buffer A	Buffer B
	"Glasgow" wash buffer	"Glasgow" wash buffer + inhibitors
PBS	+	+
Na-ortho-vanadate (Sigma)	-	1 mM
Okadaic acid	-	0.1 µM
NaF (Merck)	-	10 mM
***o*-phenan-throline (Sigma)**	-	10 mM
PMSF (Roth)	-	2 mM
TLCK (Sigma)	-	25 µM
Leupeptin (Sigma)	-	50 µM

Table 2.1: Wash buffers for harvest by the "Glasgow" procedure. The buffer with inhibitors was prepared immediately before use and kept cold. The general wash buffer was also cold (4 °C) when used. PBS: Phosphate-buffered saline; PMSF: Poly-methyl sulphonyl fluoride, TLCK: Tosyl-L-lysine chloromethyl ketone.

	Buffer C	Buffer D
	"Odense" wash buffer	"Odense" wash buffer + inhibitors
H_2O*)	+	+
HEPES (Sigma)	21 mM, pH 7.5	21 mM, pH 7.5
NaCl (Merck)	137 mM	137 mM
KCl (Merck)	5 mM	5 mM
Na-Ortho-vanadate (Sigma)	-	1 mM
Okadaic acid	-	0.1 µM
NaF (Merck)	-	10 mM
o-phenan-throline (Sigma)	-	10 mM
PMSF (Roth)	-	2 mM
TLCK (Sigma)	-	25 µM
Leupeptin (Sigma)	-	50 µM

Table 2.2: Wash buffers for harvest by the "Odense" procedure. *) The water used was either ddH_2O or HPLC water (Roth, Karlsruhe, Germany). The buffer with inhibitors was prepared immediately before use. All buffers were cooled to 4 °C before use and kept on ice during the procedure. PMSF: Poly-methyl sulphonyl fluoride, TLCK: Tosyl-L-lysine chloromethyl ketone.

	Buffer E	Buffer F
	"Odense" lysis buffer	"Glasgow" lysis buffer
H_2O[A)]	+	+
Urea (Fluka)	7 M	6 M
Thiourea (Sigma)	2 M	2 M
Tris (Sigma)	40 mM	25 mM
Detergent	1 % n-octyl-β-D-glyco-pyranoside	4 % CHAPS
$MgCl_2$(Merck)	1 mM	1 mM
Benzonase	300 u	300 u
Na-Ortho-Vanadate (Sigma)	-	1 mM
Okadaic acid	-	0.1 µM
NaF (Merck)	-	10 mM
o-phenanthroline	1 mM	1 mM
EDTA-free protease inhibitor tablet (Roche)	A fraction of a tablet, depending on the buffer volume	A fraction of a tablet, depending on the buffer volume
PhosSTOP phosphatatase inhibitor tablet (Roche)	A fraction of a tablet, depending on the buffer volume	-
Na-Pervanadate[B)]	1 mM	-

Table 2.3: Lysis buffers. [A)] The water used was UHQ-water from an ELGA-water system (Holm & Halby A/S, Brøndby, Denmark); [B)] Sodium-pervanadate was generated from activated Na-orthovanadate (Sigma) by mixing a 100 mM Na-orthovanadate solution, pH 10.0, 1:1 with an 18% H_2O_2-solution (H_2O_2 30% (Sigma-Aldrich, Steinheim, Germany) adjusted to 18% by addition of UHQ water). CHAPS: 3-[(3-cholamidopropyl)dimethyl ammonio]-1-propanesulfonate.

	Buffer G
	Membrane protein extraction lysis buffer
H_2O[A)]	+
Sucrose (Serva Feinbiochemica)	255 mM
HEPES (Sigma)	20 mM
EDTA - Titriplex (Sigma)	40 mM
$MgCl_2$(Merck)	1 mM
Benzonase	300 u
***o*-phenanthroline**	1 mM
EDTA-free protease inhibitor tablet (Roche)	A fraction of a tablet, depending on the buffer volume
PhosSTOP phosphatatase inhibitor tablet (Roche)	A fraction of a tablet, depending on the buffer volume
Na-Pervanadate[B)]	1 mM

Table 2.4: Lysis buffer for membrane protein extraction. [A)] The water used was UHQ-water from an ELGA-water system (Holm & Halby A/S, Brøndby, Denmark); [B)] Sodium-pervanadate was generated from activated Na-orthovanadate (Sigma) by mixing a 100 mM Na-orthovanadate solution, pH 10.0, 1:1 with an 18% H_2O_2-solution (H_2O_2 30% (Sigma-Aldrich, Steinheim, Germany) adjusted to 18% by addition of UHQ water). A stock solution of the HES-buffer (HEPES, EDTA, sucrose) was stored at 4 °C. The appropriate volume of HES-buffer was mixed with the different inhibitors, enzymes and catalysts just prior to use.

	Buffer H	Buffer I
	LDA-tissue release buffer	Wash buffer
ddH_2O	+	+
HEPES (Sigma)	21 mM, pH 7.5	21 mM, pH 7.5
NaCl (Merck)	137 mM	137 mM
KCl (Merck)	5 mM	5 mM
Na-Ortho-vanadate (Sigma)	-	1 mM
Okadaic acid	-	0.1 µM
NaF (Merck)	-	10 mM
o-phenan-throline (Sigma)	-	10 mM
PMSF (Roth)	-	2 mM
TLCK (Sigma)	-	25 µM
Leupeptin (Sigma)	-	50 µM

Table 2.5: Buffers for isolation of lesion-derived amastigotes. Buffer H was filter sterilised before use. The buffer with inhibitors was prepared immediately before use. All buffers were cooled to 4 °C before use and kept on ice during the procedure. PMSF: Poly-methyl sulphonyl fluoride, TLCK: Tosyl-L-lysine chloromethyl ketone.

2.2 Methods

2.2.1 Parasite cultures

Promastigotes of *Leishmania mexicana* wild type strain WT0906 and MAP kinase kinase deletion mutants (ΔLmxMKK-/- K4, ΔLmxPK4-/- HN2) were cultured in SDM medium (PAN Biotech, Aidenbach, Germany) supplemented with 10 % heat-inactivated foetal calf serum (iFCS) (Sigma, Steinheim, Germany), 7.5 µg/ml hemin (Sigma, Steinheim, Germany) and 100 U/ml penicillin/100µg/ml streptomycin (Pen/Strep; Gibco, UK). Mutant strains with green-fluorescent protein-tagged MAP kinases (GFP MPK3 Hyg20, MPK5 GFP Puro40, PK4 GFP Puro40, and MKK GFP Puro40) were cultured in similar SDM medium, but additionally had 20 µg/ml hygromycin (GFP MPK3 Hyg20) or 40 µM puromycin (MPK5 GFP Puro40, PK4 GFP Puro40, and MKK GFP Puro40) added. Cultures were incubated at 27 °C.

For amastigote cultures, stationary phase promastigotes (> 5×10^7 cells/ml) were cultured in Schneider's Drosophila medium (PAN Biotech) supplemented with, 20 % iFCS (PAN Biotech), 2 mM L-glutamine, 100 U/ml penicillin/100µg/ml streptomycin, and 20 mM 2-morpholinoethanesulfonic acid monohydrate [MES] (Serva, Heidelberg, Germany) for a final pH of 5.5. Cultures were incubated at 34 °C, 5 % CO_2, for 72 hours.

MAP kinase deletion mutant promastigotes (ΔMPK3) were provided by Maja Erdmann [15, 44, 178]. Cultures were kept like WT promastigotes.

2.2.2 Lesion-derived amastigotes

For generation of lesion-derived amastigotes, 3×10^7 late-log phase (i.e. the cell culture density is 4-5 $\times 10^7$ cells/ml) *Leishmania mexicana* promastigotes were injected into the left hind foot pad of female Balb/c mice. Development of lesions was monitored, and once the lesion size had reached 5 mm, the mice were sacrificed, and amastigotes isolated from the lesion tissue by grinding through a metal net in ice-cold **Buffer H** (see

Table 2.5). The resulting parasite suspension was transferred to a 50 ml Falcon tube on ice, and passed through a 10 ml syringe fitted with a 25 g needle to release the intracellular amastigotes. Un-lysed macrophages and other cellular debris were pelleted by differentiation at 60 g in a Beckman GS-6KR centrifuge at 4 °C for 5 minutes. The supernatants were transferred to new tubes and subjected to filtration through 5 µm Millipore filters (Millex-SV). To remove smaller contaminants, the filtrates were pelleted by 1300 g centrifugation at 4 °C for 10 minutes (Beckman GS-6KR). The pellet was resuspended in **Buffer H**, diluted 1:100 or 1:200 and amastigotes counted by use of a haemocytometer. The cell suspension was transferred to Eppendorf tubes and subjected to centrifugation in a 10K centrifuge for 30 seconds. The supernatants were discarded and pellets resuspended in ice cold **Buffer I** (see **Table 2.5**). Another round of centrifugation was added, the supernatants discarded and the pellets snap-frozen in liquid nitrogen. The snap-frozen aliquots were stored at -80 °C until lysis could be carried out.

Mouse infections and subsequent lesion monitoring were carried out by Professor James Alexander and laboratory technician Helen McGachy. Preparation of cell pellets was performed by Dr. Martin Wiese, and all subsequent sample preparation by Heidi Rosenqvist.

2.2.3 Cell count

For counting, 10 µl promastigote cell culture was typically mixed with 490 µl fixing solution (9:1, PBS:formaldehyde), and 2×10 µl of the mixture taken out to fill two counting chambers. For more or less dense cultures, the initial dilution in fixing solution was modified.

In contrast to the promastigote cultures where a few µl could easily be taken out for cell count prior to initiation of the harvest procedure, axenic amastigotes tended to cluster

too much for this to be feasible. Hence, the axenic amastigote cultures were transferred to 50 ml Falcon tubes and spun down at 2000 g, 4 °C for 10 min (Thermo Scientific Heraeus Multifuge 3 SR+ centrifuge). The supernatant was removed and the cell pellet redissolved in ice cold HEPES or PBS, collecting the contents of the Falcon tubes into one single tube pr. original culture. 400 µl of the cell suspension were taken out into an Eppendorf tube and passaged at least 6 times through a syringe (29 G ½″) before 10 µl were taken out into 90 µl fixing solution. The cells in this solution were diluted 50 times in fixing solution before counting.

While counting, the remaining amastigotes were kept in HEPES or PBS on ice, and the harvest procedure was continued like described below as soon as cell counts were ready.

2.2.4 Total protein extraction

2.2.4.1 Harvest and lysis, "Glasgow"-procedure

Cells were counted and harvested at a density of approximately 5×10^7 cells/ml. During harvest cells were consistently kept cold and the workflow kept swift. Cultures were transferred to 50 ml Falcon tubes and spun for 10 min at 4 °C (2000 g; Thermo Scientific Heraeus Multifuge 3 SR+ centrifuge). The supernatant was removed, cells respuspended in **Buffer A** (see **Table 2.1**), and aliquoted to $5 \times 10^8 – 1 \times 10^9$ cells/tube. Cells were spun down at 4 °C for 10 min (2000 g; Thermo Scientific Heraeus Multifuge 3 SR+ centrifuge), the supernatant removed and replaced with another volume of **Buffer A**. This was repeated, replacing **Buffer A** with **Buffer B**. The supernatants were carefully removed by micropipetting. Cells were snap-frozen in liquid nitrogen or redissolved in **Buffer F** and sonicated (Branson Sonifier) 3 × 1 sec with 1 min cooling on ice in between each sonication. Samples were incubated at room temperature for 10 min at gentle rotations on rolling rods, and then centrifuged for 10 min at 17,950 g to remove insoluble material. The supernatants were transferred to fresh 15 ml Falcon tubes.

2.2.4.2 Protein precipitation, "Glasgow"-procedure

Ice cold acetone was added in 4-times excess and samples were vortexed shortly prior to incubation at -20 °C for 2 h. Precipitated proteins were spun down in a step-wise manner at 4 °C, 10 min, 17,950 g. The precipitate was washed in 4 times excess of ice cold 80% acetone and spun down at 17,950 g (13,000 rpm) for 10 min at 4 °C. The acetone was removed and samples allowed to "air-dry" shortly on the bench. Precipitates were either subjected to immediate trypsin digestion (section **2.2.6**) or stored at -20 °C for later handling.

2.2.4.3 Harvest and lysis, "Odense"-procedure

Cells were counted and harvested at a density of approximately 5×10^7 cells/ml. During the harvest, cells were consistently kept cold and the workflow kept swift. Cultures were transferred to 50 ml Falcon tubes and spun at 2,000 g for 10 min at 4 °C (Thermo Scientific Heraeus Multifuge 3 SR+ centrifuge). The supernatant was removed, cells resuspended in **Buffer C** (see **Table 2.2**), and aliquoted to 5×10^8 - 1×10^9 cells/tube. Cells were spun down at 2000 g, 4 °C, for 10 minutes (Thermo Scientific Heraeus Multifuge 3 SR+ centrifuge), the buffer removed and replaced with another volume of **Buffer C**. After another round of centrifugation at 2,000 g, 4 °C, 10 min, cells were washed in **Buffer D** (see **Table 2.2**). The supernatants were carefully removed by micropipetting. At this step, aliquots were either snap-frozen in liquid nitrogen, or immediately lysed by addition of **Buffer E** (see **Table 2.3**), followed by 3 × 5 seconds of sonication (Branson Sonifier) on ice. Samples were incubated at -80 °C for at least 30 min. After defrosting, dithiothreitol (DTT) was added to a final concentration of 20 mM, and samples were incubated at 56 °C for 40 min. Samples were allowed to cool to room temperature before iodoacetamide was added to a final concentration of 40 mM, and samples incubated at room temperature in the dark for 40 min.

2.2.4.4 Protein precipitation, "Odense"-procedure

Proteins were precipitated from the cell lysates by EtOH/acetone precipitation. Ice cold EtOH was added at 4 times the sample volume, and mixed well with the sample by vortexing. The same volume of ice cold acetone was added, again mixing well by vortexing. Samples were incubated at -20 °C over night.

After over night precipitation, precipitates were spun down at 20,000g (Eppendorf 5417R centrifuge) at 4 °C for 15 min. The supernatant was discarded and pellet washed twice in 100% ice cold acetone and once in 80% ice cold acetone, spinning down at 20,000g (Eppendorf 5417R centrifuge), 4 °C, for 15 min after the final washing step. The remaining supernatant was removed and samples allowed to briefly air dry before carrying on with protein digestion.

2.2.5 Membrane protein extraction

Parasites (WTpro, WTamast, WT_{LDA}, ΔPK4pro, ΔPK4amast, ΔMKKpro, ΔMKKamast) were harvested by the "Odense" protocol (section **2.2.4.3**) and snap-frozen. For protein extraction from membrane-cytosol fractions, parasites were resuspended in buffer G (**Table 2.4**). Samples were sonicated for 3 × 15 sec on ice, and mitochondria and nuclei removed by centrifugation at 20,800 g, 4 °C, 10 min (Eppendorf 5417R centrifuge). Plasma membrane and cytosolic protein fractions were separated by ultracentrifugation (245,000 g (Sorvall M150 GX) for 2 h, resulting in one or two liquid cytosolic phases (see section **3.5.4**) as well as a membrane pellet. The liquid fractions were transferred to clean Eppendorf tubes for reduction, alkylation and protein precipitation by a procedure similar to the one employed for the total protein extracts (section **2.2.4.3**). The membrane pellet was subjected to carbonate wash with incubation in 100 mM ice cold Na_2CO_3 on ice for 45-60 min with occasional vortexing. After incubation, plasma membranes were harvested by centrifugation at 100,000 g for 30 min at 4 °C. The supernatant was discarded, and the membrane pellet carefully washed in 500 mM NH_4HCO_3 (for general LC-MS/MS analyses) or TEAB (for iTRAQ experiments) once,

and in 50 mM NH_4HCO_3 or TEAB once. Reduction and alkylation of the membrane proteins could now be carried out, similar to the procedure described in section **2.2.4.3**.

2.2.6 In-solution digestion

Precipitated proteins were redissolved in 6 M urea (Fluka, Sigma-Aldrich, Steinheim, Germany)/2M thiourea (Sigma, Steinheim, Germany) by vigorous pipetting, gentle grinding with a single-use spatula (only necessary for protein precipitates that had been stored at -20°C after acetone removal), and sonication in a water bath. If not done prior to precipitation (i.e. samples prepared by the "Glasgow" procedure), samples were reduced and alkylated by addition of DTT to a final concentration of 20 mM and incubation at 56 °C for 40 min, then addition of iodoacetamide to a final concentration of 40 mM and incubation in the dark at room temperature for 40 min. The samples were diluted 5 times with 50 mM NH_4HCO_3 (Sigma, Steinheim, Germany) and trypsin (NOVO) was added at 1:75 (enzyme-to-protein) ratios. Samples were incubated at 37 °C over night. After overnight incubation, digestion was quenched by addition of 100 % formic acid (FA, analytical grade; Merck, Damstadt, Germany) to a final concentration of 5 %. Digests not used for immediate purification and analyses were stored at -20 °C.

2.2.7 TiO_2 purification

A slurry of TiO_2 material (Titansphere, 5 µm; GL Sciences, Tokyo, Japan) was prepared in 100% MeCN and packed upon a plug of C8 Empore Disc material (3M Bioanalytical Technologies, St. Paul, MN, USA) in an ordinary p200 pipette tip. The length of the TiO_2-part of the column was approximately 5 mm. The sample was mixed thoroughly in a 5-times sample volume of loading solution (1 M glycolic acid (Fluka, Sigma, St. Louis, MO, USA) in 80% MeCN/5% TFA (sequencing grade; Aldrich, Steinheim, Germany)) and loaded onto the TiO_2 column by applying gentle air pressure through a plastic syringe fitted to the tip. The flow-through from the column was collected in a clean Eppendorf-tube and dried down in a vacuum centrifuge. The TiO_2 column was

washed in turns with loading solution, washing solution (80% MeCN/1% TFA), and UHQ water. Bound phosphopeptides were eluted to a clean Eppendorf-tube by H_2O, pH 10.5 adjusted with ammonium hydroxide solution (Sigma, Steinheim, Germany). To ensure that no phosphopeptides had been trapped within the C8 material, an additional elution with just a droplet of 30% MeCN was performed. The eluate was acidified by addition of 100% formic acid (FA) to a final concentration of 10% FA.

2.2.8 SIMAC: Sequential elution from Immobilised Metal Affinity Chromatography (IMAC)

PHOS-select™ beads (Sigma) were washed twice in 0.1% TFA and once in SIMAC loading buffer (50% MeCN, 0.1% TFA). 50 µl beads were used per 200 µg protein, and the volume of sample aliquots had a maximum of 100 µl (if necessary, volume reduction had been performed in a vacuum centrifuge prior to SIMAC). The samples were acidified to a final concentration of 0.1% TFA, and combined with the PHOS-select™ beads plus a 10 times access of SIMAC loading buffer. The samples were incubated at room temperature for 1 h with rotations at 18 rpm. After incubation, beads were spun down (10 min, 20,000 g, room temp., Eppendorf Centrifuge 5417R). The majority of the supernatant, constituting the first monophosphorylated fraction, was transferred to a new tube, and the beads resuspended in the remaining supernatant. The beads were packed into a column in a large gel-loader tip. The flow-through of the column-packing constituted the second monophosphorylated fraction. The columns were washed with SIMAC loading buffer for collection of the third monophosphorylated fraction. Columns were washed with SIMAC elution buffer 1 (20% MeCN, 1% TFA) to release the remaining monophosphorylated peptides. Finally, the multiphosphorylated peptides were eluted by 5% ammonia water, pH11.

The multiphosphorylated peptide fractions were desalted on C8/Oligo R3 (PerSeptive Biosystems) by the procedure described below (2.2.9.1 “Sample prep LC-MS/MS”) and analysed on LC-MS/MS. The different monophosphorylated peptide fractions were

volume reduced in a vacuum centrifuge, then enriched for phosphopeptides by TiO_2 as described in section **2.2.7**.

2.2.9 ESI LC-MS/MS analyses

2.2.9.1 Sample preparation LC-MS/MS

Samples for LC-MS/MS were prepared by desalting on STop And Go Extraction (STAGE)-tips [180]. A plug of C8 (phosphopeptides) or C18 (non-phosphorylated peptides) Empore disc (3M Bioanalytical Technologies, St. Paul, MN, USA) was placed in the thin end of an ordinary p200 pipette tip and washed in 100% MeCN. The reversed-phase (RP) material was equilibrated with 5% FA and the sample loaded in a 1:1 sample-to-5% FA ratio. Peptides were eluted to a clean microcentrifuge tube by 70% MeCN/5% FA and dried down in a vacuum centrifuge.

2.2.9.2 LC-MS/MS

Tryptic digests of pro- and amastigote lysates with or without enrichment for phosphopeptides were analysed on an LTQ-Orbitrap XL (Thermo Fisher Scientific, Bremen, Germany), or alternatively on a Q-TOF Premier (Waters Micromass, Manchester, UK), ESI-instrument directly coupled to an Easy-nLC™-system (Proxeon, Odense, Denmark).

Samples were redissolved in 0.4 µl 100% FA, and quickly diluted with at least 5 µl Solvent A (0.1% FA). 5 µl of the dissolved sample were transferred to a micro-titer plate and inserted in the Easy-nLC™ autosampler.

The Easy-nLC™ was fitted with a home-made analytical column (50 µm i.d.; 100 µm o.d., 15-20 cm) (ReproSil-Pur C18-AQ, 3 µm; Dr. Maisch, GmbH, Ammerbuch-Entringen, Germany), and operated with intelligent flow control (IFC). On the LTQ Orbitrap XL instruments, a gradient of 133 min with a flow rate of 250 nL/min was

employed with the peptides eluting at 0-34% MeCN in 0.1% FA from 0-100 min, and 34-100% MeCN from 100-133 min. MS analysis was carried out with an FTMS scan upon which the 10 most abundant peptides were selected for MS/MS analysis with dynamic exclusion time of 45 sec[1]. Multi-stage activation (MSA) was employed for increased sequence coverage of the phosphorylated peptides. For analyses on the Q-TOF premier, the peptides were loaded in 0.1 % FA. Peptides were eluted by a gradient of 0-38 % MeCN in 0.1 % FA over 30 min (phosphopeptides) or 100 min (un-phosphorylated peptides) at a flow rate of 250 nL/min. MS analysis was carried out with 30 sec survey MS scan upon which the 5 most abundant peptides were selected for MS/MS analysis with dynamic exclusion time of 45 sec.

2.2.10 Determination of protein concentration

2.2.10.1 Qubit™ assay

All Qubit™ protein assay solutions (Invitrogen, Eugene, Oregon, USA) were equilibrated to room temperature. The required amount of working solution was calculated (based on the number of samples, incl. standards, to analyse), and the Quant-iT™ working solution prepared by mixing of Quant-iT™ reagent and Quant-iT™ buffer 1:200 (v/v). Samples (1-20 µl) and standards (10 µl) were mixed with the Quant-iT™ working solution for a final volume of 200 µl, and vortexed for 2-5 sec. Sample and standard mixes were incubated at room temperature for 15 min before carrying out the concentration measurements on a Qubit® fluorometer. Protein concentrations in the samples could then be calculated.

---

[1] Dynamic exclusion time refers to the fact that the *m/z* values selected for fragmentation in an MS scan subsequently are excluded (i.e., they cannot be selected for fragmentation) for a period of time, while new *m/z* values are selected for fragmentation in the next MS scans. This ensures selection of less abundant peaks too, as the dominant ones that may occur in several subsequent MS scans are excluded from fragmentation once they have been selected in the first scan. If similar peaks occur in an MS scan after expiration of the exclusion time, they can be selected for fragmentation again.

2.2.10.2 Amino acid analysis (AAA)

Protein digests thought to correspond to approximately 3 µg protein were dried down in low-bind Eppendorf tubes. Samples were resuspended in acid hydrolysis solvents (6 N HCl, 0.1% phenol, 0.1% thioglycolic acid). The tubes were placed in specialised glas vial in which pressure was reduced by Argon, and incubated at 110 °C for 20 h. The hydrolysed samples were analysed on a BioChrome 30 amino acid analyser (BioChrome Ltd., Cambridge, UK) by sodium ion exchange chromatography, post-column ninhydrin derivatisation and dual-wavelength (570 nm and 440 nm) detection. The output was processed in AAAproject.

2.2.11 Quantitative analyses

2.2.11.1 iTRAQ

Amino acid analyses were carried out on tryptic digests of the clear cytosol fractions from promastigotes, axenic amastigotes and lesion-derived amastigotes to determine their concentration. Peptide labelling was performed with 4-plex iTRAQ kits (Thermo Scientific). Aliquots of 100 µg protein digest were lyophilised for reconstitution in 30 µl iTRAQ dissolution buffer. iTRAQ labels were spun down, mixed with 70 µl EtOH, vortexed and spun down again. Samples and labels were mixed as shown in **Table 2.6**. The mixtures were vortexed for 1 min, spun down and incubated at room temperature for 1 hour. After incubation, the labelled samples were spun down at 13,200g for 15 min at room temperature. Sample labelling was checked by MALDI TOF MS/MS before mixing the samples 1:1:1:1. A volume corresponding to 1-2 µg protein was taken out for strong cation exchange (SCX) to enable quantification of unphosphorylated proteins as well.

Label	Promastigotes	Axenic amastigotes	Lesion-derived amastigotes
114	WTpro WCL	WTpro WCL	-
115	ΔPK4pro clear cytosol	ΔPK4amast clear cytosol	WTpro WCL
116	WTpro clear cytosol	WTamast clear cytosol	WT_{LDA} cytosol
117	ΔMKKpro clear cytosol	ΔMKKamast clear cytosol	WTamast clear cytosol

Table 2.6: iTRAQ labels used for the different sample types for quantitative analyses of phosphopeptides in the cytosol fractions. WTpro WCL, whole cell lysate of wild type promastigotes; ΔPK4pro, PK4 deletion mutant promastigotes; WTpro, wild type promastigotes; ΔMKKpro, MKK deletion mutant promastigotes; ΔPK4amast, PK4 deletion mutant amastigotes; WTamast, wild type amastigotes; ΔMKKamast, MKK deletion mutant amastigotes; WT_{LDA}, wild type lesion-derived amastigotes.

The iTRAQ sample mix was enriched for phosphopeptides by TiO_2 chromatography as described in section **2.2.7**. TiO_2 eluates were prepared for LC-MS/MS as described in section **2.2.9.1**, and analysed on an LTQ-Orbitrap XL. The 10 most abundant peptides of the MS survey scan were selected for fragmentation by MSA and high-energy collision dissociation (HCD) to allow good sequence coverage of the phosphopeptides as well as evaluation of the iTRAQ reporter ion intensities in the low mass area.

TiO_2 flow through as well as raw iTRAQ sample mix was subjected to strong cation exchange (SCX) chromatography. Samples were vacuum dried and resuspended in 30% MeCN/1% acetic acid. Sample pH was checked and adjusted if it was not around 2.7. SCX S20 (PerSeptive Biosystems) material was packed in p200 tips to a column length of approximately 6 mm. The column was sequentially washed in 100 % MeCN, 10 mM KH_2PO_4/1% acetic acid, and 30% MeCN/1% acetic acid. Samples were loaded slowly by applying passive air pressure (a plastic syringe with the plunger pulled back fitted to the column and left on the bench for slow sample diffusion). Column was washed with 30% MeCN/1 % acetic acid, and peptides eluted by increasing concentrations of KCl in 30% MeCN/1% acetic acid (25 mM, 50 mM, 100 mM, 250 mM, 500 mM, 750 mM, and 1 M). Eluates were dried down and desalted on C18 stage tips as described in section **2.2.9.1**.

The iTRAQ samples were analysed by LC-MS/MS on an LTQ-Orbitrap XL instrument set up to apply a combination of MSA and HCD in the fragmentation process. Precursor ion selection and other instrumental settings (liquid chromatography as well as mass spectrometry) were similar to what was described in section **2.2.9.2**.

2.2.11.2 SRM

The SRM methods, choosing proteins, peptides and fragment ions to consider, were built in Thermo Pinpoint 1.0.0 (Thermo Scientific). The methods were tested on a mixture of

WTpro and WTamast TiO_2 eluates, and optimised accordingly (i.e., including only peptides/transitions that were detected in the tests, increasing sensitivity by including timed windows for when along the LC-gradient the peptides/transitions would appear, etc.). LC-MS/MS analyses were carried out with a two-column set-up on an Easy-nLC™ coupled on-line to a TSQ Vantage instrument (Thermo Scientific). The two-column set-up consisted of a trap column (ReproSil-Pur C18-AQ, 5 µm; Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) and an analytical column of the same type as described in **2.2.9.2**.

Triplicate analyses of TiO_2 eluates of whole cell lysate tryptic protein digest from WT, ΔPK4, and ΔMKK promastigotes and axenic amastigotes as well as WT lesion-derived amastigotes (WT_{LDA}) were conducted. For each targeted LC-MS/MS analysis of a selected range of protein kinases, 440 µg starting material (i.e., the concentration *prior* to TiO_2 enrichment) were used. For all the protein phosphatases the amount of starting material was 550 µg. The samples were prepared (TiO_2 enrichment, desalting, and LC-MS/MS preparation) as described in sections **2.2.7** and **2.2.9.1**. Due to the amounts in question, the TiO_2 enrichments were made separately for each of the sample replicates, but pooled prior to LC-MS/MS analysis.

The acquired data were imported into the final method file versions in Pinpoint for analysis.

2.2.11.3 MS^E

MS^E analyses were carried out by Dr. Richard Sprenger, Protein Research Group, SDU. Whole cell lysates of *Lmex* promastigotes and amastigotes, and *Lmajor* promastigotes were prepared as described in section **2.2.4**, and digested with trypsin as described in section **2.2.6**. The initial cell aliquots were different by cell numbers (*Lmex promastigotes* and *amastigote* aliquots contained 1×10^9 cells, whereas the *Lmajor* promastigote aliquot only contained 5×10^8). For comparison, protein concentration was

measured at the peptide level by application of the Qubit procedure (section **2.2.10.1**). The Qubit results were not used for aliquotation of the digested protein solutions as the results deviated far too much from the anticipated concentration (this was unfortunately the norm for Qubit concentration measurement on *Leishmania* derived samples). Aliquots of 10 µg digested protein were prepared based on anticipated sample concentration, and handed over to Richard, who performed initial sample titration (dilution with 0.1% TFA) to ensure that the expected sample concentrations were not too far off. 100 fmol BSA (Sigma®) and 50 fmol rabbit glycogen phosphorylase B (Sigma®) protein digests were added to serve as reference and internal standard. Triplicate sample analyses were carried out with nanoscale LC separation of the peptides by a NanoAcquity system (Waters Corp., Milford, MA, USA), equipped with a Symmetry C18 5 µm, 2 cm × 180 µm pre-column (Waters Corp.). Samples were loaded onto the trap in 0.1% FA (Solvent A) at a flow rate of 15 µl/min for 1 min. The peptides were separated over a gradient of 3-40% MeCN in 0.1% FA (Solvent B) for 90 min at a flow rate of 300 nl/min. Column temperature was maintained at 35 °C. The auxiliary pump of the NanoAcquity system provided 100 fmol of [Glu^1]fibrinopeptide B/µl at 300 nl/min to the reference sprayer of the NanoLockSpray source of the mass spectrometer. LC-MS/MS analyses were carried out on a Q-TOF Synapt HDMS instrument (Waters Corp., Manchester, UK), operating in positive nanoelectrospray ion mode. A typical resolution of at least 10,000 full-width half-maximum was used for the all measurements. The TOF analyser of the instrument was externally calibrated by [Glu^1]fibrinopeptide B. Post acquisition data lock mass correction was performed using the monoisotopic mass of the doubly charged precursor of [Glu^1]fibrinopeptide B as sampled by the reference sprayer with a frequency of 60 seconds. Accurate mass precursor and fragment ion LC-MS data were collected in data independent MS^E mode, alternating low energy and elevated energy mode of acquisition. Low and elevated energy MS spectra were both acquired from *m/z* 50 to 1990 for 0.98 sec each with a 0.02 sec interscan delay. In low energy MS mode, data were collected at constant collision energy of 4 eV. In elevated energy MS mode, collision energy was ramped from 15 to 40 eV during each 1.0 sec data collection cycle with one complete cycle of low and

elevated energy data acquired every 2 sec. The radio frequency applied to the quadrupole mass analyser was adjusted so that ions from *m/z* 300-1990 were efficiently transmitted; ensuring that any ions less than *m/z* 300 observed in the LC-MS data only arose from dissociations in the collision cell.

2.2.12 Bioinformatics

2.2.12.1 Raw data processing

Raw data acquired from the LTQ Orbitrap XL instruments were processed by either DTASuperCharge 1.31 (Peter Mortensen; http://msquant.sourceforge.net/#DTASCmain) or Proteome Discoverer version 1.043 (Thermo Fisher Scientific Inc.) to generate database-searchable file formats (mgf). Proteome Discoverer was also applied for analysis of iTRAQ data, combining raw data processing and database searching in a single workflow. The workflow involved two different legs, one reserved for MSA-derived spectra and one for HCD-derived spectra. The iTRAQ reporter ion ratios were calculated and normalised in Proteome Discoverer.

Data from the Q-TOF Premier were viewed in MassLynx version 4.0 (Waters Corp., Manchester, UK). Conversion of raw data to searchable pkl-files was done using MassLynx version 4.0 and/or ProteinLynx GlobalSERVER (PLGS), version 2.4 (Waters Corp., Manchester, UK).

MS^E data were processed and searched by PLGS, version 2.4. Protein identifications were obtained with the embedded ion accounting algorithm [181] of the software. Data were searched against two different *Leishmania* databases – a concatenated UniProt/TrembL database consisting of *L. major* and *L. mexicana* entries, and a predicted protein database for *L. mexicana* [182] – to which protein sequences of bovine albumin, rabbit glycogen phosphorylase B and trypsin were appended to serve as internal standard, providing the ability to address technical variation and accommodate concentration determination [138]. For protein identification and quantification, the

observed intensity measurements were normalised on the intensity measurement of the identified peptides from the digested internal standard. Database search tolerances were set automatically, typically 10 ppm for precursor and 25 ppm for product ions, cysteine carbamidomethylation was specified as fixed modification, and N-terminal acetylation, Gln → pyro-Glu (N-term Q), Glu → pyro-Glu (N-term E), and oxidation of methionine as variable modifications. Estimation of false positive identification rates was performed by searches in a randomised version of the concatenated database generated in PLGS, and typically searched at 5% protein false discovery rate per injection. Protein identifications were based on detection of 3 or more fragment ions per peptide, 2 or more peptides per protein, and identification of the protein in at least 2 of 3 replicate injections. Data was exported to Excel for further analysis. In order to normalise for differences in protein concentration between the different life stages, each triplicate run was normalised by dividing the concentration of each protein by the sum of all identified proteins after filtering.

2.2.12.2 Database searching

Mass spectrometric data were searched using a local Mascot server (http://mascot4.bmb.sdu.dk/mascot/). Database searching was either performed in line with processing of the raw data (Proteome Discoverer), or separately (mgf-files from DTASuperCharge, or pkl-files from MassLynx/ProteinLynx Global server). Data were searched against a *Leishmania mexicana* 6-frame translation library provided by Christiane Hertz-Fowler of The Sanger Institute [183] and modified by Heidi Rosenqvist (section **2.2.12.3**), a predicted *Leishmania mexicana* protein list [182] or alternatively against the *Leishmania major* protein database from GeneDB. Depending on the calibration and general standard performance of the Orbitrap on the day of data acquisition, data were searched with a peptide error tolerance of 5-10 ppm, and an MS/MS error tolerance of 0.6 Da. Q-TOF data were searched with a peptide error tolerance of 50 ppm and an MS/MS error tolerance of 0.1 Da. One missed cleavage by trypsin was allowed. Carbamidomethylation was chosen as fixed modification (samples

had been treated with iodoacetamide) and oxidation of methionine, Gln → pyro-Glu (N-term Q), Glu → pyro-Glu (N-term E), and Pyro-carbamidomethyl (N-term C) as variable modifications. For the enriched samples, phospho (S, T, and Y) were added as variable modifications.

2.2.12.3 Database modifications

Modification of the 6-frame database was performed to ease subsequent data analysis. By use of published lists of potential and/or identified protein kinases and phosphatases [30, 37] as well as comparative searches of data sets against both the 6-frame library and the *L. major* database, a number of the contigs in the 6-frame library were assigned with a protein name. This process was based on TBLASTN (protein versus translated DNA) searches on GeneDB [184], and occasionally, when a protein sequence appeared to be covered by more than one contig, theoretical protein digests in GPMAW [185] followed by Peptide Mass Fingerprint (PMF) Mascot searches against the original 6-frame library. The protein sequences used for theoretical digests would typically consist of the *L. mex.* part(s) that could be determined by the BLAST-search combined with the seemingly missing parts from the *L. major* or *L. infantum* sequence. In order not to risk losing sequence information that might be unique to *L. mex.* (i.e. parts of protein sequences that would not appear in any of the related *Leishmania* proteins), the contig sequences were not modified, even if they appeared to contain several hundred surplus amino acids in the N-terminus.

2.2.12.4 Validation

Search results for protein kinases and phosphatases were manually validated based on the spectra and their assigned ions. Peptides were only accepted if the fragmentation spectrum would display a succession of at least 3 assigned ions and a total assignment of at least 66% of the ions in an ion series, regardless the peptides score (only peptides with scores >15 were considered). For phosphopeptides, ions supporting the suggested

phosphorylation site(s) were required for validation of the site(s). If supporting ions were missing, but the peptide could otherwise be validated, the suggested phosphorylation site(s) would only be considered as plausible.

Evaluation of enrichment efficiency as well as general validation of phosphorylated peptides was performed with BioDesktop (in-house software developed by PhD-student Thomas Aarup Hansen, The Protein Research Group, Department of Biochemistry and Molecular Biology, University of Southern Denmark). Mascot search results were exported as XML-files, including “start” and “end” peptide match information as well as “MS/MS peak lists” and “Raw peptide match data”. For comparative analyses on the peptide level, sequence and position of phosphorylations were used as settings. General phosphorylation validation was performed using the assigned A-score. Phosphorylation sites with A-scores above 19 [186] were generally accepted if the phosphorylation site was supported by Mascot with a peptide score above 35. Manual validation was performed when 1) there were discrepancies between Mascot’s phosphorylation assignment(s) and that of the A-score; 2) multiphosphorylated peptides did not have signicant A-scores for all sites; and 3) phosphorylated peptides appeared to be phosphorylated on all possible residues. Manually evaluated A-scored phosphopeptides were accepted based on the criteria mentioned for protein kinases and phosphatases.

2.2.12.5 Comparative analyses

Comparison of Mascot search results for phosphorylated peptides between Mascot search results was performed in BioDesktop. Scaffold version 3_00_03 was used for general comparative analyses based on the Mascot search results. Scaffold was supplied with a FASTA-file of the 6-frame translation library and custom database parsing rules defined:

Accession Number parse rule

>(?:.*)?(Contig[^\s]*)

Description parse rule

>(.*)

Decoy parse rule

Random

The parsing rule codes were established through collaboration with Mark Turner, Proteome Software.

2.2.12.6 Statistical analyses

One-way ANOVA or 2-sample T-tests were used for statistical evaluation of the quantitative phosphoproteomics results. Calculations were carried out in Minitab 16 (Minitab Ltd., Conventry, UK).

CHAPTER 3: RESULTS

This chapter describes the establishment of a pipeline for proteomics and phosphoproteomics analyses in Leishmania mexicana. The phosphoproteomics section contains sub-sections on the potential phosphoproteins, -peptides and phosphorylation sites, as well as those that have been validated, from whole cell lysates and membrane protein extracts. Special focus has been put onto analyses of protein kinases and protein phosphatases, which will hence make up large parts of both the phosphoproteomics and subsequent quantitative phosphoproteomics sections.

3.1 Proteomics & phosphoproteomics pipeline

Building on existing knowledge and procedures, a pipeline spanning parasite harvest, differential protein extraction, phosphopeptide enrichment, LC-MS/MS- and data analyses was established step by step. To allow further evaluation of the identified proteins and peptides, different modules for quantitative analyses were incorporated. **Figure 3.1** below displays a simplified schematic of the established pipeline.

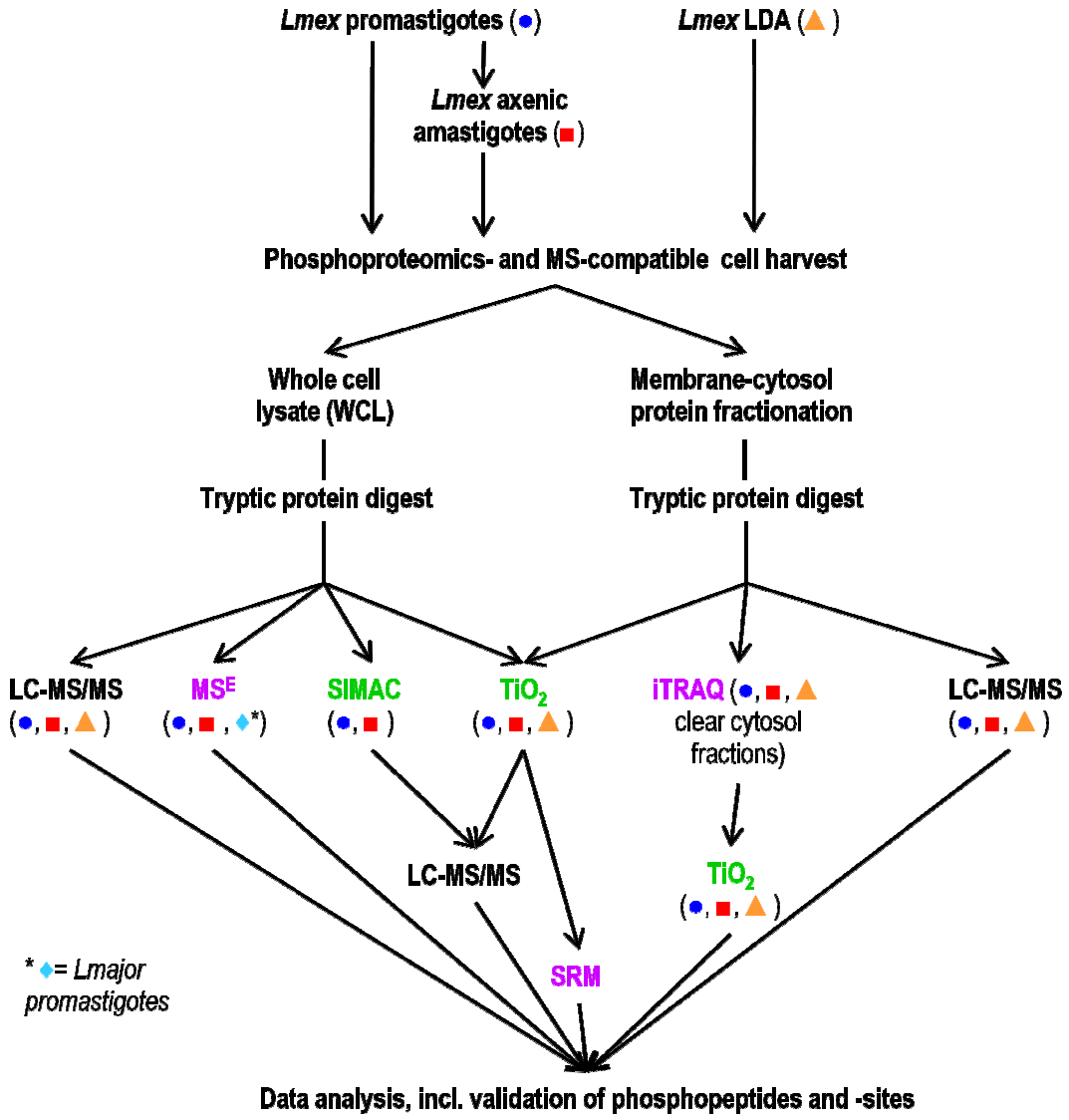


Figure 3.1: The established pipeline for bottom-up proteomics and phosphoproteomics analyses. The different sample types are marked by symbols (*Lmex* promastigotes = ●; *Lmex* amastigotes = ■; *Lmex* lesion-derived amastigotes (LDA) = ▲; and *Lmajor* promastigotes = ♦). The different techniques for phosphopeptides enrichment and quantitative analyses are colour-coded, **green** and **pink**, respectively. When a green and a pink module are directly connected, regardless their succession, they are part of peptide-based quantitative phosphoproteomics analyses.

Despite all of the procedures being well-established, some were slightly modified to accommodate the sample types and amounts to be used. Such considerations and alterations will be described in the sections dealing with the modules in question.

3.1.1 Determination of best suited protocol for harvest, lysis and protein extraction

Culturing and differentiation of *Leishmania mexicana* were well-established procedures in the Wiese-laboratory, however, the subsequent steps of harvest, lysis, and protein extraction were evaluated to determine how these would best be both compatible with phosphoproteomics MS/MS analysis as well as provide most protein yields. This evaluation was triggered by the very initial MS analyses revealing insufficient removal of BSA during parasite harvest [187]. Two different procedures were tested for harvest, lysis and protein extraction. One, here termed the "Odense" procedure, was a revised version (i.e. more washing steps) of the original Wiese protocol for harvest combined with a protocol for lysis and protein extraction derived from the Protein Research Group at University of Southern Denmark. The other procedure, here termed "Glasgow", was based on advice from Dr. Richard Burchmore's laboratory at University of Glasgow. The two procedures varied slightly by chemicals, solvents and concentrations as well as in the steps around lysis and protein extraction [187]. The initial trials for comparison were conducted on single sample runs, yielding far from conclusive results, where one comparison would favour the "Odense" procedure, and another the "Glasgow" procedure [187]. To reduce the coincidental effects that could have affected the initial comparisons, samples originating from the same culture were prepared by each of the procedures and run in triplicates for a comparative analysis. Both procedures easily passed the MS and phophoproteomics compatibility criteria. The crude comparison made with Scaffold™ showed a total of 589 proteins, 464 of which were phosphorylated. For the total number of proteins as well as the phosphorylated fraction, the overlap between the two procedures was significant (89 % and 92 %, respectively).

The remaining protein identifications were relatively evenly distributed between the procedures, with a slim majority for the "Odense" procedure at both levels (see **Figure 3.2**). Which proteins were identified by only one of the procedures appeared to be coincidental rather than based on certain characteristics or patterns. The observed differences were not significant enough to vindicate complementary use of the two procedures. Overall, the "Odense" procedure appeared to outperform the "Glasgow" procedure by the slightest possible margin, and therefore became the procedure of choice for all subsequent parasite handling.

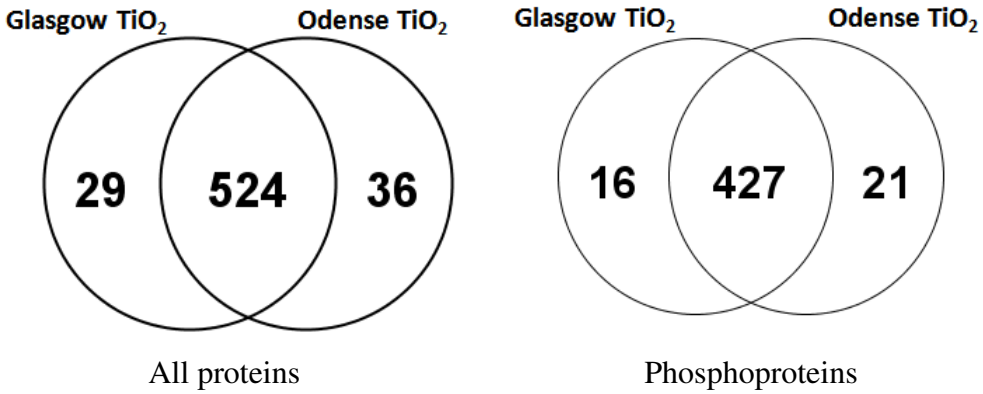


Figure 3.2: Comparison of the results obtained from the "Glasgow" and "Odense" procedure, respectively. Two Venn diagrams are displayed, one representing all identified proteins regardless of modifications (left), and one representing only those proteins identified by at least one phosphorylated peptide (right). In both cases, the overlap between the procedures is significant, but for the few proteins only identified by one of the procedures, the "Odense" procedure wins by a very tight margin.

3.1.2 Phosphopeptide enrichment strategies

Another important component in the phosphoproteomics pipeline was maximization of the detection of phosphorylated peptides. The ionisation ability as well as amount of unmodified peptides far exceeds those of modified peptides. Hence, efficient analyses of modified peptides, such as phosphopeptides, require some kind of prior enrichment steps to concentrate the peptides of interest in the sample. While this enrichment can be performed at both protein and peptide level (see section 1.3.1), or even combined for enrichment at both these levels, the current study only employed peptide-level enrichment. The phosphoproteomics analyses presented in this thesis were conducted after TiO_2 chromatography or sequential immobilised metal affinity chromatography (SIMAC). The outcome as well as applied modifications for these protocols will be described below.

3.1.2.1 TiO_2 chromatography

TiO_2 chromatography was performed off-line, prior to LC-MS/MS analyses. The initial analyses of whole cell lysates led to selection of 330-333 µg digested protein as a suitable amount of starting material. While less starting material was applicable, it was at the cost of significantly reduced identification of low abundance phosphoproteins, including many of the protein kinases and phosphatases of interest (section **3.5.3**). However, this rather large amount of digested proteins entailed relatively large sample volumes as well. For TiO_2 chromatography, the sample needs to be diluted 5 times in loading solution (see section **2.2.7**). The prevalent procedure for TiO_2 chromatography involved packing the TiO_2 material in GeLoader or p10 pipette tips. However, these tips have a maximum volume of 30 µl, which was far too little to accommodate a dilution of 330 µg protein digest. Hence, the aliquots of tryptic digests to be enriched by TiO_2 were volume-reduced to 20-25 µl by vacuum centrifugation. TiO_2 material was packed in p200 tips to allow a larger amount of TiO_2 material as well as the possibility to load 120-150 µl diluted sample. Similarly, the recommended volumes of the subsequent washing steps were increased. This proved to be just as successful as enrichment of small sample

amounts in GeLoader or p10 tips. The TiO_2 enrichment procedure proved repeatedly successful, and so even when a batch mode for TiO_2 enrichment became the norm in the PR Group, no changes in the procedure were made. Compared to the batch mode of TiO_2 enrichment, the column-based procedure does not require prior knowledge of the sample protein concentrations, or subsequent weighing out of the TiO_2 material, making it a simple and almost equally fast alternative (depending on the number of samples to be enriched). To evaluate if the amount of TiO_2 material used for the amount of starting material was sufficient, a dual-enrichment trial was carried out (**Figure 3.3**). This involved an additional TiO_2 chromatography step of the flow-through from the initial TiO_2 chromatography. The eluates of these two enrichment steps were analysed separately and the results compared. The difference between the total phosphopeptide output of the singly versus the dual TiO_2 chromatography strategy was shown to be minimal, and the second TiO_2 enrichment step did not prove worthwhile [187]. With its ease and relative reproducibility (see section **3.6.1.1**), the customised traditional procedure for TiO_2 chromatography (**Figure 3.3 A**) became the method of choice for the different phosphoproteomics experiments carried out throughout this study.

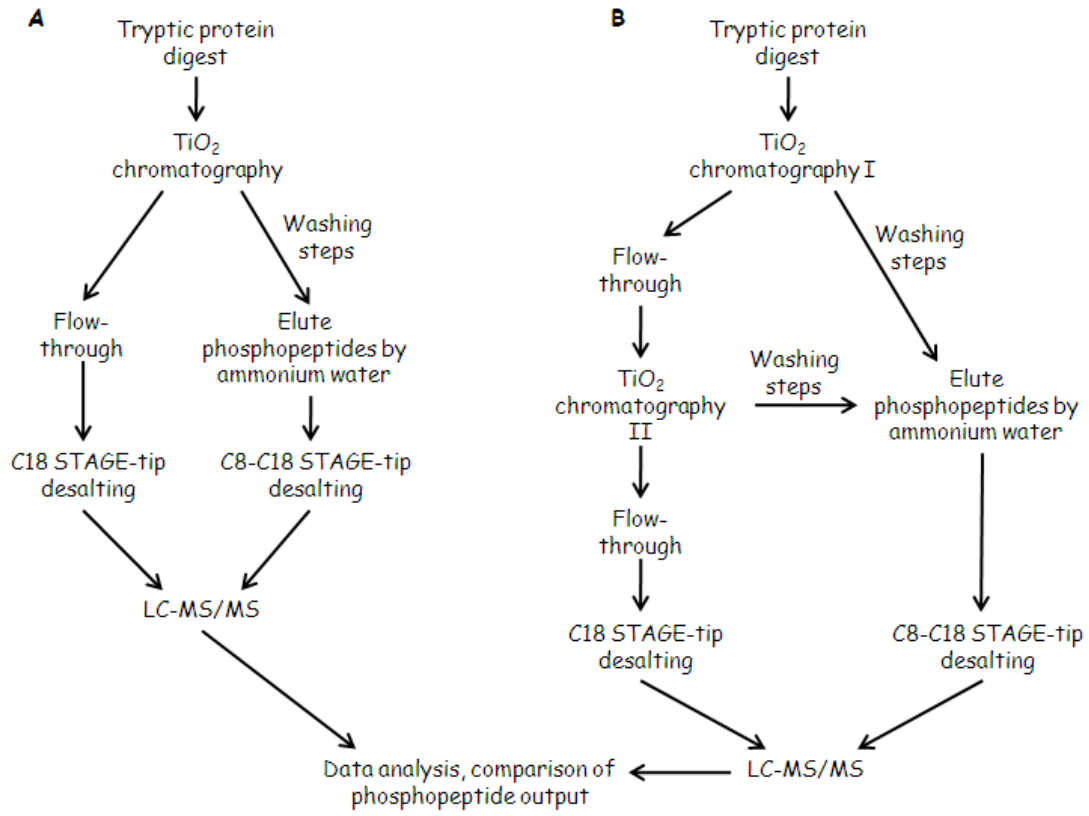


Figure 3.3: Evaluation of TiO_2 chromatography for phosphopeptide enrichment. The traditional TiO_2 chromatography approach (**A**) was carried out with larger TiO_2 columns (prepared in p200 tips). The majority of phosphorylated peptides would be retained on the TiO_2 column for elution with ammonium water. The flow-through of the TiO_2 column would predominantly contain un-phosphorylated peptides. The flow-through as well as TiO_2 eluates were desalted on Stop And Go Extraction (STAGE) tips and analysed by LC-MS/MS. In the dual-TiO_2 chromatography approach (**B**), an additional step of TiO_2 chromatography was applied to the flow-through of the first TiO_2 chromatography step to evaluate if the amount of TiO_2 material used in the columns was sufficient for the columns not to be overloaded. All subsequent steps were similar to those of the traditional TiO_2 chromatography approach. Final comparison of the enriched phosphopeptide samples revealed insignificant differences, thus indicating that the traditional approach was well-balanced. A more extensive description of the TiO_2 chromatography procedure can be found in section **2.2.7**.

3.1.2.2 Sequential immobilised metal affinity chromatography (SIMAC)

TiO_2 chromatography has a reputation of discriminating in favour of singly phosphorylated peptides, and thus other phosphopeptides enrichment strategies are often considered to increase the yield in phosphoproteomics analyses. This was also the case in the current study, even though a number of multiply phosphorylated peptides, e.g. from the MAP kinases, seemed readily identifiable by TiO_2 chromatography [187, 188]. To assess the enrichment of multiply phosphorylated peptides, a simplified version of sequential elution from IMAC (SIMAC) [12] was tested. In SIMAC, the strengths of IMAC (high affinity for multiply phosphorylated peptides) and TiO_2 (better for singly phosphorylated peptides) are combined. In large-scale phosphoproteomics analyses, the singly phosphorylated peptide fractions of SIMAC are often subjected to hydrophilic interaction liquid chromatography (HILIC) prior to TiO_2 enrichment. The objective of this test, however, was to evaluate identifications of multiply phosphorylated peptides by IMAC as opposed to TiO_2, to decide whether or not SIMAC should be permanently involved in the study. Hence, the HILIC step was omitted in SIMAC as well as up-front the traditional TiO_2 chromatography procedure. SIMAC was carried out 3 times with wild type promastigotes, and 3 times with wild type axenic amastigotes, each time using 1,000 µg starting material. The experiments were not carried out at once, rather one at a time, assessing the outcome and making slight changes for the next experiment. Thus, the reproducibility of the SIMAC procedure will not be evaluated, suffice it is to say that despite several tests, the procedure did not appear to be fully optimised as the fraction of multiply phosphorylated peptides generally contained far too many un- or singly phosphorylated peptides. The timing of the first elution step from IMAC is critical, and it seems to be a very delicate balance not to lose the multiphosphorylated peptides while also not retaining too many singly or un-phosphorylated peptides.

3.1.2.3 TiO_2 vs. SIMAC

To fully evaluate the applicability of the two enrichment procedures, results of the SIMAC enrichments were compared to the results obtained by TiO_2 enrichment. As the starting material for each single SIMAC trial was 1,000 µg, thus equating a total of 3,000 µg from triplicate experiments on promastigote and amastigote proteins, respectively, several analyses of TiO_2 eluates were merged to allow comparison of identifications from similar amounts of starting material. TiO_2 chromatography was generally carried out on 333 µg of starting material, thus requiring the results of 9 analyses to be pooled. For amastigotes this meant triplicate TiO_2 analyses from 3 different samples, whereas for promastigotes all TiO_2 analyses could be retrieved from the same sample, though originally analysed as a sixplicate and a triplicate, respectively. For the comparative analysis, the mgf-files generated from the different SIMAC and TiO_2 raw data files were searched against the newest version of the Lmex6-frame database. The resulting Mascot result files were uploaded to Scaffold 3_00_03 in two separate analyses, one representing the amastigote SIMAC and TiO_2 results and the other representing the promastigote SIMAC and TiO_2 results. To get a comprehensive view of identifications made by the two different enrichment procedures, the Mascot search files were treated as were they derived from MuDPIT experiments [189], combining all SIMAC files and all TiO_2 files, respectively. The overall results of the analyses are shown in **Figure 3.4** below.

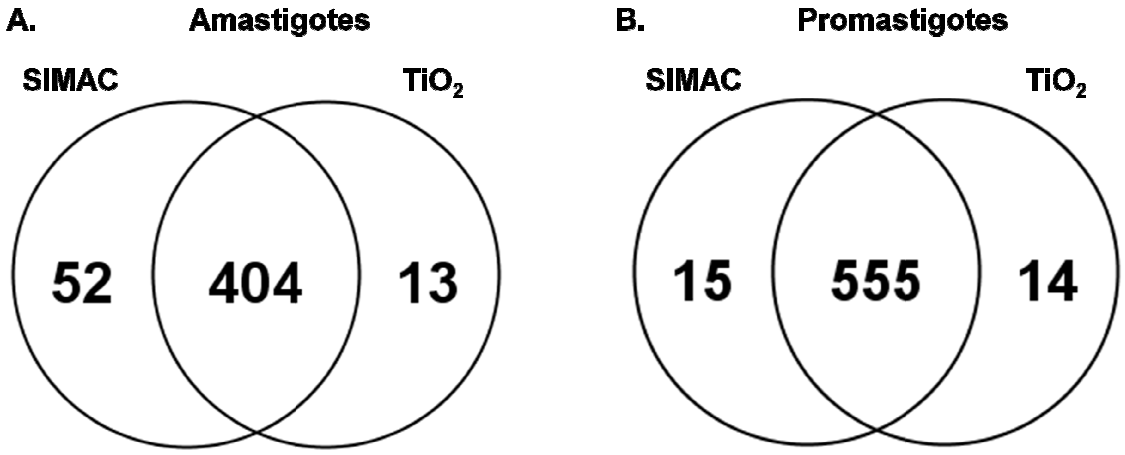


Figure 3.4: Scaffold Venn diagrams of the phosphoprotein identifications made by SIMAC and TiO_2 enrichments. For both life stages, the majority of phosphoprotein identifications are shared between SIMAC and TiO_2. The identifications unique to one of the enrichment procedures are very similar for promastigotes (**B**). SIMAC has significantly more unique identifications in amastigotes (**A**). This may be explained by the fact that the amastigote SIMAC experiments were the last ones to be carried out, thus potentially reflecting a slight improvement in the application of the procedure.

Figure 3.4 suggests that SIMAC may indeed lead to more protein identifications, so the question was then if these additional identifications would be derived from identifications of multiphosphorylated peptides. Venn diagrams of unique peptide identifications can also be displayed in Scaffold. Unfortunately, these peptides are listed without any indications of modifications or their protein origin, making for a very tedious manual search to locate these hundreds, or even thousands, of peptides and determine their phosphorylation status. At the protein level (**Figure 3.4**), the numbers are much lower and the information sufficient to allow location of the proteins to assess the identified peptides. Thus, to get a preliminary idea of the distribution of multi- vs. mono-phosphorylated peptides identified by the two procedures, the proteins identified solely by each were scrutinised. **Table 3.1** summarises the distribution of proteins identified with only singly phosphorylated peptides or with at least one multiply phosphorylated peptide.

	SIMAC			TiO_2		
	Multiphos.	**Monophos.**	**Total**	**Multiphos.**	**Monophos**	**Total**
Promastigotes	7	8	**15**	1	13	**14**
Amastigotes	6	46	**52**	10	3	**13**

Table 3.1: Distribution of proteins identified with peptides carrying one or more phosphorylation sites for SIMAC and TiO_2, respectively. Multiphos.: Proteins identified by at least one peptide carrying 2 or more phosphorylations. Monophos.: Proteins identified with singly phosphorylated peptides.

While an evaluation based only on protein identifications unique to each procedure is by no means ideal for assessment of the potential of SIMAC to add more identifications of multiply phosphorylated peptides, it does indicate that the impression of required optimisation of the procedure is valid. This will be discussed further in section **4.1.2**. If anything, the numbers of multi- versus monophosphorylated peptides listed in **Table 3.1** suggest that enrichment of these by SIMAC and TiO_2, respectively, is relatively coincidental. For the promastigote samples, almost equally many multi- and monophosphorylated peptides were identified by SIMAC, whereas the distribution for these peptides by TiO_2 perfectly matched the anticipated dominance of singly phosphorylated peptides. For the amastigote samples, however, the vast majority of the phosphopeptides unique to SIMAC were singly phosphorylated, whereas those unique to TiO_2 were multiphosphorylated.

3.2 Mass spectrometry – which techniques to fit into pipeline

The Protein Research (PR) Group at University of Southern Denmark has a very well-equipped mass spectrometry laboratory at disposal. Several different instruments were considered for use in the LC-MS/MS couplings of the proteomics pipeline. Evaluation was based not only on performance, but also on availability and data analysis considerations.

The initial LC-MS/MS analyses, evaluating the first sample handling and preparation steps, were performed on a Q-TOF type of instrument (Waters Q-TOF Premier). While the resolution of the acquired Q-TOF spectra typically exceeded that of Orbitrap spectra, the overall results depended on whether the analysed peptides were phosphorylated or not. The Q-TOF instrument does not carry the ability to accommodate the typically very significant neutral loss of pS- and pT-containing phosphopeptides, impacting the identification rate. On the other hand, the Orbitraps available would allow for phosphopeptide-customised MS/MS analyses by MSA, HCD, or a combination of the two. The end results of Orbitrap analyses would, expectedly, far exceed those of Q-TOF analyses, not only by numbers of identifications (see **Figure 3.5**), but also on the sequence coverage of the phosphopeptides, thus facilitating phosphorylation site validation. The differences in numbers of identified peptides also show by overall protein sequence coverage. In an un-enriched sample, one of the most dominant proteins identified by both Q-TOF and Orbitrap, heat-shock protein 83-1, achieved 21% sequence coverage from the Q-TOF analyses, but 42% sequence coverage in the Orbitrap analysis. These differences were from analyses run with standard instrumental settings, but further optimisation of the instrument methods on the Orbitrap were conducted for subsequent analyses. The standand instrumental settings of the Q-TOF and Orbitrap, respectively, are comparable in terms of precursor selection, thus the increased number of identifications by the Orbitrap was not just a matter of more peptides being selected for fragmentation. The majority of the discovery experiments as well as the iTRAQ quantitative analyses were conducted on Orbitrap instruments. This, however, reduced the flexibility of when to carry out the analyses as any of the Orbitraps

typically had to be booked two months in advance, and booking rules prevented more than one booking ahead at any one time.

Figure 3.5 – legend on p. 101.

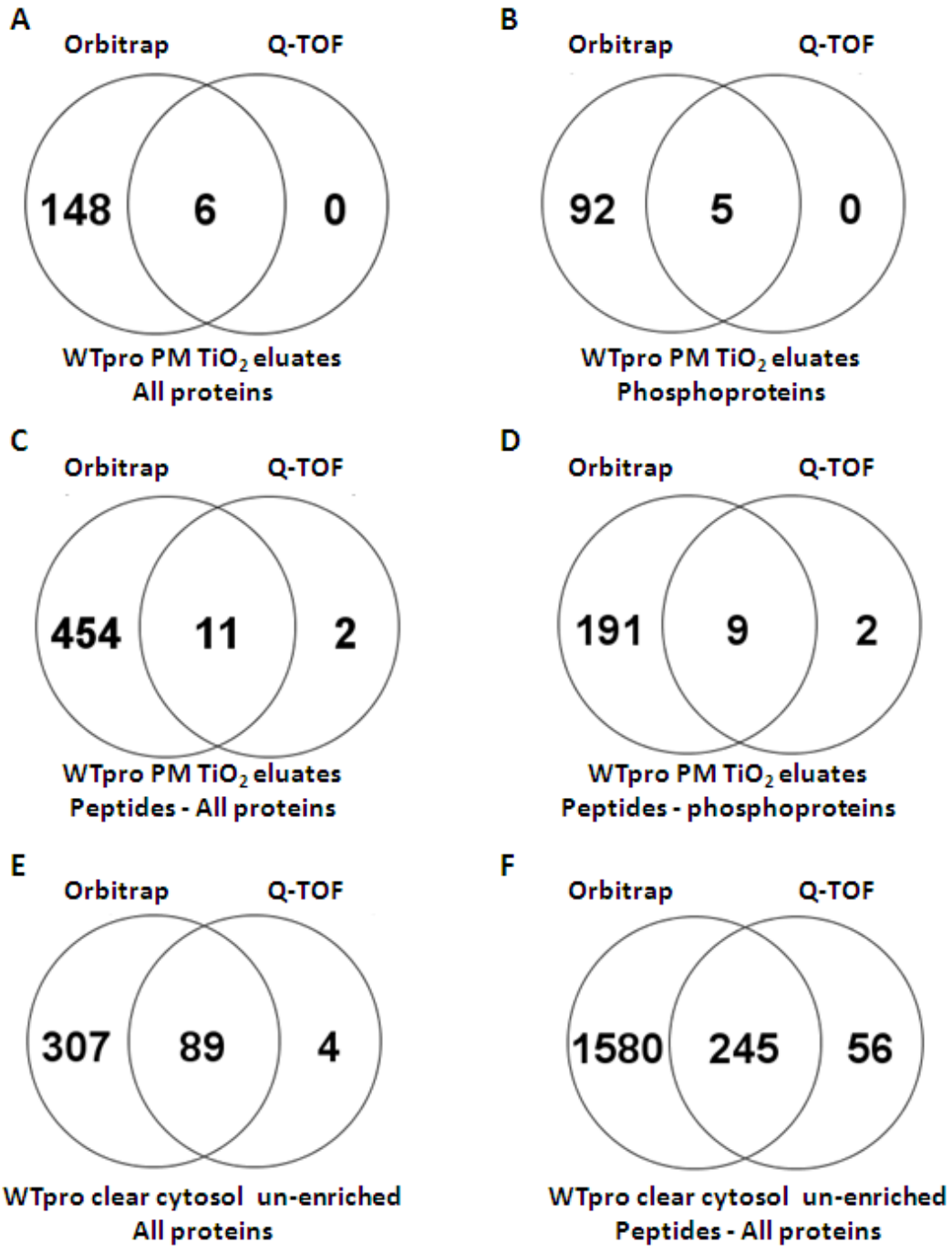


Figure 3.5: Scaffold comparison of Q-TOF and Orbitrap performance for proteomics and phosphoproteomics analyses, respectively. Comparisons were made with standard instrument-specific settings for both phosphopeptide enriched samples (**A-D**) and un-enriched samples (**E-F**). Scaffold settings were 99.0% protein confidence, minimum 1 peptide/identification, and 95% peptide confidence. Diagrams **A** and **B** display the distribution of proteins identified under these conditions by the two different instruments in analysis of phosphopeptides enriched from wild type promastigote membrane fractions. Diagrams **C** and **D** displays the distribution of unique (to the samples, red.) peptides in the aforementioned sample type. Diagrams **E** and **F** display the distribution of identified proteins (**E**) and sample-specific unique peptides (**F**) from an un-enriched sample of wild type promastigote clear cytosol fraction analysed on the two instruments. At both protein and peptide levels, regardless if phosphorylated or not, the Orbitrap appeared far superior to the Q-TOF instrument in terms of reliable identifications.

Apart from iTRAQ, the quantitative analyses undertaken in this study had their own requirements with regard to mass spectrometer performance. Label-free quantitative analyses could in principle be carried out on most of the available instruments; however the targeted analyses (SRM) were obvious tasks for a triple quadrupole instrument, in this case a TSQ Vantage (Thermo). MS^E analyses are currently a matter for the Q-TOF type of instruments, and in this case analyses were carried out on a Waters Synapt HDMS.

MALDI TOF MS and MS/MS analyses were also applied intermittently, typically for initial sample quality checking and assessment of iTRAQ labelling efficiency.

3.3 Database improvements

With pipeline components of harvest, lysis, protein extraction and phosphopeptide enrichment established, proteomics and phosphoproteomics analyses were enabled, though initially compromised by the database status.

At the beginning of this project, no true species-specific protein databases were available for *Leishmania mexicana*. Instead, the initial analyses were conducted with a concatenated database of all *Leishmania* protein sequences available in the SwissProt/UniProt repository, or a *Leishmania major* database. Neither database choice was optimal. The redundancy in the concatenated database was extensive, as it primarily contained the most abundant *Leishmania* proteins (e.g. the tubulins, heat shock proteins, and ribosomal proteins) from numerous different *Leishmania* species displaying minimal sequence variation. Of the more than 25,000 protein sequences in the concatenated database, only 139 actually derived from *Leishmania mexicana*. The *Leishmania major* database on the other hand, while being more correctly diverse, suffered from those minor, yet in terms of correct peptide assignment, significant sequence variations

existing between *L. major* and *L. mexicana*. Hence, the possibility of either missing protein/phosphopeptide identifications or identifying some that would not be present in *L. mexicana* due to sequence variations was an impending risk. This is futher elaborated upon in section **4.2**.

In July 2008, however, a 6-frame translation library of *Leishmania mexicana* assembled contigs was released by Sanger [183]. This was installed on the in-house Mascot server at SDU, Odense, to use for mass spectrometric data analysis. The 6-frame translation library was annotated by the corresponding DNA contig names, making it challenging to get an overview of the search results in terms of interesting proteins identified [188]. Additionally, searches against this database yielded 3-4 times the number of protein identifications as could be achieved with any of the previously used databases. Hence, extensive homology analyses were initiated to "annotate" the contigs. *Leishmania mexicana* protein kinases and phosphatases were the first to be matched to contigs by protein-DNA BLAST searches. A number of annotations also originated from comparative analyses where the same raw MS-data were searched against the databases initially used and the 6-frame translation library. Subsequently, the MS search results would be the base of contig annotation by conducting protein-protein BLAST searches of the contig sequences against the predicted protein sequences of *L. major*, *L. infantum*, and *L. braziliensis*. In some cases, a contig would only match part of a proposed protein sequence. In these cases, the contig sequence would be merged with the remaining protein sequence of the top BLAST hit, imported into GPMAW [185] for a theoretical tryptic digest, and the generated mass list would then be used for Peptide Mass Fingerprint (PMF)-analysis against the 6-frame translation library in Mascot. This way, more than 2,000 contigs have been annotated, enabling more than 90% of the hits in a typical search result to now carry a potential protein name.

The size and complexity of the *Leishmania mexicana* 6-frame translation library database affects not only the search time, but also the ability to predict false discovery

rates, as the database is already approximately 50% non-sense. Because it is virtually impossible to assess which of the contig sequences are non-sense and which are true, even if just covering minor protein sequence parts, the release of the *Leishmania mexicana* predicted protein database by Sanger in the summer of 2009 was anticipated with hope. The web-based version of this database [182] is continuously updated to also include protein functions, and has thus improved dramatically over the past year. On the other hand, the downloadable version has not been modified since its initial release last year (Lmex.pep.gz on ftp://ftp.sanger.ac.uk/pub/pathogens/L_mexicana/). This implies that some sequences are still not complete, as indicated by *'s and x's, and no protein functions are included. Additionally, it was found that in some cases the predicted protein sequences would be short of N-terminal parts not present in the Scaffold organism (*L. major*), but in *L. mexicana* as well as *L. braziliensis* and/or *L. infantum* (see [188] as well as section **3.5.4** for examples of this). Thus, the *L. mexicana* predicted protein database has predominantly been employed in conjunction with the 6-frame translation library, and annotation of the latter not ceased.

A separate annotation effort emanated from the MS^E experiment (section **3.7.3**), and involved bioinformatics analyses by Veit Schämmle, post doc in the PR Group. Veit managed to annotate 2994 of the sequences (**Appendix B**, http://dl.dropbox.com/u/3011619/Appendix%20B.xlsx) in the *Lmex* predicted protein list, not including any hypothetical proteins. These annotations were used to get an overview of the potential biological implications of the MS^E results (see section **3.6.3**).

3.4 “All proteins” identified

The arrival of the 6-frame translation library, and later a *Leishmania mexicana* predicted protein database [182], enabled more extensive analyses of the acquired MS-data. Focusing on protein kinases and phosphatases with little (e.g. membrane protein extraction) or no pre-fractionation of the total cell lysates, meant that the bulk of the search results would be anything but kinases and phosphatases. To not waste all this precious information, creation of a library of all the proteins identified in the different experiments was initiated. While the current version of this library (see **Appendix C**, http://dl.dropbox.com/u/3011619/Appendix%20C.xlsx) only contains entries from approximately 10% of all the acquired data, it still lists more than 2,000 proteins. The proteins in the library have been identified from both phosphopeptide enriched samples, and samples of raw cell lysate, membrane or cytosolic protein fractions, thus containing a mixture of potentially phosphorylated and un-phosphorylated protein. In addition, some proteins are identified by more than one contig, where each covers part of the entire protein sequence. The current library version has not been filtered for these as they add another level of information to the overall results with crude locations of the potential phosphorylation sites. A manual review of the library, subtracting all obvious redundancies and adding membrane transporters (section **3.5.4**), protein kinases and phosphatases (section **3.5.3**) identified from datasets that have yet to be included in the library, sums up to 2201 proteins. A crude classification of these can be seen in **Figure 3.6**.

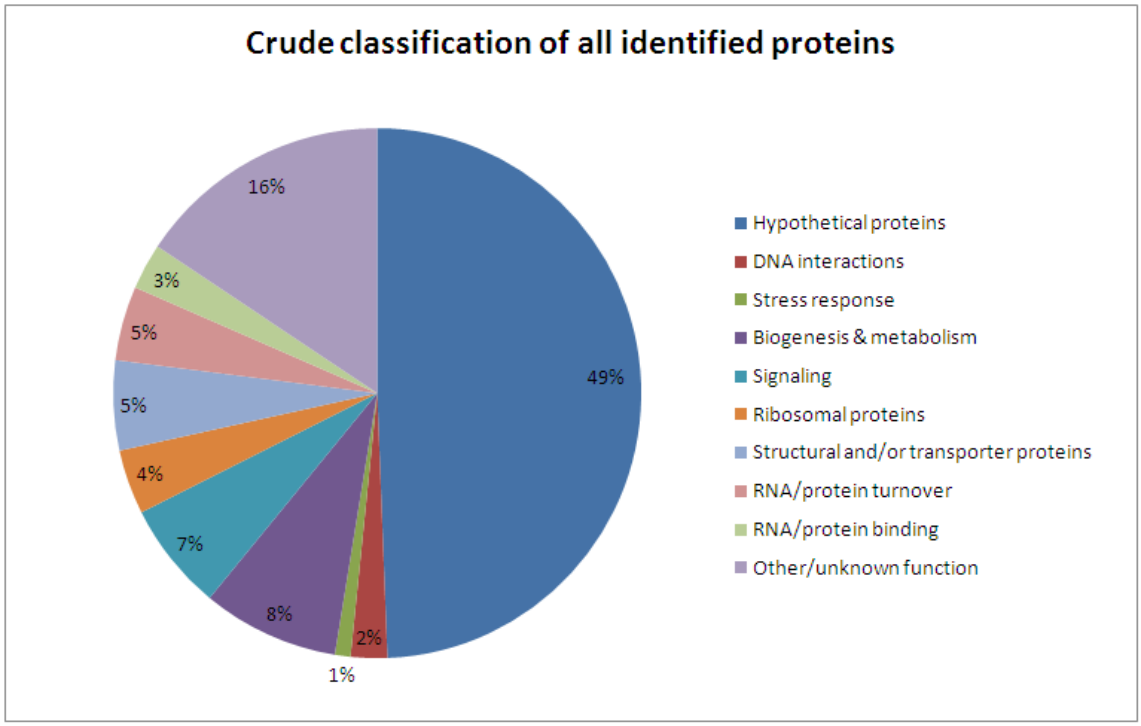


Figure 3.6: Crude classification of all identified proteins. Assembly of redundancy-filtered protein identifications from the all identified proteins library along with membrane transporters, protein kinases and protein phosphatases that have yet to be included in the library. The proteins have been grouped according to general functions/characteristics, only partly based on available GO-annotations.

Not surprisingly, given the fact that close on 68% of the genes found *Leishmania* have no known function (i.e., their predicted proteins are dubbed hypothetical) [39], the protein classification in **Figure 3.6** reveals that almost half of the identified proteins are hypothetical, i.e. they do not resemble known proteins to any great extent and have not been experimentally evaluated to suggest their function. Some of these proteins are likely involved in signalling pathways in one way or another, and could thus contribute to the 7% fraction currently covering proteins involved in signalling. The vast majority of those proteins making up the signalling group are further analysed and evaluated in sections **3.6** and **4.3**. Another group of proteins worth noticing in the chart is the ribosomal proteins. The ribosomal proteins, involved in the cellular process of mRNA translation and thus protein synthesis, were among those proteins found to be most profoundly regulated between the life stages in the MS^E analysis (section **3.6.3**). The proteins identified in the MS^E analyses were not included in the "all identified proteins" list or chart. The "all identified proteins" library is being built upon information from regular LC-MS/MS analyses, and while this in itself entails some limitations as to the identification of peptides for subsequent protein identification (dynamic range of the instrument, how many precursor ions are selected in each MS survey scan, etc.), the findings can still indicate potential differences between the samples in question. In relation to life stage-specific differences, the "all identified proteins" library suggests that approximately half of the proteins are shared between both life stages, while the other half is equally distributed between promastigotes and amastigotes, respectively (**Figure 3.7**). In relation to this, it should be mentioned that of those 22 datasets currently making up the library, 13 are from amastigotes and 9 from promastigotes.

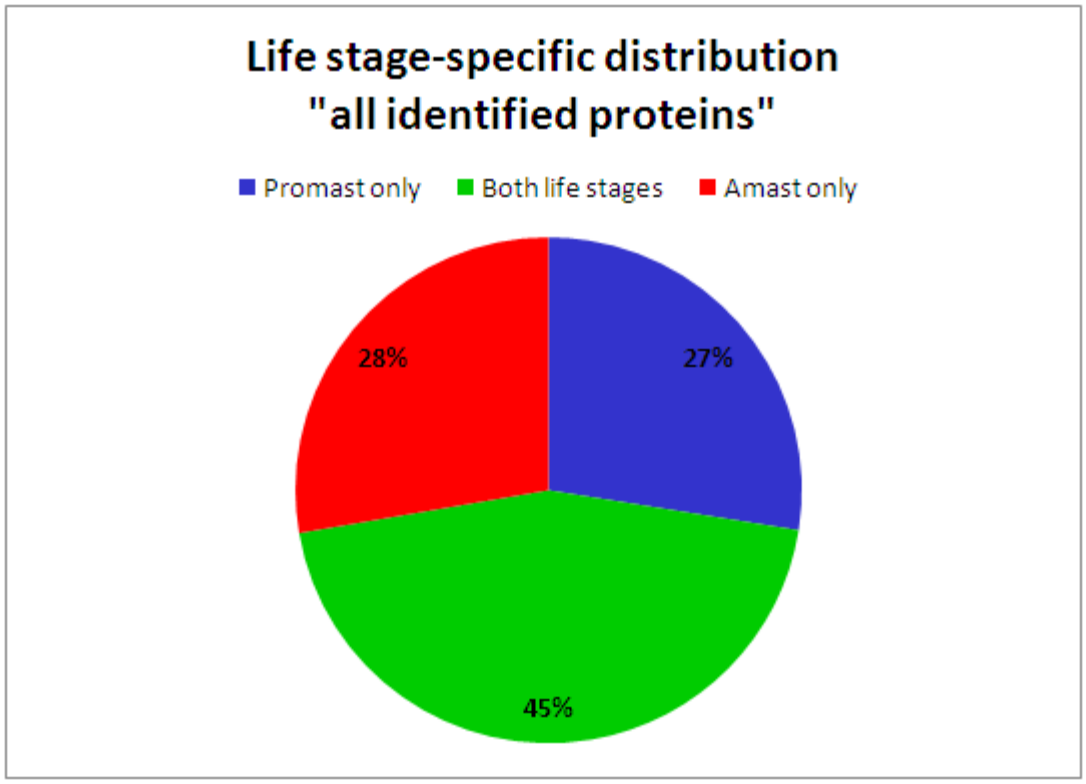


Figure 3.7: Life stage-specific distribution of the proteins in the "all identified proteins"-list. This assessment is based solely on the protein level, as at the peptide level more differences between the life stages may occur, especially by modifications. Approximately half of the identified proteins were found in both life stages, and the rest fairly equally distributed between promastigotes and amastigotes. In this diagram, no differentiation between axenic and lesion-derived amastigotes has been made.

3.5 *Leishmania mexicana* phosphoproteomics

The proteomics analyses were concentrated on the analyses of phosphorylated proteins. Hence, the majority of the MS-analyses were conducted on samples that had been enriched for phosphopeptides. Different procedures were applied for phosphopeptide enrichment (section **3.2**). The database issues described in section **3.3** and further discussed in **4.1** and **4.2**, however, complicated a large-scale phosphoproteomics analysis, as the mere phosphopeptide information would be rather lost in context, if the proteins to which the peptides belonged were unknown, or if phosphopeptides belonged to possible non-sense proteins (e.g., contigs that are too short to confidently be matched to a protein).

3.5.1 Potential phosphoproteins, phosphopeptides and phosphorylation sites

The "all identified proteins" list currently contains 1381 potential phosphoproteins. Among those are also some of the validated protein kinases and phosphatases (section **3.5.3**). Looking at the potentially phosphorylated peptides, 5,127 different peptide sequences with one or more different phosphorylation sites were found in the bulk of the whole cell lysate and membrane-cytosolic protein extract phosphoproteomics analyses. If counting phosphopeptides as different if they, despite having the exact same amino acid sequence, have the phosphorylation site(s) differently positioned, an estimated 10-20,000 phosphopeptides need to have their validity scrutinised, not only in terms of correct phosphorylation site assignment, but also whether they belong to a "true" protein sequence. With the improved databases this is now possible for most of these peptides, yet a very time consuming task that remains to be completed.

3.5.2 Validated phosphopeptides and phosphorylation sites

The identification of peptide sequences and potential modifications provided by Mascot and other proteomics search engines are based on the best possible hit (probability-based), and rated by how well it complies with the search algorithm [190, 191]. Yet, these identifications are not always true, and especially when PTMs are involved, the results may be dubious, not only due to the way the proteomics search engine works, but just as much due to the quality of the spectra. Therefore, proper identification of phosphopeptides and phosphorylation sites requires inclusion of one or more validation steps. Validation may be performed by manual inspection of spectra, use of programs or scripts for automated evaluation, or a combination of these.

The current study applied manual validation for the identified protein kinases and phosphatases (section **3.5.3**), as well as a combination of automated and manual validation for a number of full datasets. Eleven datasets from whole cell lysate TiO_2 analyses and two datasets from TiO_2 analyses of cytosolic protein fractions have been evaluated by use of A-score [186], accepting scores >19 unconditionally, while adding additional requirements and occasional manual validation for phosphopeptides with A-scores <19, but with high Mascot and/or Mascot delta scores [192], to be accepted.

The 11 whole cell lysate TiO_2 datasets resulted in a total of 4493 phosphopeptides with 4902 phosphorylation sites validated by A-scores. These numbers, however, are highly redundant, as the same peptides may have been identified and accepted by A-score a number of times, both in each individual dataset, but also between the different datasets. Also, the fact that one missed cleavage was allowed in the database searches means that the same phosphorylation site may be counted a number of times. If missed cleavages are accepted in the definition of individual peptides – even if they carry phosphorylation site(s) identical to their peers without any missed cleavages – the eleven datasets count

1694 different phosphopeptides with 2070 phosphorylation sites. Removing the redundancy for the phosphorylation sites is more difficult, as a lot of cross-checking is necessary to ensure that the phosphorylation site in question has indeed been counted more than once.

A-scoring of the 2 cytosolic protein fractions gave 294 phosphopeptides with 356 redundant phosphorylation sites. Filtering these results, left 243 different phosphopeptides with 281 phosphorylation sites (redundancy significantly reduced).

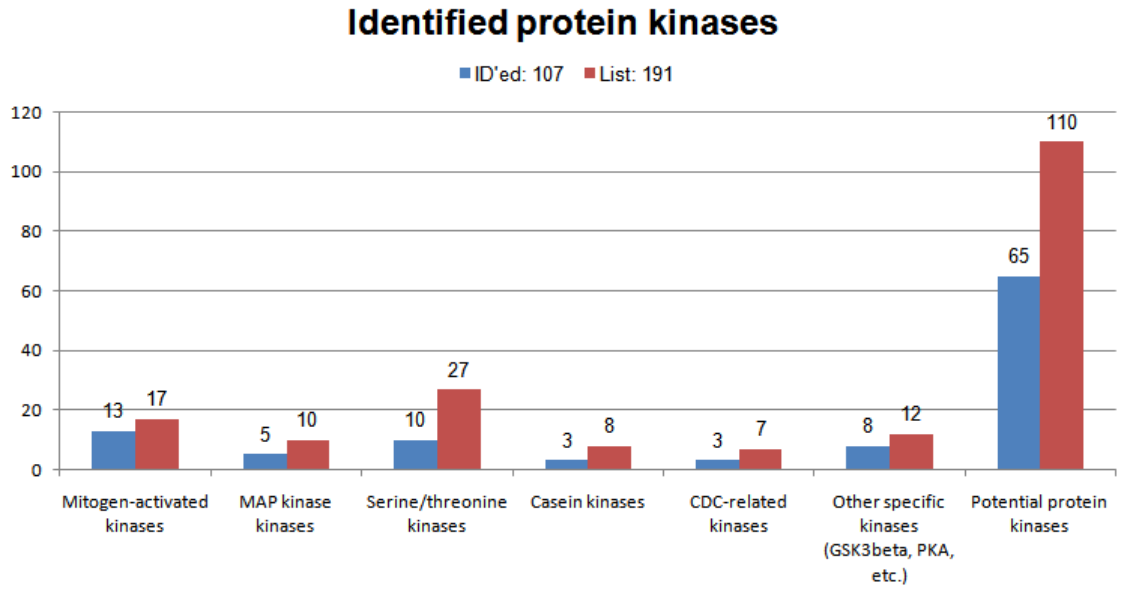
The collection of redundancy-reduced phosphopeptides and phosphorylation sites validated by A-score can be seen in **Appendix D** (http://dl.dropbox.com/u/3011619/Appendix%20D.xlsx).

3.5.3 Protein kinases and phosphatases

With the original objective of the project being that of signalling networks in *Leishmania mexicana* assessed by phosphoproteomics, the main focus of this study became phosphoproteomics analyses of protein kinases and protein phosphatases. These proteins are among the top conductors of signal transduction orchestration, yet this fact typically also entails them being far less abundant than their substrates as the effects of their actions are amplified by subsequent substrates. Analyses of *Leishmania major* led to publication of the proposed *Leishmania* kinome and phosphatome in 2005 [30] and 2007 [37], respectively. The lists provided by these two publications were used in a homology set-up to deduce the sequences of protein kinases and protein phosphatases in *Leishmania mexicana*. The DNA sequences of the *L. major* proteins were employed for BLAST searches against the preliminary *L. mexicana* genome, giving rise to a list of 191 protein kinases and 89 protein phosphatases in the latter species. The information about which contigs made up these protein sequences were then used to survey the MS search

results for their presence. Hundreds of LC-MS/MS analyses of whole cell lysates as well as membrane protein and cytosolic protein extracts from promastigotes, axenic amastigotes and lesion-derived amastigotes, led to the identification of 107 of the protein kinases and 36 of the protein phosphatases in wild type and different kinase deletion mutant parasites (**Figure 3.8**).

A.



B.

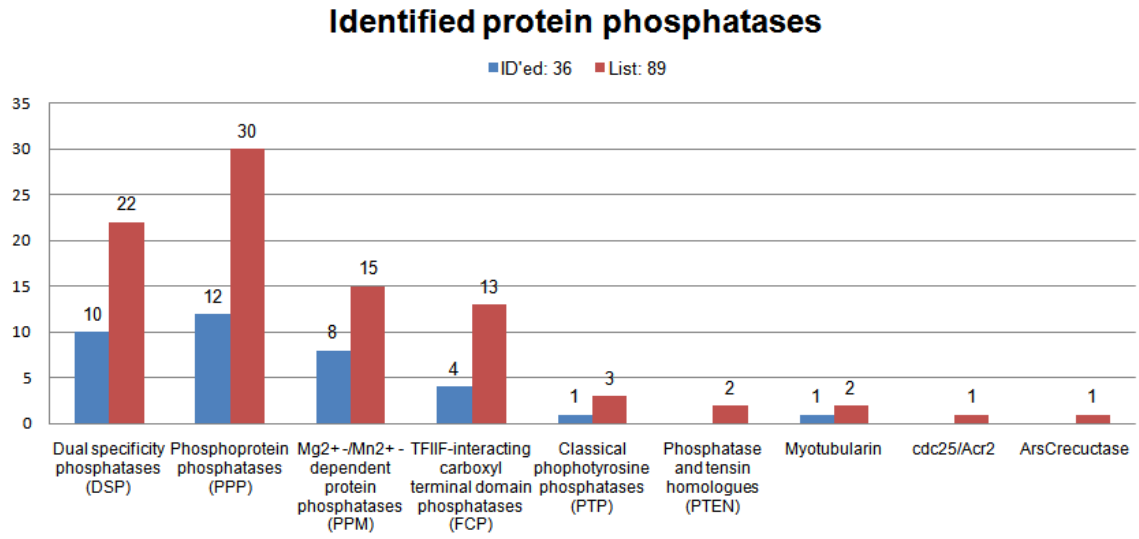


Figure 3.8: Identified protein kinases (A) and protein phosphatases (B) in *Leishmania mexicana*. The blue bars represent the proteins identified whereas the red represent the proteins present in the lists of proposed protein kinases and protein phosphatases in *Leishmania mexicana*.

The 318 validated protein kinase phosphopeptides carried a total of 385 phosphorylation sites, 338 being non-redundant, while the 81 validated protein phosphatase phosphopeptides carried 95 phosphorylation sites, of which 85 were non-redundant. The position of phosphorylation sites were, in some cases, expected. The MAP kinases are known to contain a TXY-motif, which is often phosphorylated on the tyrosine (Y) and/or threonine (T) residue(s) to regulate the activity these MAP kinases. Thirteen MAP kinases (MPK1, MPK3, MPK4, MPK5, MPK6, MPK8, MPK9, MPK10, MPK11, MPK12, MPK13, MPK14, and MPK15) were identified in this study, and with the exceptions of only MPK8 and MPK15, all were identified with one, or typically two, phosphorylations in this motif. When digested with trypsin only, the sequence around the TXY-motif in MPK8 generates a huge peptide (44 aa residues), thus making it impossible to analyse by ordinary CID-type LC-MS/MS.

Initially, analyses were only carried out on wild type promastigotes, axenic amastigotes and lesion-derived amastigotes, but in assistance of a separate project, LC-MS/MS analyses were conducted on whole cell lysates of ΔMPK3 and ΔMKK promastigotes. The number of protein kinases and phosphatases as well as their phosphorylation pattern differed, as expected, between the wild type and kinase deletion mutants (**Figure 3.9**). The effects seemed most pronounced in the ΔMKK promastigotes, and triggered further analyses of the MKK-deletion mutant as well as another MAP kinase-kinase deletion mutant strain, ΔPK4. Some protein kinases, e.g. MPK9, MPK10, and GSK3β, were readily identified in the majority of the samples, whereas other, e.g. PK4 and MKK, were not detected in any of the general analyses. This led to application of alternative approaches to get an idea of the proteins' presence and phosphorylation status. Genetically modified parasites with green fluorescent protein (GFP)-tags of 5 MAP kinases and MAP kinase-kinases were available. MPK5 GFP, MPK1 GFP, MKK GFP, PK4 GFP and GFP MPK3 promastigotes were cultured for alternative analyses on the side. The main objective for the work with the GFP-tagged parasites was to investigate the potential of alternative protein-specific enrichment procedures, so that if it worked

the enriched kinases could be subjected to LC-MS/MS. Apart from fluorescent microscopy analyses indicating the cellular location of the different MAP kinases, and magnetic immunoprecipitation, which proved less successful (results not included), whole cell lysate of samples from these mutants were also analysed by LC-MS/MS. These whole cell lysates were derived from perfect late log-phase promastigotes only in the case of the MKK GFP mutant, whereas the rest were of stationary or borderline stationary (GFP MPK3) promastigotes. Phosphorylations in MPK1, MPK3 and MPK5 had already been identified from previous LC-MS/MS analyses of wild type samples. Due to the unknown effect of samples being derived from stationary promastigotes, it was decided to initially only analyse one sample derived from stationary promastigotes (MPK1) and one derived from borderline stationary promastigotes (MPK3), along with the MKK sample. MPK1 was not detected, but MPK3 and MKK were. These results indicated that the samples derived from stationary cells would not improve detection, as the GFP-tagged proteins were only detected in samples from late log-phase or borderline stationary cells. With numerous other samples of higher priority being destined for LC-MS/MS analyses and only limited instrument time, the remaining GFP samples of stationary promastigote origin were precluded from further analysis. Whether the GFP tag in any way improved detection of the kinases in whole cell lysate is impossible to assess by identification in just a single good-quality sample type.

The validated phosphopeptides and phosphorylation sites of *Leishmania mexicana* protein kinases and phosphatases are listed in **Appendix E** (http://dl.dropbox.com/u/3011619/Appendix%20E.xlsx).

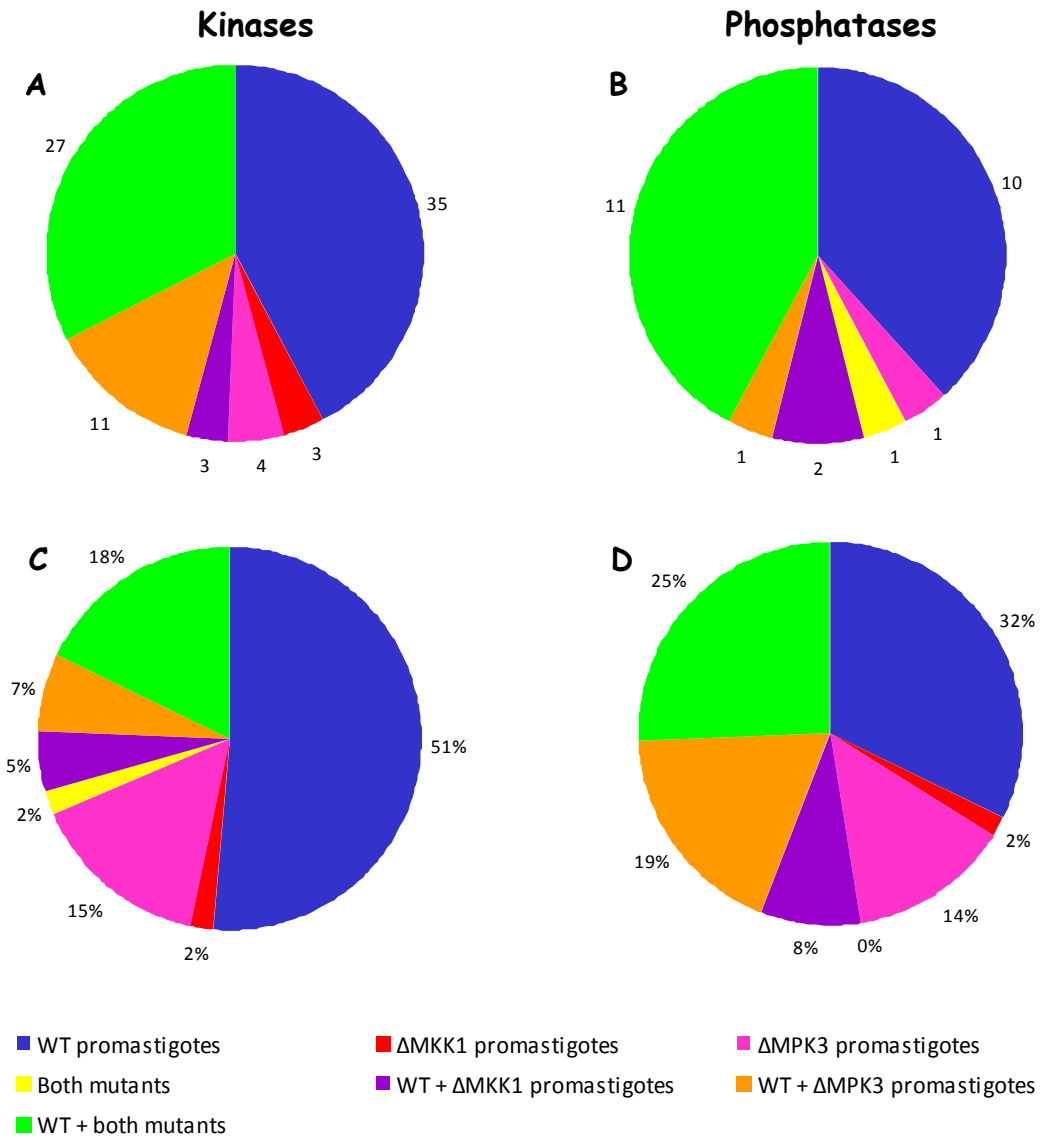


Figure 3.9: Differences between wild type and MAP kinase deletion mutants in initial experiments. Charts **A** and **C** represent the distribution of protein kinases and their phosphopeptides, respectively, and charts **B** and **D** the distribution of protein phosphatases and their phosphopeptides, respectively. The numbers of identified protein kinases and phosphatases is lower than the numbers stated in the beginning of this section because these experiments were conducted 1½ years ago (see 20 month report [188].

3.5.4 Membrane protein extracts

Proteins associated with cellular membranes, whether those be the plasma membrane or organellar membranes, are vital in many signalling networks [32]. A typical eukaryotic phosphorylation-controlled signalling pathway is initiated by activation of transmembrane receptor tyrosine kinases, which then attracts substrate kinases [32]. *Leishmania* are devoid of traditional receptor tyrosine kinases [25], but this does not necessarily exclude these parasites from having other membrane-lodged or –associated protein kinases. To investigate if it was possible to identify such potential kinases, or other proteins that might play a role in signal transduction initiation, extraction of membrane proteins by ultracentrifugation was employed. Extraction of membrane proteins simultaneously generated fractions consisting of cytosolic proteins as well proteins of the nucleus and mitochondria, respectively. Apart from some expected protein-specific differences described later in this section, membrane protein extraction also revealed another curious difference between axenic and lesion-derived parasites: Upon ultracentrifugation, the axenic parasites (promastigotes as well as amastigotes) would generate a two-phased supernatant, with one “cloudy” phase in the top, and a clear phase below. For the lesion-derived amastigotes, the supernatant of soluble proteins was all clear as anticipated (**Figure 3.10**).

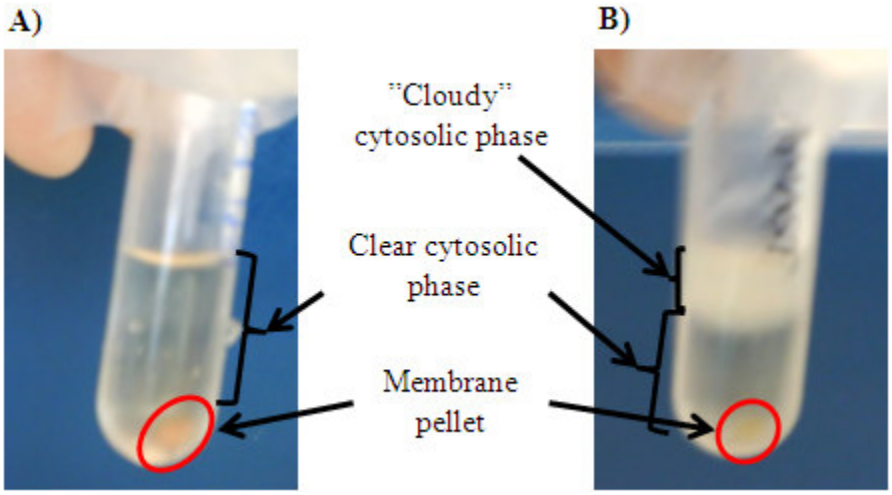


Figure 3.10: Differences in membrane protein extraction from lesion-derived (A) and axenic amastigotes (B) after ultracentrifugation. Ultracentrifugation of any of the axenic promastigote or amastigote samples would result in a small membrane protein pellet with a supernatant containing cytosolic proteins in a clear as well as a "cloudy" phase (**B**). In contrast to this, the supernatant of lesion-derived amastigote samples would be all clear (**A**). The relatively poor resolution of the pictures is due to the fact that they were snapped with a mobile phone camera in the hurry of a busy experimental plan. Apart from cropping to fit the figure, the pictures have not been processed or manipulated in any way.

The main objective with the different protein fractions generated by the membrane protein extraction procedure was to carry out iTRAQ analyses (section **3.6.1**), but not least due to the finding of different phases of soluble proteins in the axenic parasite samples, the fractions were also analysed by TiO_2 chromatography alone. Both the "clear" and the "cloudy" fraction of cytosolic proteins observed in the axenic samples contained proteins. A comparison of these fractions from WTpro and WTamast showed a significant overlap in the identifications (**Table 3.2**). For promastigotes, this comparison encompassed 6 datasets from "clear" cytosol fraction and 6 datasets from the "cloudy" cytosol fraction (4 TiO_2 enrichments and 2 un-enriched samples for both fractions). For amastigotes, only 4 datasets were available for each cytosol fraction (in both cases 2 TiO_2 enrichments and 2 un-enriched samples).

	All proteins				Phosphoproteins			
	Clear	**Shared**	**Cloudy**	**Total**	**Clear**	**Shared**	**Cloudy**	**Total**
Pro	5	529	14	**548**	1	171	2	**174**
Amast	9	346	5	**360**	2	79	1	**82**

Table 3.2: Distribution of identified proteins between the "clear" and "cloudy" cytosolic protein phases in axenic promastigotes and amastigotes. The "clear" and "cloudy" columns, respectively, represent the number of proteins identified only in these fractions, whereas the "shared" column represents proteins identified in both fractions.

The presence of the “clear” and “cloudy” fractions is not readily explainable by protein differences in the two fractions. The proteins unique to each fraction (**Table 3.2**) seemed rather coincidental as there were no specific characteristics suggesting otherwise, and these proteins were also identifiable in the whole cell extracts. Other potential explanations could be reactions between remnants of culture medium-derived additives and the reagents used in the membrane protein extraction procedure, or a different lipid profile of axenic amastigotes generating a lipid-protein slurry surfacing the otherwise soluble cytosolic protein phase.

In contrast to the “clear” and “cloudy” fractions only presenting few proteins that could not be found by WCL analyses, the membrane protein fractions added a number of new identifications. The proteins identified in the membrane fractions included numerous transporter proteins, e.g. glucose and amino acid transporters, but also protein kinases such as MPK14 that had not been identified in any of the previous WCL analyses. **Table 3.3** lists the different phosphorylated transporter proteins identified in the membrane protein fractions.

Table 3.3 – legend on p. 125

	Protein	Lmex Contig	Lmx accession	Similar to	Phospho-sites
1	Vacuolar-type proton translocating pyrophosphatase 1, putative	Contig_0002385_294	LmxM30.1220	LmjF31.1220	S215, S747
2	Glucose transporter, LmGT1	Contig_0002327_1372	LmxM36.6300	LmjF36.6300	T592, S601, S602, S603, S605
3	Folate/biopterin transporter, putative	Contig_0001524_180	LmxM10.0370	LinJ10_V3.0390	S16*, S305, S309, S311, S669, S670, S673
4	Nucleobase transporter	Contig_0001398_178	LmxM13.1210	LinJ13_V3.1110	S297, S301, T316, S318, S320, S330
5	P-type H+-ATPase, putative	Contig_0001005_196 + Contig_0001004_51	LmxM18.1510	LmjF18.1510	S21, T26, S939, S941, S942, S944, S957, S963
6	Calcium motive p-type ATPase, putative	Contig_0001945_132	LmxM34.2080	LmjF35.2080	S143, T163, S869
7	Nucleoside transporter 1, putative	Contig_0001776_393	LmxM36.1940	LmjF36.1940	S300, S303, T330
8	Vesicle-associated membrane protein, putative	Contig_0001251_315	LmxM08.0030	LmjF08.0030	T97, S99, S158, T160
9	Calcium channel protein, putative	Contig_0000458_217	LmxM33.0480	LinJ34_V3.0500	T732, S848, S1283

10	Cation-transporting ATPase, putative	Contig_0002253_2781	LmxM07.1050	LinJ07_V3.1210	T320
11	ABC transporter, putative	Contig_0002047_769	LmxM25.0530	LinJ25_V3.0540	T569, S574
12	P-glycoprotein e	Contig_0001550_415	LmxM30.1270	LinJ31_V3.1290	S967, T970, S1389, S1390
13	ADP/ATP mitochondrial carrier-like protein	Contig_0000977_440	LmxM14.0990	LinJ14_V3.1050	S227
14	Chloride channel protein, putative	Contig_0002274_36	LmxM04.1000	LmjF04.1000	S711, S734, S950, S952
15	Amino acid permease, putative	Contig_0001372_415	LmxM22.0230	LinJ22_V3.0100	S23, T25, S28
16	1-acyl-sn-glycerol-3-phosphateacyltransferase-like protein, putative	Contig_0001875_1476	LmxM31.1960	LmjF32.1960	S146
17	Transporter, putative	Contig_0001436_281	LmxM19.0760	LinJ19_V3.0760	S205, S206, S355, T358
18	ATP-binding cassette protein subfamily C, member 2, putative	Contig_0001810_542	LmxM23.0220	LinJ23_V3.0240	S224, S639, S649, S651, S906
19	ATP-binding cassette protein subfamily C, member 8, putative	Contig_0000028_1872 + Contig_0000027_91	LmxM33.0670	LmjF34.0670	T356, S1746, S1750, S1952, S1956
20	Glucose transporter, lmgt2, putative	Contig_0002325_540	LmxM36.6290	LmjF36.6290	S14
21	Tricarboxylate carrier, putative	Contig_0001928_285	LmxM01.0570	LinJ01_V3.0590	S315
22	Pteridine transporter ft6, putative	Contig_0001526_1091	LmxM10.0360	LmjF10.0360	S267, S269, S633
23	Amino acid transporter aATP11, putative	Contig_0002382_1330	LmxM30.0580	LinJ31_V3.0610	S40, S109

24	Na/H antiporter-like protein	Contig_0000006_932	LmxM23.0830	LinJ23_V3.1000	S660, S1099, S1494
25	Dolichyl-P-Man:GDP-Man5GlcNAc2-PP-dolichyl alpha-1,2-mannosyltranslocase, putative	Contig_0001505_1669	LmxM28.2410	LmjF28.2410	S93
26	Phospholipid transporting ATP-like protein, putative	Contig_0001991_289	LmxM09.0890	LmjF09.0890	T22, S23, S233
27	Cation transporter, putative	Contig_0001304_473	LmxM15.1310	LmjF15.1310	S387, S415
28	Hypothetical protein, conserved (ion channel?)	Contig_0001885_1548	LmxM14.0530	LmjF14.0530	T361, S364, T383, S388
29	P-type ATPase, putative	Contig_0001077_588	LmxM17.0600	LinJ17_V3.0660	S442, T444
30	Amino acid transporter aATP11, putative	Contig_0001841_452	LmxM30.0350	LinJ31_V3.0370	S56, S59, S69
31	Pteridine transporter ft5, putative	Contig_0001524_226	LmxM10.0400	LmjF10.0400	S301, S332, S336
32	MFS transporter, putative	Contig_0001587_116	LmxM03.0410	LmjF03.0410	S72, S309, S316, S318, S350, S363
33	Hypothetical protein, conserved	Contig_0001135_700	LmxM34.3580	LinJ35_V3.3630	S16, S20, S522
34	Phosphate-repressible phosphate permease-like protein	Contig_0001530_604	LmxM10.0030	LinJ10_V3.0010	S267, T269, S272, T273, S297, S300, S302, S305
35	Amino acid permease	Contig_0000173_15	LmxM30.1800	LmjF31.1800	S213, S644
36	Transporter, putative	Contig_0001826_2094	LmxM34.0080	LinJ35_V3.0080	T304, S313

37	Aminophospholipid translocase, putative	Contig_0000543_115	LmxM33.3220	LinJ34_V3.3000	T544
38	ABC transporter-like protein	Contig_0001808_1224	LmxM23.0380	LinJ23_V3.0430	S531, T549
39	Cation transporter, putative	Contig_0001169_1628	LmxM19.1380	LinJ19_V3.1420	S463, S467, S493, S498
40	MGT2 magnesium transporter	Contig_0002056_549	LmxM25.1090	LinJ25_V3.1130	S20
41	ATP-binding cassette protein subfamily A, member 2, putative	Contig_0001450_24	Lmx11.1220	LinJ11_V3.1210	S1552
42	ATP-binding cassette protein subfamily C, member 1, putative	Contig_0001811_2032	LmxM23.0210	LinJ23_V3.0230	S630, S640, S642, S883, S885
43	Qb-SNARE protein, putative	Contig_0000021_61	LmxM23.1740	LmjF23.1740	S130, T132, S212

Table 3.3: Transporter proteins identified in the membrane fractions. Forty-three membrane-associated phosphorylated proteins with transporter functions were identified in the membrane protein extracts. The proteins are listed with information about which *L.mexicana* contig(s) cover(s) their sequence, the *Lmex* predicted protein accessions, as well as *L.major* or *L.infantum* protein accessions for their orthologues, which occasionally have been used to name the *Lmex* proteins. The position of the phosphorylation sites are given based on the *Lmex* predicted protein sequences, with the exception of protein 3, where the first phosphorylation site (marked with *) is positioned in a contig sequence part not included in the predicted protein sequence. Instead, the position of this phosphorylation site is relative to the first methionine residue in the contig (there are 2 Met residues prior to the one initiating the predicted protein sequence). The positions of the subsequent phosphorylation sites in this protein are also relative to this initial Met residue of the contig.

3.6 Quantitative analyses

To assess the significance of the different proteins identified as well as the apparent differences between the two life stages and between wild type and kinase deletion mutants, a number of quantitative proteomics and phosphoproteomics analyses were conducted.

3.6.1 iTRAQ

Isobaric tags for relative and absolute quantitation (iTRAQ) were used to compare amounts of phosphopeptides in cytosolic protein fractions as well as evaluate the reproducibility of the TiO_2 phosphopeptide enrichments.

3.6.1.1 Evaluation of TiO_2 chromatography reproducibility by iTRAQ

To evaluate the reproducibility of the applied TiO_2 procedure for phosphopeptide enrichment, 4 × 100 µg whole cell lysate protein digest from wild type promastigotes were labelled with iTRAQ 4-plex. The 4 differently labelled samples were enriched for phosphopeptides by TiO_2 prior to mixing 1:1:1:1 for LC-MS/MS analysis. Labelling of the samples had been checked by MALDI MS/MS (1 µg/sample) prior to mixing. The final mixture thus corresponded to 396 µg protein, which was split between two LC-MS/MS analyses (2h gradient). The timed ion chromatograms of the two LC-MS/MS runs were very similar, and this was confirmed by subsequent one-way ANOVA analysis (see **Figure 3.11**) of the iTRAQ ratio distribution. The raw data were processed in Proteome Discoverer where ratios were calculated with the 114-labelled sample as reference. In theory, all ratios should be just around 1.0, or at least similar, however this was only the case for approximately one third of the identified and quantified peptides. Two different patterns dominated the remaining two thirds of the data: One where the 115/114 and 116/114 ratios would be similar and the 117/114 ratio then significantly higher, and another where 115/114 < 116/114 < 117/114. Overall, the results indicated

that the 115-labelled enrichment was the least effective, with the 114- and 116-labelled enrichments being rather equal and the 117-labelled enrichment the most effective, which is also seen of the means comparison chart in **Figure 3.12**. There are a number of possible reasons for this, incl. 1) different peptide contents of the different aliquots used; 2) minor differences in labelling efficiency too small to observe in the pre-mixing test of a few individual peptides, but subsequently adding up; 3) differences in the TiO_2 columns (length, amount, how tightly they were packed, etc.) too small to assess by the bare eye, etc.

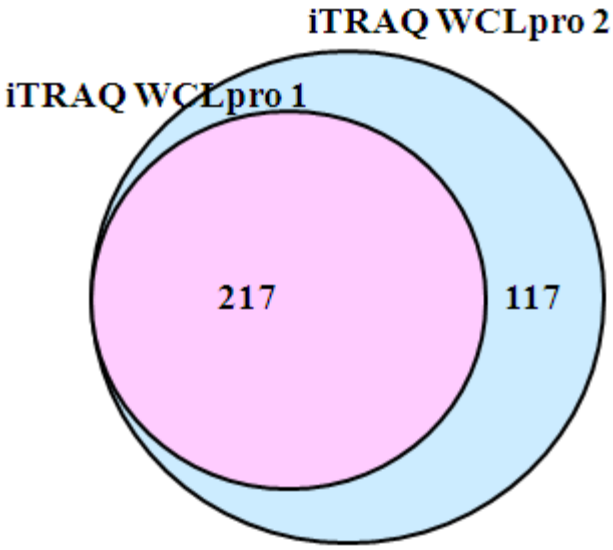


Figure 3.11: Overlap between the two LC-MS/MS runs of iTRAQ WCLpro. The similarity of the chromatograms was reflected in the peptide identifications (only phosphopeptides assessed) by 100% overlap between the first and the second run. The second run led to identification of an additional 117 phosphopeptides, which could be due to low abundant phosphopeptides retained on the analytical column after the first run being eluted together with their peers in the second run to become abundant enough to be selected for fragmentation. Another possible explanation relates to the spacial distribution of peptides in the sample volume positioned in the autosampler micro titer plate, where peptides might sink lower into the well concurrently with the passing of time since being placed in there. Thus, when time passes for other samples to be analysed before the current one, the spacial distribution of peptides may no longer be even and differences arise in the analyses.

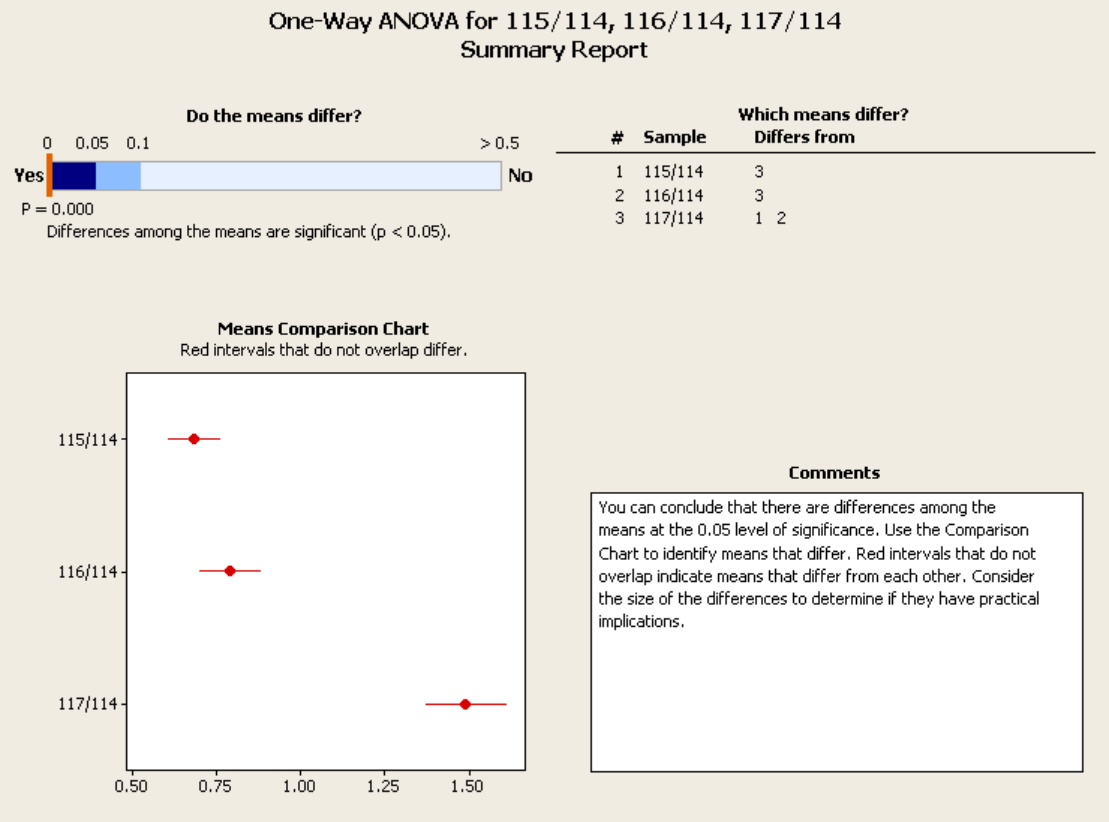
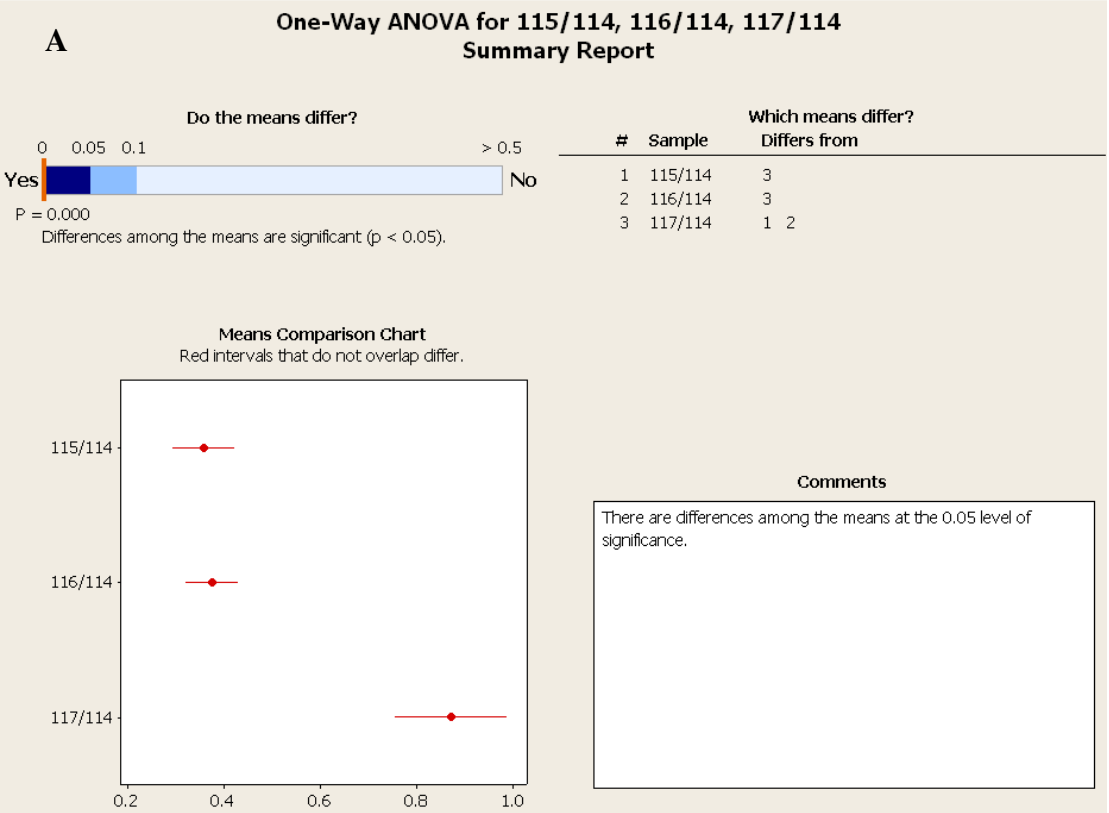


Figure 3.12: Reproducibility of TiO_2 chromatography-based phosphopeptide enrichment assessed by iTRAQ. Four aliquots of wild type promastigote whole cell lysate were iTRAQ-labelled (114, 115, 116, and 117, respectively), and subjected to individual TiO_2 chromatography prior to sample mixing. Duplicate LC-MS/MS analyses were conducted on the samples. The data were processed in Proteome Discoverer, and manually filtered to only contain quantified phosphopeptides with expected value thresholds less than 1E-05, translating into 51-55 high confidence phosphopeptide identifications for each of the ratios. A one-way ANOVA test with 95 % confidence was used to assess differences in ratio- as well as LC-run-specific means. The ANOVA test showed that the results of the two LC-runs are almost identical, and that the different TiO_2 enrichments were related like this: $115 < 114 \approx 116 < 117$. The statistics analysis was carried out in Minitab 16.1.1 [193].

3.6.1.2 iTRAQ analysis of promastigote clear cytosol fractions

Aliquots of 100 µg protein digest (as determined by amino acid analyses) from clear cytosol fractions of WTpro, ΔPK4pro, ΔMKKpro and WTpro WCL were labelled with iTRAQ reagents as displayed in **Table 2.6**. The WTpro WCL sample was used as a common reference in all iTRAQ cytosol analyses (sections **3.6.1.3** and **3.6.1.4** as well). Labelling efficiency was checked by MALDI MS/MS prior to mixing the differently labelled samples 1:1:1:1. The mixture was subjected to TiO_2 chromatography for phosphopeptide enrichment. Triplicate LC-MS/MS analyses of the C8-C18 desalted TiO_2 eluates were conducted from a total of 792 µg protein mixture, and data were processed in Proteome Discoverer. Three hundred and thirteen peptides were identified, 208 of these phosphorylated, and 86 unique phosphopeptides were quantified. One-way ANOVA analysis with a significance level of 0.05 for mean differences of 2 or above showed that the ΔMKKpro clear cyt/WTpro WCL-ratio sticks out (**Figure 3.13A**), and the reason for this becomes apparent when looking at the distribution of 117/114-ratio values in **Figure 3.13B**. The distribution of ratios for WTpro clear cyt (116) and ΔPK4pro clear cyt (115) (**Figure 3.13B**) indicates that differences at the phosphopeptide level between these two sample types are minute. This is even more apparent if comparing ΔPK4pro clear cyt and ΔMKKpro clear cyt to WTpro clear cyt (**Figure 3.14**), where it is seen that the ΔPK4-WT relation is very even, whereas the ΔMKK-WT one is more scattered. These distributions, however, mask the fact that ratio differences are present, e.g. the ratios of ΔPK4/WT are 0.07-1.82, indicating potentially quite significant differences in phosphopeptide abundance between the two sample types.

Figure 3.13 - legend on page 132.



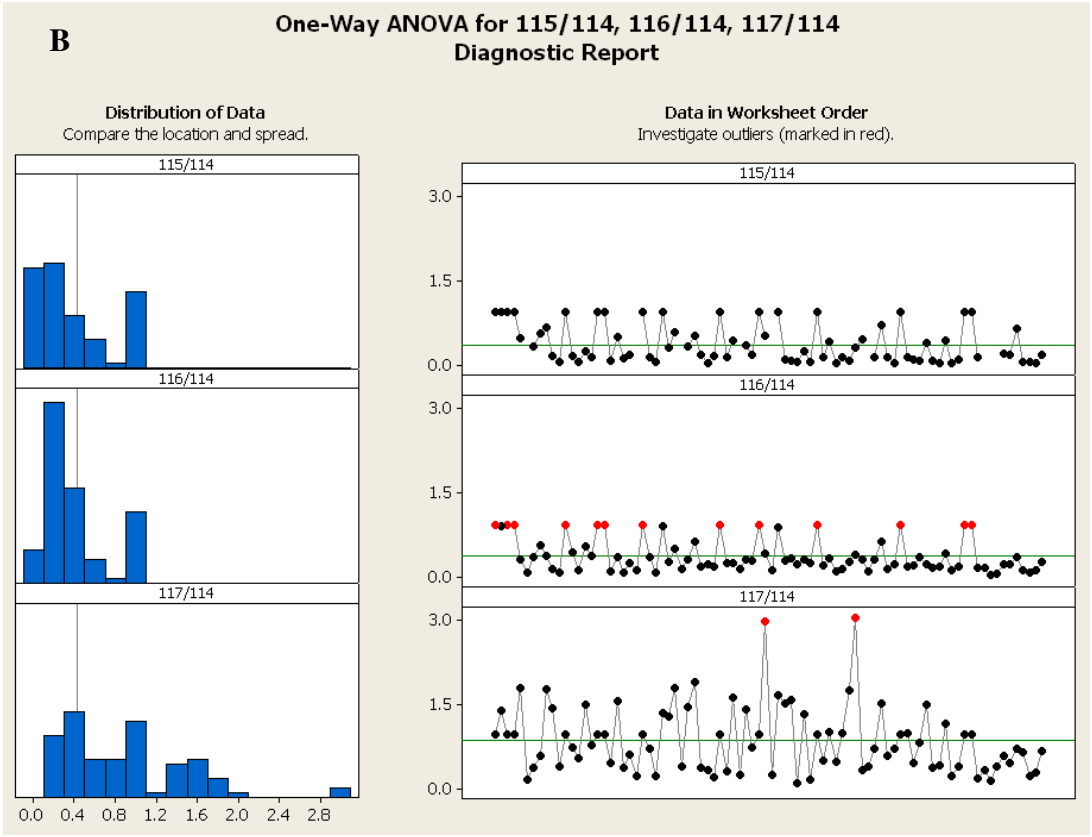


Figure 3.13: Statistical analysis of the promastigotes clear cytosol iTRAQ experiment. The means comparison chart in **A** shows that the ratios of ΔMKKpro clear cyt/WTpro WCL significantly differ from those of WTpro and ΔPK4pro. The reason for this difference is apparent in the outliers report in **B**, where it is seen that the ratio distribution in 117/114 is very different from those of 115/114 and 116/114.

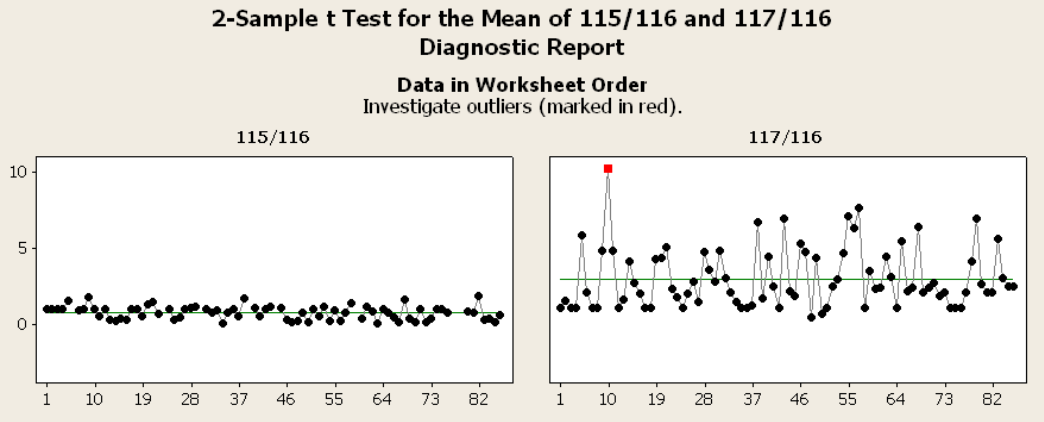


Figure 3.14: 2-sample t-test showing the distribution of ratios for the promastigote clear cytosol iTRAQ experiment, when WTpro clear cytosol is used as reference. 115: ΔPK4pro clear cyt; 116: WTpro clear cyt; 117: ΔMKKpro clear cyt. From the charts it is apparent that the ΔPK4pro-WTpro ratios all fall within a relatively tight area (0.07-1.82), whereas the ΔMKKpro-WTpro ratios are a lot more scattered (0.46-10.22).

The phosphopeptides displaying the most significant differences, in relation to WTpro WCL as well as WTpro clear cyt, are listed in **Table 3.4**. Only phosphopeptides displaying a ΔMKK/WTpro ratio > 2 have been included. The three peptides are derived from three different hypothetical proteins, neither of which have any functional information listed (to *Lmajor* orthologue) at GeneDB [182].

Peptide	Protein	115/114	116/114	117/114
qIDkVTESVA**t**ISMHEEPSETR	Hypothetical protein, conserved (LmxM25.1060)	0.507	0.431	**3.006**
nLP**s**VDDGLYPk	Hypothetical protein, conserved (LmxM27.1730)	0.290	0.398	**3.058**
		115/116	**117/116**	
kAPIADSD**s**DDDEPVR	Hypothetical protein, conserved (LmxM34.4380)	0.993	**10.217**	

Table 3.4: Peptides displaying significant regulation compared to the others in the individual series. The lower case letters indicate modified amino acid residues (iTRAQ for N- and C-terminal residues, and phosphorylation for intra-sequence serine or threonine residues). 114: WTpro WCL; 115: ΔPK4pro clear cytosol; 116: WTpro clear cytosol; 117: ΔMKKpro clear cytosol. No functional information is stated for any of these hypothetical proteins.

3.6.1.3 iTRAQ analysis of amastigote clear cytosol fractions

Analyses similar to those described in section **3.6.1.2** above were also carried out on amastigote clear cytosol fractions. Labels can be seen in **Table 2.6**. Five hundred fifty peptides were identified, 467 of these phosphopeptides, and a total of 216 unique phosphopeptides were quantified. Statistics analysis confirmed the notion that there were significant differences between the ratios. **Table 3.5** shows which peptides were consistently up-regulated (more than 2-fold) in the amastigote clear cytosol fraction compared to the WTpro WCL reference.

Peptide	Protein	115/114	116/114	117/114
vTESVATI**s**IHEEPSETR	Hypothetical protein, conserved (LmxM25.1060)	**4.202**	**4.434**	**2.591**
kLSPSEPNVA**y**IcSR	GSK3β (LmxM18.0270)	**12.906**	**11.9**	**6.716**
lSTA**s**QGEGLEDLLLk	Hypothetical protein, conserved (LmxM06.0390)	**7.736**	**5.34**	**3.216**
vRSEDD**s**SADMV	Hypothetical protein, pseudogene (LmxM33.3640)	**3.076**	**6.492**	**5.78**
lSHS**s**EMSR	Heat shock protein HSP70 (LmxM26.1960)	**7.086**	**5.763**	**4.446**

Table 3.5: iTRAQ phosphopeptides consistently up-regulated in amastigote clear cytosols compared to promastigote whole cell lysate reference. All peptides showed a more than 2-fold up-regulation. 114: WTpro WCL; 115: ΔPK4amast clear cytosol; 116: WTamast; 117: ΔMKKamast clear cytosol.; GSK3β: glycogen synthase kinase 3 beta.

Table 3.5 reveals that even for peptides shared between all sample types, abundance variation is still present. For the first two peptides (of LmxM25.1060 and LmxM18.0270), their abundance appear relatively constant between WTamast and ΔPK4amast, whereas ΔMKKamast can only muster half the amount. The opposite relation is seen for the fourth peptide (of LmxM33.3640), where WTamast and ΔMKKamast display similar ratios about twice as high as for ΔPK4. In general, few peptides displayed significant regulation in only one of the sample types. ΔPK4 amastigotes only had 2 peptides with more than 2-fold regulation that did not show as high regulation in WT and ΔMKK, and ΔMKK amastigotes had 6. The number of peptides with ratios >2 in ΔPK4 and ΔMKK compared to wild type amastigotes also tells that phosphorylation is impacted in both the deletion mutants (**Table 3.6**).

Table 3.6 – Legend on p. 140.

Peptide	Protein	115/114	116/114	117/114
eYDLSGLFDGH**s**PR	Coronine-like protein (LmxM23.1165)	1.115	**2.011**	**2.371**
lMVGQLGDSL**t**AEDGk	Endoribonuclease L-PSP (pb5), putative (LmxM23.0200)	0.099	**4.892**	0.084
eVSGFAE**s**FEETGGAPHTPR	Hypothetical protein, conserved (LmxM09.111)	0.289	**6.529**	0.505
dAGAS**s**PTTPPPR	Hypothetical protein, conserved (LmxM21.0825)	1.592	**2.514**	**4.195**
s**D**s**P**DIHDTPTPLR	Hypothetical protein, conserved (LmxM22.0730)	0.989	1.572	**2.889**
iTESVATI**s**IHEEPNESR	Hypothetical protein, conserved (LmxM25.1060)	**2.786**	1.745	1.891
vTESVATI**s**IHEEPSETR	Hypothetical protein, conserved (LmxM25.1060)	0.1	**3.45**	0.257
yLEGLQ**s**EPGTGR	Hypothetical protein, conserved (LmxM29.0770)	-	**7.868**	0.401
s**P**s**N**HSVAAPVGR	Hypothetical protein, conserved (LmxM32.1035)	0.54	**2.845**	0.129
iDELEQSIDNLMQQ**s**GGDQGEkPAAR	Hypothetical protein, conserved (LmxM34.2680)	**2.494**	1.96	1.05
kAPIAD**s**D**s**DDDVPVR	Hypothetical protein, conserved (LmxM34.4380)	0.87	**2.727**	**2.788**
iEELVAEVDGMA**s**ENRR	Hypothetical protein, unknown function (LmxM27.0240)	**2.786**	**4.965**	0.822

aPADLS**s**YESVYAk	Microtubule-associated protein, putative (LmxM05.0380)	0.748	**6.924**	0.703
aDELAcWTSHSV**s**QIYE	Nucleoside diphosphate kinase b (LmxM34.3870)	0.773	**6.072**	1.139

Table 3.6: Phosphopeptides displaying significant fold changes. Protein levels in amastigote clear cytosol fractions of ΔPK4 (115), WT (116), and ΔMKK (117) in relation to WTpro whole cell lysate (114). Fold changes above 2 are in bold. The lower case letters in the peptide sequences represent modified amino acids. All peptides are iTRAQ modified on the N-terminal amino acid residue and some on the C-terminal amino acid residue as well. Lower case letters not positioned in the peptide terminals are either phosphorylated (s), carbamidomethylated (c), or iTRAQ labelled (k).

3.6.1.4 iTRAQ analysis of axenic and lesion-derived amastigote cytosol

The final iTRAQ analyses involved the axenic amastigote clear cytosol fraction and lesion-derived amastigote cytosol fraction, again with WTpro WCL as the reference. Labels can be seen in **Table 2.6**. A total of 647 peptides, hereof 315 phosphopeptides, were identified, and 326 of these were unique and quantifiable. Filtering of these data left 50 phosphopeptides. The general trend for all these peptides is that the axenic amastigotes will display higher ratios than the lesion-derived amastigotes. A 2-sample T-test confirmed this observation. **Table 3.7** displays those phosphopeptides that express the most significant ratio differences.

Table 3.7 – Legend on p. 143.

Peptide	Protein	116/115	117/115	116/117
kAPIADSD**s**DDDVPVR	Hypothetical protein, conserved (LmxM34.4380)	4.237	5.921	0.716
yLEGLQ**s**EPGTGR	Hypothetical protein, conserved (LmxM29.0770)	22.314	14.881	1.500
eVTGkEADD**s**DGEDDSTAAFIk	Hypothetical protein, conserved (LmxM28.2170)	3.184	**9.812**	**0.325**
aPIADSD**s**DDDVPVR	Hypothetical protein, conserved (LmxM34.4380)	3.099	5.781	0.536
vR**s**EDDSSADMV	Hypothetical protein, pseudogene (LmxM33.3640)	3.710	4.482	0.828
TTEKEV**t**DEDEEEAk	Heat shock protein 83-1 (LmxM.32.0312)	2.415	3.043	0.794
qIDkVTESVATI**s**MHEEPSETR	Hypothetical protein, conserved (LmxM25.1060)	2.036	2.164	0.941
eRSD**s**PDIHDTPTPLR	Hypothetical protein, conserved (LmxM22.0730)	2.256	2.578	0.875
sAQcGPDE**s**DDEMR	Amastin-like protein (LmxM28.1400)	2.065	1.951	1.058
vTESVA**t**ISIHEEPSETR	Hypothetical protein, conserved (LmxM25.1060)	0.772	**5.102**	**0.151**
vTESVATI**s**IHEEPSETR	Hypothetical protein, conserved (LmxM25.1060)	0.975	**4.629**	**0.211**
eVSGFAE**s**FEETGGAPHTPR	Hypothetical protein, conserved (LmxM09.111)	0.968	**10.359**	**0.093**
lMVGQLGD**s**LTAEDGk	Endoribonuclease L-PSP (pb5), putative (LmxM23.0200)	0.594	**9.672**	**0.061**
iTESVATI**s**IHEEPNESR	Hypothetical protein, conserved (LmxM25.1060)	1.523	**8.067**	**0.189**
tASEDGHD**s**DVAANDAPGEQR	Hypothetical protein, conserved (LmxM27.0240)	1.493	**5.306**	**0.281**

eTGYYNALGV**s**PDASEDEIkR	Heat shock protein DNAJ, putative (LmxM27.2400)	0.939	**4.402**	**0.213**
iTESVA**t**ISIHEEPNESR	Hypothetical protein, conserved (LmxM25.1060)	1.131	**5.573**	**0.203**
aPADLS**s**YESVYAk	Microtubule-associated protein, putative (LmxM05.0380)	0.298	**12.970**	**0.023**
aDELAcWTSHSV**s**QIYE	Nucleoside diphosphate kinase b (LmxM34.3870)	0.673	**7.652**	**0.088**
eVtGKEADD**s**DGEDDSTAAFIk	Hypothetical protein, conserved (LmxM28.2170)	1.484	**4.702**	**0.316**
tL**s**DYNIQk	Ubiquitin-fusion protein (LmxM30.1900)	0.926	**7.760**	**0.119**

Table 3.7: A summary of iTRAQ-ratios displaying significant differences between axenic and lesion-derived amastigotes. Ratios of phosphopeptide levels of lesion-derived amastigotes cytosol (116) and axenic amastigotes clear cytosol (117) to WTpro WCL (115) showing consistent regulation are listed. Comparing the ratios of axenic and lesion-derived amastigotes reveals that the axenic amastigote ratios are more than 2-fold larger in 60% of the cases (ratios in bold). This is also illustrated by the lesion-derived-to-axenic amastigotes ratios predominantly being less than 0.5. The lower case letters in the peptide sequence indicate modified amino acid residues, the bold ones being the phosphorylated ones.

3.6.1.5 iTRAQ for protein kinases and phosphatases

The original plan for quantitative analyses with iTRAQ was to include all the fractions obtained by membrane protein extraction, i.e., the clear and cloudy cytosolic protein fractions as well as the membrane protein one. It was thought that the additional sample fractionation step would improve detection of at least some of the protein kinases and phosphatases of interest, thus enabling quantitation of these along with a range of "background" proteins. The iTRAQ experiments carried out on whole cell lysate as well as cytosolic protein fractions, however, did not provide many quantitative identifications of protein kinases or phosphatases. Only about a handful of kinases and a couple of phosphatases were routinely identitified and quantified by iTRAQ. One possible explanation to this is a low amount of the different protein phosphatases combined with the relatively low amount of starting material. None of the iTRAQ analyses were carried out on the same sample amounts as was used in the non-quantitative discovery experiments. Just as important might well be some of the iTRAQ related issues, described in section **1.5.1.1**, impairing the overall number of possible protein identifications. Because more extensive quantitative knowledge about protein kinases and phosphatases was desired, iTRAQ was discontinued even before the membrane protein fractions were analysed. The data obtained by the iTRAQ experiments still provides information about many other phosphoproteins of *Leishmania*, though.

3.6.2 Targeted phosphoproteomics on protein kinases and phosphatases

The identifications of protein kinases and protein phosphatases in promastigotes and amastigotes indicated differences in their life stage presence as well as phosphorylation levels/sites. Because the iTRAQ analyses did not provide much information (section **3.6.1.5**), these differences were assessed by selected reaction monitoring (SRM) of TiO_2 enrichments from whole cell lysates. A number of different SRM experiments were carried out, initially testing the potential of the procedure, then focusing on a specific protein kinase where the discovery findings significantly opposed statements in the literature, and finally large-scale experiments were set up to analyse the levels of all protein phosphatase phosphopeptides as well as a selected range of protein kinase phosphopeptides.

3.6.2.1 Initial testing

The initial tests of SRM for relative quantification of protein phosphorylation in protein kinases and phosphatases were performed on TiO_2 enriched samples from WT promastigotes (WTpro) and WT axenic amastigotes (WTamast). From 27 random datasets that had been processed and searched (Mascot) in Proteome Discoverer, a list of phosphopeptides and transitions was build for all present protein kinases plus a few other kinases. This resulted in a list of 64 proteins represented by 207 peptides and 1753 transitions. Optimisation of the method led to final triplicate analysis of 105 phosphopeptides from 51 proteins (593 transitions) in WTpro and WTamast. Little more than half of the proteins (28 of 51) had more than one peptide or phosphorylation site targeted. From proteins with more than a single phosphopeptide, it was apparent that phosphorylation sites are indeed very life stage specifically regulated, as 19 of the 28 proteins displayed variations in the life stage presence of the different peptides (i.e., one peptide could be significant in promastigotes while another be significant in amastigotes

and a third be present at equal levels). **Figure 3.15** displays a graphical overview of the quantitative results assessed at the peptide level. The figure shows that of those peptides displaying more than 1.5-fold regulation (red and blue columns, respectively), the majority are more abundant in promastigotes. What was also apparent from the initial testing was that multiply phosphorylated peptides were equally well targeted as their singly phosphorylated peers. This was relevant to know as many of the protein kinase phosphopeptides identified in the discovery experiments carried more than one phosphorylation site.

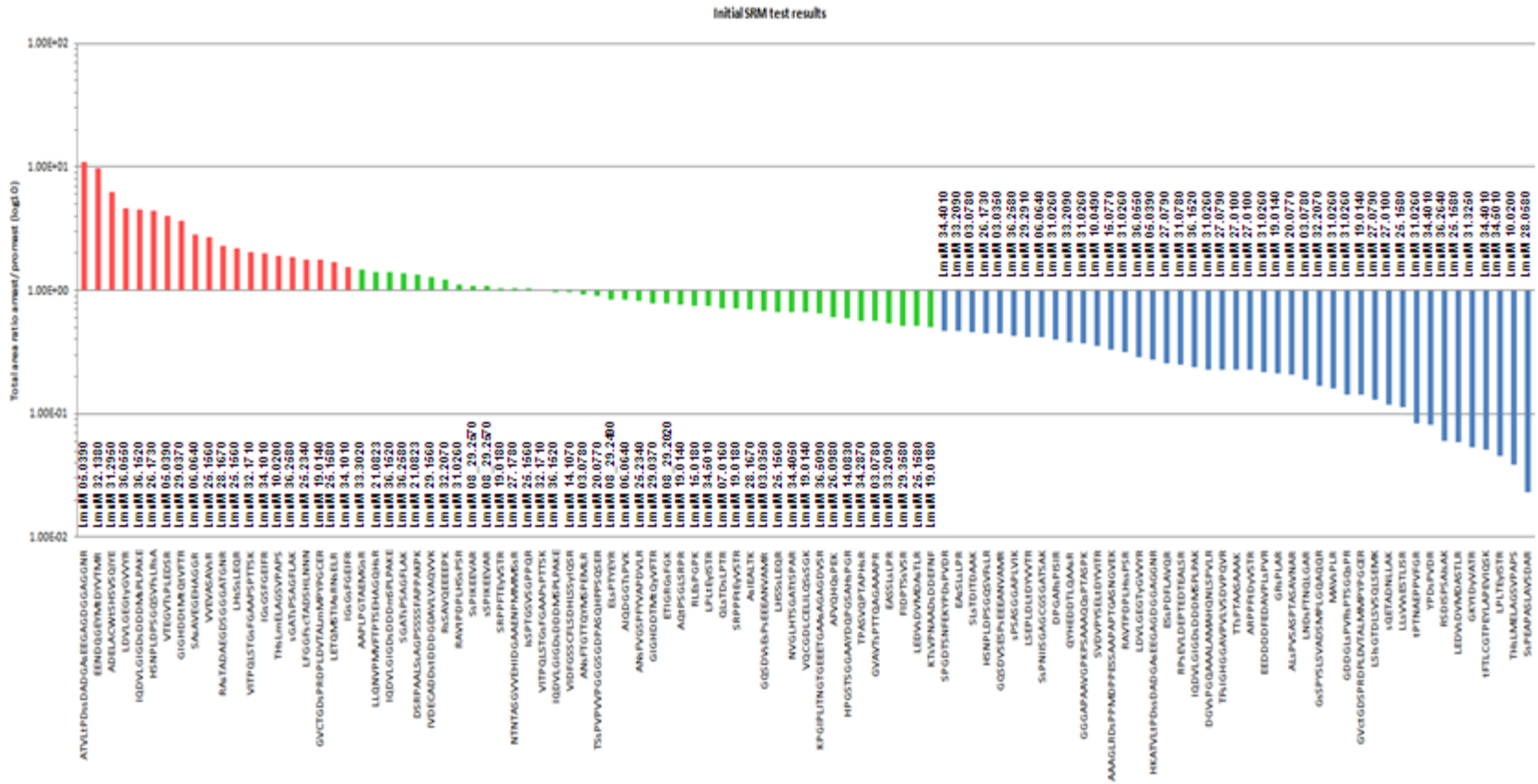


Figure 3.15: Graphical display of initial SRM testing. The red columns represent phosphopeptides more abundant in amastigotes and the blue ones phosphopeptides more abundant in promastigotes. Phosphopeptides of equal presence (ratios 0.5-1.5) are green. The lower case letters in the peptide sequences signifies modified amino acids, i.e. phosphorylation for s, t, or y, carbamidomethylation for c, and oxidation for m. *Lmex* protein accession numbers are indicated for all peptides.

3.6.2.2 MPK10 trial

The results of the initial test inspired an even more focused trial. Several phosphopeptides of *Lmex* MPK10 had been identified in the discovery experiments. In contrast to the literature notion that MPK10 phosphorylation is amastigote-specific [134, 194], the majority of the phosphopeptides were found in promastigote samples (see **Table 3.8**). This obviously encouraged quantification of the identified MPK10 phosphopeptides in promastigotes and amastigotes.

	Promastigotes	Amastigotes
TH**Y**VTHR	×	
ED**T**ADANKTH**Y**VTHR	×	
EDTADANK**T**H**Y**VTHR	×	
TH**S**LMELAGSVPAPS	×	×

Table 3.8: MPK10 phosphopeptides detected in the discovery experiments. The phosphorylated residues are shown in bold red.

The first MPK10 SRM experiment was carried out with just a single WT sample from each life stage. All four phosphopeptides mentioned in **Table 3.8** were targeted, and the results clearly indicated their presence in both life stages, even at very different levels. A mere glance at the chromatograms (**Figure 3.16**) of the two different samples confirms that the phosphorylated peptides are present in both life stages, but also indicates that variants may be present, especially in the promastigote sample. Such variants could for instance be triply phosphorylated versions of the EDTADANKTHYVTHR-peptide. It is possible to investigate such presences by SRM, even without discovery data to support the analysis, but it is no trivial task, so only one possible triply phosphorylated version of this peptide was targeted in the large-scale SRM analyses (section **3.6.2.3.1**).

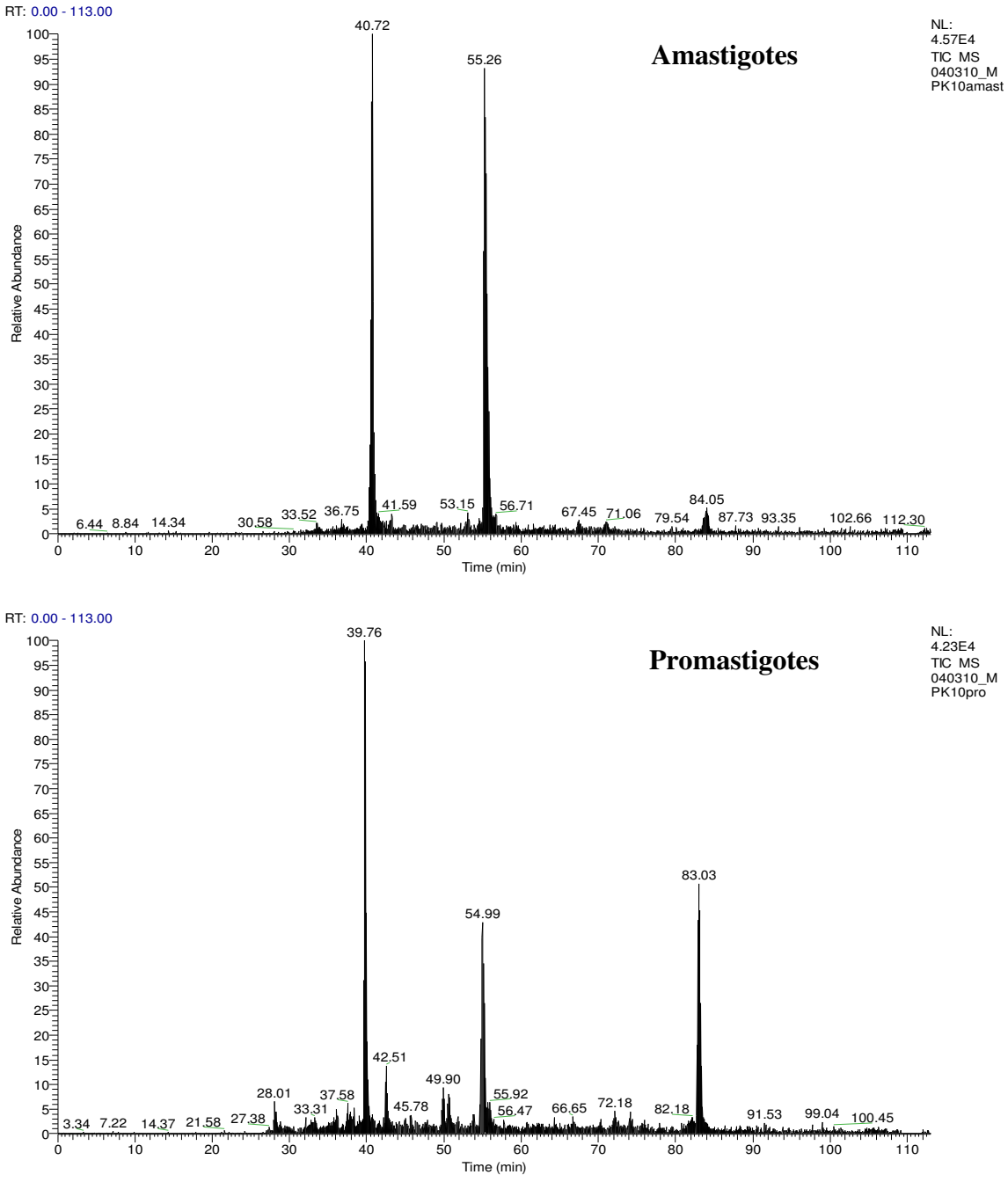


Figure 3.16: Chromatograms of the initial MPK10 SRM analysis. The MPK10 phosphopeptides (see **Table 3.8**) targeted for relative quantification by SRM were separated chromatographically. Each peak in the chromatograms corresponds to a different peptide. Differences in peak patterns and intensities are visible, as is a slight drift in retention time, which is likely LC- and not sample related.

The relative amounts of the different phosphopeptides in the two life stages are determined by integration of the SRM transition peaks. **Figure 3.17** displays the results of the relative quantification of the MPK10 phosphopeptides.

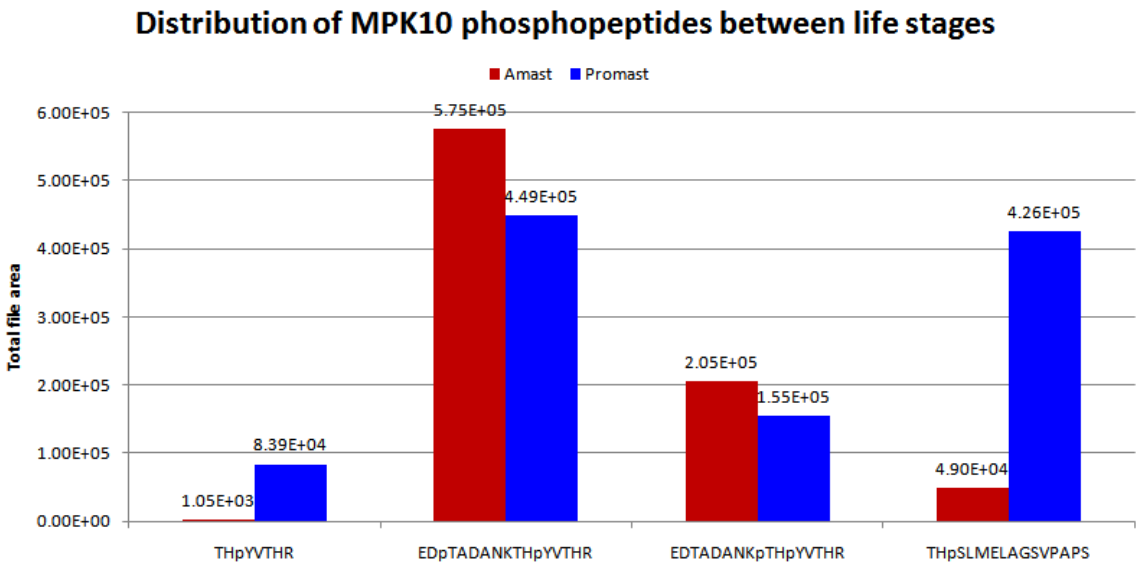


Figure 3.17: Relative quantification of MPK10 phosphopeptides. The red bars represent the peptide amount in amastigotes whereas the blue bars represent peptide amounts in promastigotes. In the peptide sequences, the phosphorylation sites are indicated by a small case p in front of the phosphorylated residue.

The significant differences in the levels of the different phosphopeptides were since confirmed in replicate analyses in the large-scale SRM analysis of selected protein kinases (section **3.6.2.3.1**).

3.6.2.3 Large-scale SRM analyses of protein kinases and phosphatases

Having proved a valuable tool for phosphopeptides quantification and possible validation, SRM was employed for thorough analyses of phosphopeptides from protein kinases and protein phosphatases identified in the discovery experiments. The relatively low number of identified protein phosphatases enabled SRM analysis of the entire collection of their phosphopeptides. Due to the large number of phosphopeptides identified for protein kinases, the initial aim was to conduct multiple SRM analyses to cover the entire collection. However, time and material constraints made this aim too ambitious, so instead analyses ended up encompassing just a selected range of the protein kinases, analysed with a larger amount of starting material. Details and results of the analyses will be mentioned in separate sections for the protein kinases and protein phosphatases, respectively, as there were some slight differences between these.

3.6.2.3.1 SRM - protein kinases

Selected reaction monitoring (SRM) of protein kinases was carried out in order to obtain relative quantitative information about, and potentially validate a few questionable phosphorylation sites in the two life stages as well as wild type and kinase deletion mutants. The selected range of protein kinases for the large-scale SRM experiment was determined based on general interest (e.g. MAP kinases and casein kinases), life stage and WT-mutant specific differences indicated by the discovery experiments, as well as the initial SRM training experiments (e.g. certain putative protein kinases). This led to a list of 59 different protein kinases, representing 167 phosphopeptide targets specified by 1287 transitions. Each phosphopeptide had 6-10 transitions to enable validation as well

as quantitation. Based on fragment ion intensities in the discovery experiments, the 2-3 most dominant ions were selected as primary transitions, and the method was tested on a mixture of WTpro and axenic WTamast WCL TiO_2 eluates. The testing indicated some issues with triggering the secondary transitions as this only happened rarely and inconsistently. There was, however, not enough instrument time or testing material to fully trouble shoot this issue, so instead all transitions were ranked as primary for those peptides of special interest (validation- or quantitation-wise). For the remaining peptides, 2-3 primary transitions were kept when the testing had indicated good signals from these. This latter group of peptides were mainly derived from putative protein kinases. TiO_2 eluates from WTpro, WTamast, ΔPK4pro, ΔPK4amast, ΔMKKpro, ΔMKKamast and WT lesion-derived amastigotes (WT_LDA) were prepared from 220 µg starting material. For each sample type, TiO_2 eluates were pooled to correspond to the phosphopeptide fraction of 440 µg starting material per LC-injection. The experimental plan aimed for triplicate analysis of all samples, however due to an unforeseen LC issue, two of the WT_LDA samples were flawed.

The analyses detected 57 of the targeted protein kinases and 149 of their phosphopeptides (**Appendix F**, http://dl.dropbox.com/u/3011619/Appendix%20F.xlsx). Ratios were calculated with WTpro as the reference for all samples as well as with WTamast as reference for the ΔPK4amast, ΔMKKamast and WT_LDA samples. These calculations displayed the differences in phosphorylation site abundance between the samples. **Table 3.9** summarises the findings.

	WTAA/ WTpro	PK4pro/ WTpro	MKKpro/ WTpro	PK4AA/ WTpro	MKKAA/ WTpro	LDA/ WTpro	PK4AA/ WTAA	MKKAA/ WTAA	LDA/ WTAA
Not detected in numerator sample	11	7	15	13	31	26	7	25	26
Ratio < 5E-1	41	33	59	43	65	64	25	74	55
5E-1≤ratio≥2.0E0	48	89	57	53	30	38	98	38	48
Ratio > 2.0E0	38	9	7	29	12	15	8	0	9
Only detected in numerator sample	11	3	3	9	7	5	3	2	5

Table 3.9: Summary of SRM results from protein kinases. The numbers in the table refer to phosphopeptides belonging to the different groups. Ratios were calculated on the total peak areas of the 2 most dominant (primary) transitions for low abundant phosphopeptides and phosphopeptides off the high-interest list, and up to 6 dominant transitions for the high abundant/high-interest ones. Abbreviations used: WTAA, wild type axenic amastigotes; WTpro, wild type promastigotes; PK4pro, PK4 kinase deletion-mutant promastigotes; MKKpro, MKK kinase deletion-mutant promastigotes; PK4AA, PK4 kinase deletion-mutant axenic amastigotes; MKKAA, MKK kinase deletion-mutant axenic amastigotes; LDA, wild type lesion-derived amastigotes.

From the table, it is apparent that fairly equal numbers of phosphopeptides are more than 2-fold up- or down-regulated in WT amastigotes compared to promastigotes (1st column in **Table 3.9**). Of those phosphopeptides more than 2-fold up-regulated, 12 belong to MAP kinases. Phosphorylation(s) in the TXY-domains of MPK3, MPK4, MPK5, MPK11 and MPK13 are all up-regulated from wild type promastigotes to amastigotes. On the other hand, CRK3, PKA1C, MPK6 and MPK14 display more than 2-fold down-regulation in WT amastigotes. MPK10 and MPK12 both hold phosphopeptides belonging each of these two categories. For MPK10, the doubly and triply phosphorylated versions of the EDTADANKTHYVTHR peptide are significantly up-regulated while the phosphopeptide displaying a single phosphorylation of tyrosine in the TXY-domain, as well as another peptide, is significantly down-regulated.

The lesion-derived amastigotes predominantly had lower amounts of the different phosphopeptides than the axenic amastigotes (90 phosphopeptides were at least 2-fold less abundant in LDAs compared to axenic amastigotes). Thirteen peptides were at least 2-fold more abundant in LDAs than in axenic amastigotes, though. Among these were 3 casein kinase peptides, two from MPK10 (singly and doubly phosphoryation of the TXY-domain) and one from MPK6.

Comparing WT amastigotes to those of the deletion mutants display some interesting findings. In wild type parasites, the singly phosphorylated version of the MKK SQESLENDVK peptide is predominant in both life stages (2.9 times more abundant than its doubly phosphorylated relative in amastigotes, and 7.5 times more abundant in promastigotes). In the PK4 knock-out mutant, however, the two versions of this phosphopeptide are almost equally present in promastigotes, while the doubly phosphorylated version is more than 2-fold down-regulated (compared to WT) and the singly phosphorylated version more than 2-fold up-regulated in amastigotes.

In the MKK knock-out mutant amastigotes, GSK3β phosphopeptides along with a range of MAP kinase-derived phosphopeptides are significantly down regulated. For MPK1, MPK3, MPK4, MPK10, MPK11, and MPK12 the affected phosphorylation sites are within the TXY-domain of the phosphorylation lip.

In promastigotes, both knock-out mutants show more than 2-fold up-regulation, compared to WT promastigotes, of the MPK4 tyrosine phosphorylation and the MPK5 threonine phosphorylation sites in the TXY-domains.

3.6.2.3.1 SRM - protein phosphatases

The background of the protein phosphatases SRM analyses was similar to what was just described for the protein kinases. Because the number of protein phosphatases and corresponding phosphopeptides in the discovery experiments was relatively low, they could all be included in the SRM set-up. The protein phosphatase SRM method contained 35 protein targets[2], representing 86 peptide targets and 592 transitions. The discovery experiments as well as a previous SRM trial (results of this not included here) had indicated that the vast majority of the identified protein phosphatases were low-abundant, even in phosphopeptide enriched fractions. Hence, the complete protein phosphatases SRM analyses were carried out with TiO_2 eluates of 550 µg starting material for every LC injection. For all samples – WTpro, WTamast, ΔPK4pro, ΔPK4amast, ΔMKKpro, ΔMKKamast, and WT_LDA – SRM analyses were carried out in triplicates.

Seventy nine phosphopeptides from 24 different protein phosphatases, i.e. all the phosphorylated phosphatases identified in the discovery experiments, were detected in the SRM analyses (**Appendix G1** (all axenic) and **G2** (WTpro-WTamast-WT_{LDA})

[2] The observant reader may notice that this number, taken directly from Pinpoint (see section **1.6.4**), is higher than the number of protein phosphatases previously stated to be identified as phosphorylated. The main reason for this is that even the slightest variation in the protein description line in Pinpoint will make for a new protein target count. Hence, some protein phosphatases are counted twice because they 1) are covered by more than one contig in the 6-frame translation library; 2) not all phosphopeptides to be analysed from a specific phosphatase would be found in the same search result file, thus enabling database updates (e.g. typo corrections in protein descriptions) as well as typos when manually entering a the protein description for a peptide to the list to play a role; and 3) not all phosphopeptides were identified in datasets based on the 6-frame translation library, the initial concatenated *all Leishmania species* database also contributed. When all these sources of ambigous contributions are sorted out, the number of targeted protein phosphatases will match the number of phosphatases identified with at least one phosphopeptide.

http://dl.dropbox.com/u/3011619/Appendix%20G1.xlsx
http://dl.dropbox.com/u/3011619/Appendix%20G2.xlsx). Their ratios in the different sample types analysed are listed in **Table 3.10**. The table confirms the general notion from the discovery experiments that overall phosphorylation in the ΔMKK mutants (both life stages, though more pronounced in amastigotes) is significantly impaired not only compared to the wild type, but also the ΔPK4 mutants.

A general trend for the protein phosphatases is that the phosphopeptides regulated more than 2-fold compared to the other life stage or the wild type, are predominantly down-regulated. This is easily seen with the ΔMKK mutants where no phosphopeptides would be more than 2-fold up-regulated compared to the wild type, but no less than 45 phosphopeptides would be more than 2-fold down-regulated.

	WTAA/ WTpro	PK4pro/ WTpro	MKKpro/ WTpro	PK4AA/ WTpro	MKKAA/ WTpro	LDA/ WTpro	PK4AA/ WTAA	MKKAA/ WTAA	LDA/ WTAA
Not detected in numerator sample	3	2	6	8	14	11	5	11	7
Ratio < 5E-1	25	13	33	34	32	35	18	34	30
5E-1≤ratio≥2.0E0	35	55	30	22	24	23	44	24	30
Ratio > 2.0E0	9	2	3	8	2	10	2	0	7
Only detected in numerator sample	0	0	0	0	0	0	0	0	1

Table 3.10: Summary of SRM results from protein phosphatases. The numbers in the table refer to phosphopeptides belonging to the different groups. Ratios were calculated on the total peak areas of all selected transitions (6-8/peptide). Abbreviations used: WTAA, wild type axenic amastigotes; WTpro, wild type promastigotes; PK4pro, PK4 kinase deletion-mutant promastigotes; MKKpro, MKK kinase deletion-mutant promastigotes; PK4AA, PK4 kinase deletion-mutant axenic amastigotes; MKKAA, MKK kinase deletion-mutant axenic amastigotes; LDA, wild type lesion-derived amastigotes.

What cannot be seen from the table is that, for those peptides detected in all sample types, the promast-amast profile registered in the wild type, e.g. peptides predominant in promastigotes and less significant in amastigotes, would also be seen in the mutants, though generally at lower levels in these. For an illustration of this see **Figure 3.18**.

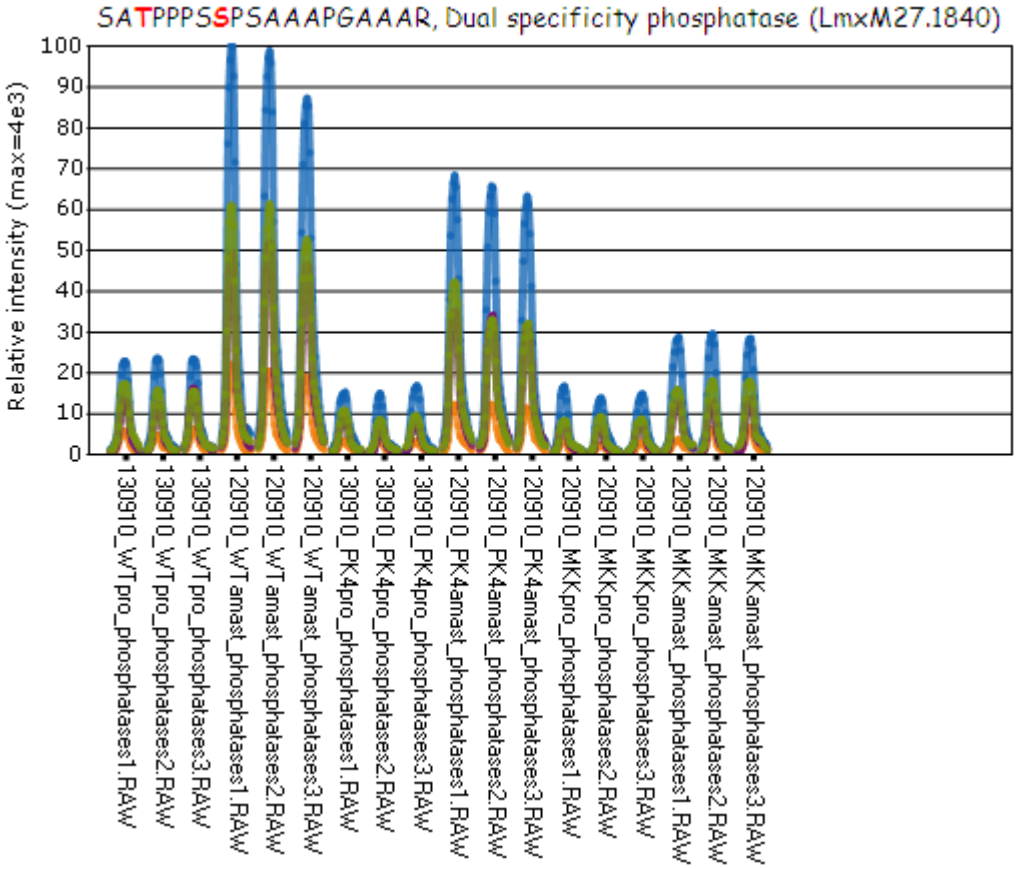


Figure 3.18: SRM phosphopeptide pattern between life stages. A screenshot from Pinpoint (see section **1.6.4**) displaying the intensities of different transitions (differently coloured peaks) in one of the targeted protein phosphatases. The pattern of phosphopeptide levels detected in wild type (WT) promastigotes and amastigotes would typically be similar in the two kinase deletion mutants. If the wild type parasites showed a 1:4 promastigote-amastigote phosphopeptide level ratio, as in this figure, similar ratios would typically be seen for the mutants. Often the wild type would display the highest levels, followed by ΔPK4 and ΔMKK. This is also the case for the present peptide, although here the promastigote-amastigote phosphopeptide ratio for the ΔMKK parasites is only about half as high as those of WT and ΔPK4.

3.6.3 MS^E

MS^E analyses were carried out with a dual purpose of 1) improving database annotation by obtaining protein information from *L. major* identifications sharing peptides with *L. mexicana*; and 2) acquiring general quantitative information about proteins in promastigotes and amastigotes. Hence, whole cell lysate of *L. mexicana* promastigotes and axenic amastigotes as well as *L. major* promastigotes were subjected to in-solution digestion with trypsin and aliquoted based on theoretical protein concentrations. Qubit protein concentration measurements were carried out for comparison, but not used for aliquoting as by experience, the protein concentrations provided by Qubit were notoriously off, regardless the measurements being carried out before or after protein digest. Samples were handed over to Richard Sprenger, post doctoral fellow in the Protein Research Group, SDU, who conducted the one-dimensional LC-MS/MS analyses and subsequent processing. Unfortunately, the size of the dataset, due to the exhaustive MS/MS analyses, and limited computer power, made data processing and searching against the 6-frame translation library database impossible. Instead, the data were searched against a concatenated Swiss-Prot database containing the available *L. major* and *L. mexicana* protein sequences as well as the *L. mexicana* predicted protein database, yielding slightly different results (**Table 3.11**). Expectedly, use of the *L. mexicana* predicted protein database provided the highest number of unique protein identifications in the *L. mexicana* samples. Of the 463 proteins detected by the predicted protein database in *L. mexicana* promastigotes and/or amastigotes, 141 were only detected in promastigotes and 89 only in amastigotes; another 26 proteins being more than 2-fold more abundant in amastigotes (**Appendix H**, ratio + sorted tabs, http://dl.dropbox.com/u/3011619/Appendix%20H.xlsx). The most interesting, and somewhat surprising, finding in the MS^E dataset was the significant down-regulation or even absence of ribosomal proteins in the amastigote life stage. Looking at the data from the concatenated database, the presence and levels of ribosomal proteins in the two life stages displayed like this:

- 36 of 107 proteins not detected in amastigotes were ribosomal proteins
- Ribosomal proteins constituted 19 of 30 proteins less abundant in amastigotes (1.52-3.55-fold down-regulation)
- Only 5 ribosomal proteins seemed to be present at relatively equal abundance in promastigotes and amastigotes (amastigote/promastigote-ratios of -1.44 – 1.47)
- 1 ribosomal protein was only detected in amastigotes

The potential biological implication of these finding are discussed in section **4.4.2**.

Table 3.11 – Legend on page 167.

	SwissProt+Trembl concatenated database				*L. mexicana* ORF database (predicted proteins)			
	***L. major* promast**	***L. mex* promast**	***L. mex* amast**	**Total unique proteins**	***L. major* promast**	***L. mex* promast**	***L. mex* amast**	**Total unique proteins**
Total protein number[3]	346	241	194	446	242	372	320	502
Total peptide number	3328	2319	2002		2277	3952	3452	
Average coverage (CV)	27.2% (32%)	26.0% (33%)	23.7% (31%)		24.4% (35%)	27.1% (33%)	25.1% (35%)	
Average peptides/protein (CV)	9.8 (27%)	9.7 (27%)	10.6 (28%)		9.6 (25%)	10.7 (25%)	11.0 (27%)	
Total amount on column, ng (CV)	646 (21%)	659 (24%)	488 (21%)		535 (20%)	946 (23%)	646 (24%)	
Calculated protein amount (nmol/10^9 cells)	96.6	101.9	39.5		75.9	135.4	50.5	

[3] Counted if detected in at least 2 replicates, with 2 or more unique peptides per protein

	SwissProt+Trembl concatenated database				*L. mexicana* ORF database (predicted proteins)			
	***L. major* promast**	***L. mex* promast**	***L. mex* amast**	**Total unique proteins**	***L. major* promast**	***L. mex* promast**	***L. mex* amast**	**Total unique proteins**
Calculated protein amount (µg/10^9 cells)	4013	4079	1733		3321	5852	2295	
Estimated concentration	2500 µg/ml	4500 µg/ml	1200 µg/ml		2500 µg/ml	4500 µg/ml	1200 µg/ml	
Measured MS^E	2675 µg/ml	4662 µg/ml	1980 µg/ml		2470 µg/ml	7210 µg/ml	2940 µg/ml	
Measured Qubit	474 µg/ml	598 µg/ml	496 µg/ml		474 µg/ml	598 µg/ml	496 µg/ml	

Table 3.11: MS^E results by use of different databases. The table lists the number of proteins and peptides identified with either the concatenated SwissProt-Trembl database mainly consisting of *L. major* entries, or the *L. mexicana* predicted proteins database. The protein amounts, as well as measured concentrations, are based on the identifications, and thus differ between the databases for the different sample types. The estimated concentrations are based on the notion that 10^6 promastigotes should contain 4 µg protein and 10^6 amastigotes 1 µg protein, relative to the number of cells in the aliquots from which the samples derived. CV, coefficient of variance.

Since the MS^E analyses were conducted on un-enriched samples, only 5 protein kinases and phosphatases were identified, 3 kinases and 2 phosphatases. Of these, only a single protein kinase and a single protein phosphatase were detected in both life stages, but simulated protein ratios were calculated to account for the fact that no detection does not necessarily mean that the protein is completely absent. These simulated ratios, as well as the standard ratios, when present, are shown in **Table 3.12**.

Lmex accession	Protein	MS^E standard, relative ratio amast/promast	Simulated ratio, fold change amast/promast
LmxM25.0750	Protein phosphatase, putative (Contig_0002051_164 [9379-10656])	→1.22	→1.22
LmxM05.0100	Phosphoprotein phosphatase, putative (Contig_0000670_257 [15840-17990])	∞	↑12.77
LmxM08_29.2570	Protein kinase, putative (Contig_0001468_413 [21354-19741] Reverse sense)	Not detected in amastigotes	↓5.90
LmxM18.0270	GSK3β (Contig_0001383_347 [24722-26119])	→0.78	↓1.28
LmxM25.2340	Serine/threonine protein kinase, putative (PKc-like superfamily) (Contig_0002077_114 [7839-9200])	∞	↑8.16

Table 3.12: Protein kinases and phosphatases detected in the MS^E analyses. The MS^E standard, relative ratio is calculated from the measured amounts of protein in the amastigote sample triplicate relative to the promastigote sample triplicate. The simulated ratio fold change is an extrapolated ratio calculated from the amounts measured in one or both life stages.

CHAPTER 4: DISCUSSION

This chapter will sum up the results and matters affecting their interpretation, as well as try to place them in a biological context. The latter part is not as easy as it may sound, so speculations have also made their way into some of these sections. To conclude the chapter, a view on future perspectives and research within this area is given.

4.1 *Leishmania* proteomics and phosphoproteomics

More phosphoproteomics analyses of trypanosomatids have been published [134, 174, 175, 195] within the past couple of years, but the overall results are still far from comparable with what has been achieved in studies of well-characterised organisms [74, 83, 196]. In the current study, an effective pipeline for qualitative and quantitative proteome and phosphoproteome analyses was established in *Leishmania mexicana*. The pipeline was based on established methods for cell handling, harvest and lysis as well as protein extraction and enrichment procedures, but each level was evaluated to find the best combination for effective and reliable LC-MS/MS results. *Leishmania* proteomics and phosphoproteomics are impacted by the relative lack of suitable protein databases for all but a few species (see section **4.2** for further discussion of this issue). The continued work on the genomics front will aid as will proteomics with contributions to validate predicted proteins [176], characterise expression profiles [100, 194, 197] and post-translational modifications [100, 134, 174, 175] as well as provide proof of inter-species differences that may not be picked up by the traditional homology-based sequence predictions [188]. The analytical pipeline presented in the current study allows for all these. More than 2,000 proteins have been identified (sections **3.4**) and 463 quantified (section **3.6.3**), with thousands of indicative (section **3.5.1**) or validated phosphorylation sites (section **3.5.2**). Focus on specific protein families led to identification of 107 protein kinases and 36 protein phosphatases (section **3.5.3**) with comprehensive phosphorylation site characterisation and quantification (sections

3.6.1.2-**4**, **3.6.2.1**, and **3.6.2.3.1**-**2**). The phosphoproteomics results pointed to *Leishmania*-specific sequence differences (section **3.5.4** and [188]), which may help improve the *L. mexicana* protein databases. The established pipeline can stand further modifications and optimisation, e.g. inclusion of further fractionation steps prior to MS-analysis. This also allows for its customisation for other *Leishmania* (phospho-) proteomics research projects with different sample perspectives as well as instrumental and software availability.

4.1.1 Approaching the Leishmania phosphoproteome?

Especially in the phosphoproteomics area, the pipeline promises significantly improved output compared to previous studies, e.g. [134, 174, 195]. A-scoring validated phosphorylations in 879 contig sequences. The overlap in contigs with validated phosphorylation sites from the 2 membrane and 11 WCL datasets evaluated by A-score was 105, i.e., 695 different contig sequences in total. This likely does not correspond to 695 different proteins as some protein sequences are covered by more than one contig. If we assume that 1 in 20 of these contigs corresponds to the same protein[4] – without taking isoforms or gene copies of the same protein into account – this would give 660 different phosphoproteins. Separately, 97 protein kinases and 25 protein phosphatases were found in a phosphorylated state. The overlap between these and the A-score validated phosphoproteins is 41 to be subtracted from the phosphoprotein total. This results in 741 presumably different phosphoproteins identified and validated in this study. Would this then correspond to the *Leishmania mexicana* phosphoproteome? No, certainly not, it will only be part of the phosphoproteome of this species. The current study identified more than 1,300 potential phosphoproteins from 15 datasets, none of

[4] In the list of protein kinases and phosphatases, 17 of 143 proteins (11.8%) are covered by 2 contigs. However, only a single of the 41 kinases/phosphatases also found in the A-score datasets was represented by both contigs. Hence, even if more than 1 in 20 proteins are covered by more than a single contig, both contigs need not always be found containing a phosphorylated peptide

which were among those evaluated by A-score (section **3.5.2**). Thus, many more phosphoproteins are likely to be added to the list of validated phosphoproteins in *Leishmania mexicana*. Publications within phosphoproteomics typically contain a lot of numbers: How many phosphoproteins/phosphopeptides were identified, how many unique phosphorylations sites, etc. Occasionally, publications display "the phosphoproteome" of this species or that organelle. This was recently the case in another trypanosomatid, namely the bloodstream form of *Trypanosoma brucei* [195]. While it may be tempting to use the term "phosphoproteome" when a certain higher number of phosphoproteins and phosphorylation sites have been identified, it is technically not correct as in reality nobody knows the exact size nor composition of an entire phosphoproteome in any organism. Due to the very dynamic nature of protein phosphorylation as well as the limitations of mass spectrometry (e.g. sensitivity, dynamic range, etc.) and sampling (e.g. biological origin of samples, enrichment procedures, etc.) what phosphoproteome research studies will provide is rather a glimpse of phosphorylated proteins at a given time point and/or state.

4.1.2 Phosphopeptide enrichment

Two different approaches for phosphopeptide enrichment prior to LC-MS/MS analysis were evaluated for use in the phosphoprotomics pipeline. TiO_2 chromatography is frequently used in phosphoproteomics studies, and the procedure has been evaluated, improved and modified several times [11, 74, 75, 79, 80]. The procedure was also modified to fit into the current study set-up (section **3.1.2.1**), and generally proved very effective with a decent degree of reproducibility (section **3.6.1.1**). In combination with IMAC, TiO_2 chromatography is also part of the SIMAC procedure [12], which should improve the yield of multiply phosphorylated peptides compared to TiO_2 alone. The performance of SIMAC and TiO_2 was also evaluated in the current study, in an attempt to further improve the detection of multiply phosphorylated peptides. The comparison of phosphopeptide enrichment from 3 × 1,000 µg protein digests by SIMAC and TiO_2

chromatography, respectively, proved that while SIMAC could indeed lead to identification of more phosphopeptides, it was no easy feat. The gain of multiply phosphorylated peptides by SIMAC proved insignificant, presumably due to sub-optimal elution-timing and conditions. The first elution step in the SIMAC procedure (section **2.2.7**) serves to elute the remaining singly phosphorylated or un-phosphorylated peptides, but if the elution is carried out too fast, a lot of these peptides will remain on the column. On the other hand, if the elution is dragged on too long, the multiphosphorylated peptides will also start to desorb, and thus be lost into the fraction of otherwise singly phosphorylated peptides. The very delicate balance of achieving the "perfect" conditions for SIMAC is a major drawback and the cause of discontinuation of SIMAC use for phosphopeptide enrichment in the current study. Even on its own, TiO_2 chromatography proved a very effective phosphopeptide enrichment measure, though this does not preclude other enrichment and/or sample fractionation procedures (section **4.1.3**) to be able to contribute and improve the overall output of phosphoproteomics in *Leishmania*.

4.1.3 Sample fractionation

In most phosphoproteomics studies, the current one included, it is of priority to organise the experiments in a way that enables high yields. Still, the overall experimental as well as biological conditions of the study need considerations and possible compromises. State-of-the-art phosphoproteomics typically exploits the forces of one or more pre-fractionation steps as well as enrichment procedures. The advantages of this strategy are that a complex sample can be split into multiple less complex fractions to allow for detection of more low-abundance proteins and tailoring of the mass spectrometric analyses. Apart from sample material and sophisticated preparation strategies, successful large-scale phosphoproteomics studies also depend on high quality protein databases. Regardless of noble experimental efforts, a poor protein database will impair the phosphoproteomics outcome. In the current study, the available databases have been

among the main issues. In the beginning, it was the lack of a species-specific, curated database, then interpretation of identified species-specific sequences. Because of these issues, it was reasoned that it would be more sensible to focus on fewer, high-quality, annotated phosphoprotein/-peptide identifications than aiming for achieving several thousand identifications only revealing anonymous contig-identity. Hence, the current study can be seen as a surface scanning of the *Leishmania mexicana* phosphoproteome. Trawling or deep-water fishing by application of phosphopeptide-specific pre-fractionation steps such as HILIC [198] and/or SAX/SCX [65, 199-201] can be embarked upon with improvement of the database(s) (both sequence- and annotation-wise). Such pre-fractionation steps, possibly in combination with 2D-LC-MS/MS analyses, should greatly increase the number of phosphopeptide/-protein identifications in *Leishmania mexicana* as even more low-abundant phosphopeptides may be subjected to fragmentation for subsequent identification. Yet, even aiming for less detailed analyses, it was still a priority to gain as much information on phosphorylations in protein kinases and phophastases as possible. In pursue of this, membrane protein extraction was undertaken to generate fractions of membrane proteins, soluble cytosolic proteins, and nuclear and mitochondrial proteins. This did indeed lead to identification of supplemental protein kinases and phosphatases, though not at any great extent. Still, valuable information could be retrieved from the membrane protein analyses, as numerous membrane transporters invisible in the WCL analyses were identified, some highly phosphorylated (section **3.5.4**). The importance of these findings is still unclear, especially as coupling to specific protein kinases remains to be characterised, but they do indicate that membrane proteins of *Leishmania mexicana* are regulated by phosphorylation and thus may well be part of signalling networks.

4.1.4 Phosphorylation motif analysis

Identification of sequence motifs surrounding phosphorylated residues is commonly undertaken in phosphoproteomics studies [60, 202, 203]. This is because it may assist in placing the identified phosphoproteins in the context of relevant interaction partners,

including the kinases responsible for the phosphorylations, thus contributing to confirmation or dismissal of the proposed hypotheses. In organisms where the signalling network of interest is poorly characterised, like in *Leishmania*, motif analyses may be the first step towards identifying kinase-substrate relations. Motif-X [39] was tested in the current study, but its use was found greatly impaired by the fact that the sequences under investigation were not derived from any typical organism/database. Other tools, e.g. WebLogo [39] should be tested for general motif analysis, but so far analyses have been confined to comparison between different *Leishmania* species, which in itself has proved valuable for evaluation of the degree of phosphorylation profile variation across species. **Appendix I** (http://dl.dropbox.com/u/3011619/Appendix%20I) shows the variation of the protein sequences among the different *Leishmania* species where phosphorylation sites have been identified in *L. mexicana*. Only phosphopeptides derived from protein kinases that have been identified from wild type parasites are shown. As can be seen from this appendix, only 30% of the phosphorylation sites have a common sequence motif between the different species, obviously most predominant for the MAP kinases where phosphorylations were frequently observed in the TXY-domain. Thus, trying to identify *L. mexicana*, *L. infantum*, or *L. braziliensis* proteins and their phosphorylation sites solely from the *L. major* database will only work in about 30% of the cases.

4.2 Database implications

Proteomics studies of any kind are highly dependent on suitable databases in order to achieve the best possible evaluation of the acquired MS-data. As has already been mentioned repeatedly, the available *Leishmania* protein databases are of very varying quality. Unless being keen on *de-novo* sequencing from MS/MS spectra [204, 205], researchers may resolve to generate customised databases [206], and apply them in conjunction with databases from related species if available [207]. This latter approach is similar to what was applied in the current study. While the use of the recently generated 6-frame translation library and predicted protein list for *Leishmania mexicana*

posed certain challenges, the end results should pay off. This is perfectly illustrated by the phosphopeptides identified for *L. mexicana* in the current study. A comparison of a number of these peptide sequences to their relatives in *L. major*, *L. infantum*, and *L. braziliensis* showed that if the 6-frame translation library or predicted *L. mexicana* protein list had not been available, almost 70% of the phosphopeptide and phosphorylation site identifications would not have been possible from the *L. major* or other related *Leishmania* protein databases (section **4.1.4**). This very much underscores the importance of having species-specific databases available, and partly explains the success rate of the proteomics and phosphoproteomics analyses in the current study. If no species-specific databases had been available, the protein identifications of the current study would only have been around 660 (i.e. 30% of the 2201 proteins mentioned in section **3.4**), and with 60% of those being phosphorylated this would correspond to 395 phosphoproteins. These numbers are much more in line with another recent phosphoproteomics study identifying 445 phosphoproteins in *L. donovani* [174].

4.2.1 False discovery rates

In proteomics studies, specification of false discovery rates (FDR) for protein and/or PTM identifications is expected – and with good reason. The purpose of FDR listing is to prove the validity and trustworthiness of the presented results. This is a necessity due to the numerous factors influencing the results – from the quality of the acquired mass spectrometric data over the choice of database and search criteria to the base of results' acceptance. Depending on the experimental setup and criteria for acceptance of database search results, it is very easy to present thousands of identifications and quantitations, even in phosphoproteomics studies [202]. Due to the proportions of typical large-scale proteomics studies, the output results are often processed at least semi-automatically via different software applications. These applications often allow one to filter the results on one or more different levels, e.g. peptide score, peptide/protein probability, FDR, etc. FDR thresholds are often used, but the reliability of these depends largely on how they have been calculated, which may well differ from one application to another. If the data

have been searched against a decoy database – often a reversed version of the "real" database – the number of identifications with this database is often compared to the identifications with the real database to get a measure of the FDR. This may be a very robust way of assessing FDRs if the choice of the true database is well considered and mainly consists of validated entries. In the current study, where the main database used was a 6-frame translation library, this approach however is futile as the database already consists of up to 50% non-sense sequences. As a consequence, no false discovery rates are reported for the identifications presented in this thesis. For the analyses performed in Scaffold, false discovery rates were stated, but not included in this thesis, the reason for this being elaborated here. Scaffold operates with two different approaches for FDR calculations, an empirical one when a decoy database has been applied, and a probabilistic method when no decoy database results are available. The FDRs available for the *L. mexicana* data analysed by Scaffold were based on the latter method where the protein identification probabilities in Scaffold are summed and divided by the highest possible probability (i.e. 100%) for each identification [208]. Thus, the results of the FDR calculations are largely dependent on the settings chosen for protein and peptide probability filtering. Yet, even with relatively strict settings for these filters, there may be flaws affecting the FDR calculations. For whatever reason, it appears that some proteins will be listed with 100% identification probability in the results overview, even if no peptides are listed for these proteins. **Figure 4.1** shows a screenshot from the Scaffold analysis of the membrane protein fractions where the protein has 100% identification probability scores in all samples even though no peptides apparently have passed the set criteria in 5 of 6 samples.

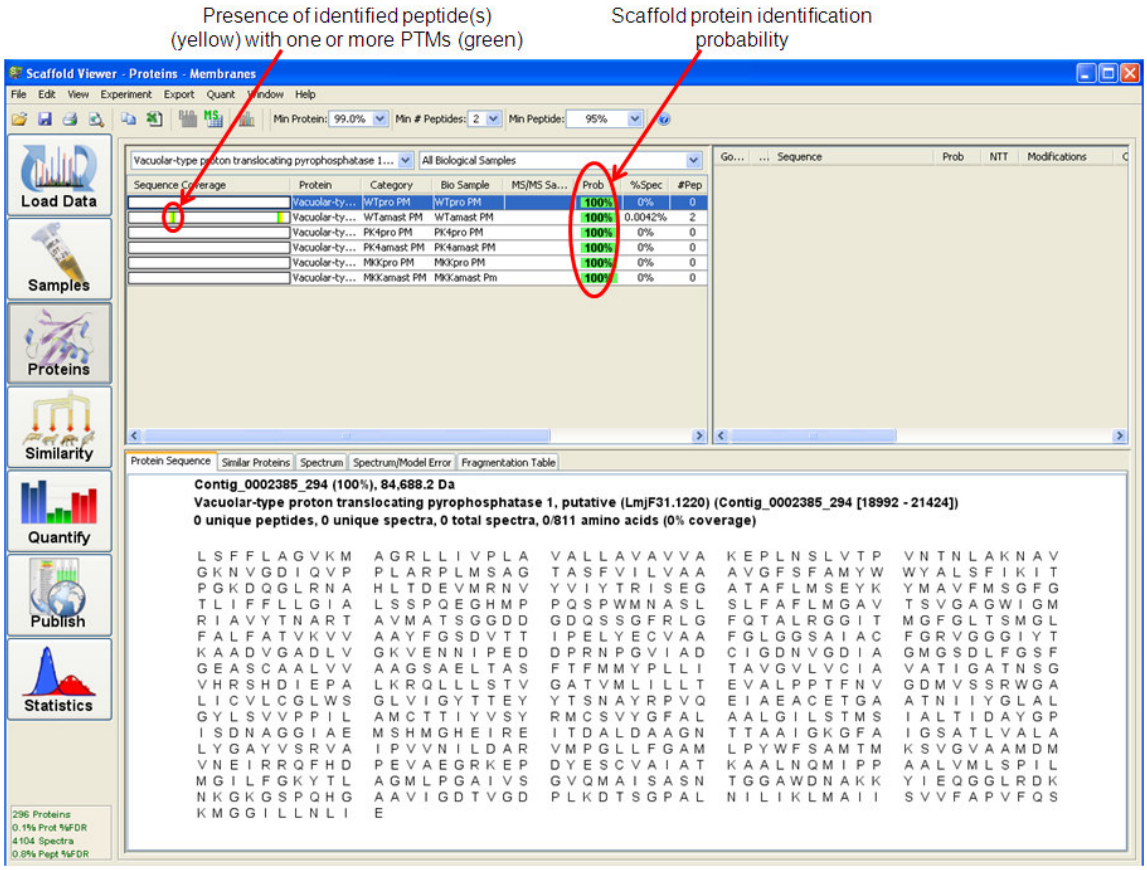


Figure 4.1: Screen shot from Scaffold displaying an apparent bias in the probabilistic method for FDR calculation. In this example, the protein has 100% identification probability assigned for all 6 samples, even though peptides are only identified in one of the samples (WTamast PM marked by yellow and green colours in the "sequence coverage" column).

The lack of FDRs in the current study may be regarded as a shortcoming, and could have been circumvented to a certain degree by using the predicted *L. mexicana* protein database which would allow for generation of a decoy search. This was not persued though, as it would still entail a certain bias because not all peptides would be identifiable from this database. True FDR evaluation for *L. mexicana* proteomics results would still benefit from improved databases.

4.3 Quantitative analyses

Three different types of quantitative analyses were conducted, assessing differences between life stages as well as between wild type and kinase-deletion mutant parasites. They supplemented and supported each other in different ways, but also displayed certain limitations. The method-specific advantages and limitations will be discussed individually in the following sections, while the general ones will be summed up and evaluated in section **4.3.4**.

4.3.1 iTRAQ

iTRAQ analyses were carried out on the clear cytosolic protein fractions of WT, ΔPK4 and ΔMKK promastigotes and axenic amastigotes as well as lesion-derived amastigotes' cytosolic proteins. The number of identifications was generally lower than in the non-quantitative analyses. The initial perspective of the analyses was to expand the number of identifications of phosphorylated peptides by starting from less complex samples, while simultaneously obtaining quantitative information to assess life stage as well as wild type-kinase-deletion mutant-specific differences. The number of new identifications made by iTRAQ was minimal to non-existing, but the quantitative analyses indicated some surprising facts. The overall perception, established from the discovery experiments prior to both iTRAQ and SRM, was that the MKK knock-out mutant would show less pronounced phosphorylation. This did not always seem to be the case, when judging by the iTRAQ results. In the amastigote clear cytosol iTRAQ

analysis, the ΔMKK samples would generally have the same number of identified phosphopeptides as the WT, while ΔPK4 would miss out on a few of them. The pattern of peptide levels being WT > ΔPK4 > ΔMKK, as observed with the kinases and phosphatatases SRM analyses, did not display as clear in the iTRAQ analyses either.

4.3.1.1 Choice of reference

For proper use of iTRAQ-derived results it is important to have an appropriate reference to compare the changes of the different samples against. A perfect example would be the choice of t=0 in a time-course experiment, tracking changes over time. In the current study, iTRAQ was applied to compare different sample types in different life stages, making the choice of a reference sample less easy. The choice fell on WTpro WCL as reference in all experiments, mainly because it was easily available, it would allow results comparison across the different iTRAQ experiments if necessary and because previous experiments had shown relatively consistent results of phosphopeptide enrichment. It was, however, a choice with certain reservations, as it did not seem like the most appropriate choice for amastigote-specific analyses, since anticipated life stage-specific differences could mask changes between the samples in question simply because ratios can not be calculated if the phosphopeptide is not present in the reference life stage. To get a measure of the effect of a promastigote sample reference compared to an amastigote one for amastigote-specific experiments, ratios were also calculated with the WTamast clear cytosol sample as the reference. The statistical analyses clearly proved that this made a difference. An example of this difference can be seen in **Figure 4.2** below.

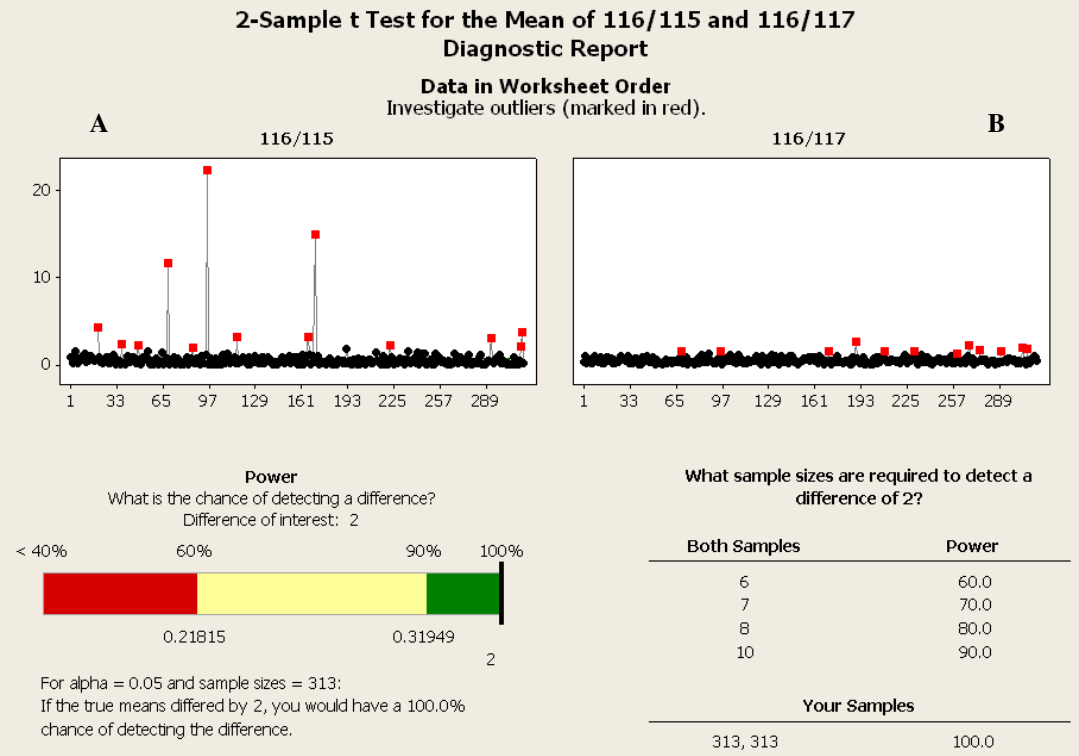


Figure 4.2: The effect of WTpro WCL or WTamast clear cyt as iTRAQ reference sample in axenic amastigotes-lesion-derived amastigotes comparison. Chart **A** and **B** clearly shows that the distribution of data points (ratios) differs when WTpro WCL is chosen as reference (**A**) instead of WTamast clear cytosol (**B**). When WTpro WCL is chosen as reference, the number of outliers (red points) as well as their differences to the general population of data is greatly increased.

4.3.2 SRM

In contrast to iTRAQ, SRM proved extremely useful in the characterisation of life stage as well as wild type versus kinase-deletion mutant differences in protein kinase and protein phosphatase phosphorylations. Relative quantification of 228 phosphopeptides from protein kinases (149) and protein phosphatases (79) was achieved, indicating anticipated as well as unexpected differences between the different types of samples analysed. These will be discussed in the sections below. While the SRM analyses proved very efficient in quantitative analyses of protein kinases and phosphatases, the quantitation is still relative. A great improvement would be to include synthetic, heavy isotope labelled peptides of sequences identical to those "real" ones being scrutinised. This would not only improve confidence in choosing the correct retention time in cases where there seem to be more opportunities, it would also allow absolute quantitation of the analysed phosphopeptides. Combined with analyses of the total protein concentration, this could add yet another level of information to the individual phosphorylation sites. E.g. assessing the importance of a given phosphorylation by comparing the fold-change of the phosphopeptide to the fold-change of the protein from which the given peptide is derived.

4.4 Life stage-dependent variance

The discovery as well as subsequent quantitative analyses indicated some significant differences between the two major life stages of *Leishmania mexicana*. Differences specific for the amastigote life stage may be of importance for future investigation of potential drug targets. Hence, focus in this discussion will be on these differences. Amastigote-specific differences were detected for both kinase and phosphatases (SRM analyses), and at the general protein level (primarily MS^E).

4.4.1 Amastigote-specific kinase phosphorylation

Amastigote-specific differences related to signalling are of high interest in the quest for better understanding of parasite life stage regulation as well as for potential future drug targets. At the protein level, such differences can be qualitative as well as quantitative, both generally and PTM-wise. A previous phosphoproteomics study in *Leishmania*, looking at both promastigotes and amastigotes by a combination of 2D electrophoresis and mass spectrometry, found MPK10 and nucleoside diphosphate kinase b to mainly be phosphorylated in the amastigote life-stage [134]. The current study identified phosphorylation sites in both these kinases in both life stages. Especially for MPK10, the statement of phosphorylation predominantly being an amastigote-specific event seemed a bit dubious as 4 different phosphopeptides were identified in promastigotes, and only one of these also in amastigotes (**Table 3.7**). To further investigate the presence of these kinases, they were included in different SRM set-ups. Nucleoside diphosphate kinase b was analysed in the initial SRM test, while MPK10 was involved in both this initial SRM test, but also extensively evaluated in an MPK10-specific SRM analysis as well as the selected kinases SRM analysis. The different MPK10 phosphopeptides were repeatedly shown to be present in both life stages, some more significantly in promastigotes, others in amastigotes. An estimation of the total phosphorylated MPK10 amount in the two life stages, as calculated by Pinpoint (see section **1.6.4**) and based on the targeted phosphopeptides, indicates fairly equal levels. In the MPK10-specific and selected kinases SRM analyses, the MPK10 phosphoprotein ratio was 1.0:7.46E-1 – 1.0:8.5E-1 (promastigotes:amastigotes). In the initial SRM test study, only the THpSLMELAGSVPAPS peptide in two versions (+/- methionine oxidation) was included, thus making this ratio of 1.0:5.1E-1 very biased. The single phosphopeptide analysed for nucleoside diphosphate kinase b, showed a 6.2 times up-regulation in amastigotes.

MPK10 was not the only protein kinase to show differential phosphorylation positions and amounts between the two life stages. Phosphopeptides of 8 MAP kinases were shown to be 2-fold more abundant in WTpro than in WTamasts, whereas 7 showed more than 2-fold less abundance. Among those MAP kinases with phosphopeptides being more abundant in amastigotes were MKK and MPK3, both of which are thought to be significantly down-regulated (protein level) in amastigotes [44, 179]. This could indicate that the phosphorylations identified as up-regulated in these proteins in amastigotes are significantly more important as it could be interpreted as a greater percentage of the MPK3 and MKK pools are phosphorylated. The fact that both are found regulated in the same way for at least some of their phosphopeptides, also supports their interactions as were described in [44].

4.4.2 The ribosomal proteins' effect

The MS^E experiments indicated that ribosomal proteins are very differently regulated between the two life stages. The vast majority of the identified ribosomal proteins were either significantly less abundant, or completely absent, in axenic amastigotes. The exact reason for this is unknown, but one can speculate that different types of ribosomal proteins are essential in the different life stages. This would of course have to be investigated by further experiments, e.g. trying to quantitatively map the distribution of the different ribosomal proteins between the life stages. It is also possible that the fact that the analyses were done on axenic amastigotes could have biased the results slightly, but that is not to know until this question has been assessed by an experiment including lesion-derived as well as axenic amastigotes. Should it turn out to be a true and significant difference, it might be exploited towards the identification of new potential drug targets, whether directly approaching the ribosomal proteins, or affecting processes that regulate these proteins. Several studies have already shown an effect of *Leishmania* ribosomal proteins as potential vaccine agents [209-212].

4.5 WT vs. kinase deletion-mutants

Kinase deletion-mutants were included to investigate the effects of protein phosphorylation compared to the wild type state. While the mutants were viable, they were clearly not unaffected by the lack of potentially important protein kinases. Especially the ΔMKK mutants seem affected with a significantly changed morphology as well as impaired proliferation rate. When knock-out of presumably important higher order protein kinases is not lethal, it can have different explanations. The easiest is obviously that the given protein kinase is not essential, but it is also possible that the viability of the deletion-mutants attests to the parasites being extremely clever survivors, capable of substituting important kinase function just enough. It was expected to see somewhat lower degrees of protein phosphorylation in the deletion-mutants, which was also generally confirmed. Overall, MAP kinase phosphorylation seemed decreased in MKK deletion mutants, which fits well with the notion that MKK is responsible for phosphorylation of one or more down-stream MAP kinases [44]. On the other hand, some phosphopeptides seemed more abundant in the MKK deletion mutant, indicating that MKK may not only serve to activate down-stream kinases, but also exerts some kind of inhibitory regulation. The relationships between MKK and the kinases showing more or less abundant phosphorylations in the absence of MKK need to be evaluated by further analyses, including sequence motif assessments.

The other deletion mutant, ΔPK4, did not appear to have phosphorylation levels as severely affected as ΔMKK. Still, a number of MAP kinases as well as MKK phosphopeptides showed differential regulation in amastigotes when compared to the wild type. The doubly phosphorylated version of the MKK SQESLENDVK-peptide was significantly less abundant in the ΔPK4 amastigotes, whereas the singly phosphorylated version of the peptide displayed just above 2-fold increase in abundance. Thus, these kinase kinases may share some substrates as well as up-stream kinases, which can then

account for differential regulation of phosphorylation of either MAP kinase kinase once the other is missing, perhaps to try to make up for the missing kinase. Another MAP kinase showing significantly lower abundance in ΔMKK compared to ΔPK4 and wild type was MPK3. MPK3, by interaction with MKK, was recently shown to be involved in regulation of flagellar length in promastigotes [44], and so it makes perfectly sense to see down-regulation of MPK3 phosphorylation in the MKK knock-out mutant, especially when the phenotype of the MKK deletion mutant also displays an extremely short flagellum.

One protein kinase that was detected in all the different quantitative experiments was GSK3β. The MS^E analysis indicated a fairly equal abundance of the protein between the two life stages, and the SRM analyses of the tyrosine phosphorylated peptide showed a similar distribution for the wild type as well as for the ΔPK4 deletion mutant. In the ΔMKK deletion mutant, however, GSK3β tyrosine phosphorylation was detected as significantly (i.e. more than 2-fold) less abundant in both promastigotes and amastigotes. This finding was similar for both the SRM and the iTRAQ analyses. It could indicate that ΔMKK and GSK3β are involved in the same pathway, if not directly then indirectly via other up- or down-stream kinases.

4.6 Axenic versus lesion-derived amastigotes

Axenic amastigotes, when available, currently serve as a model of amastigotes *in vivo*. The environment of axenic amastigotes, i.e. culture medium deprived of other cell types, however is very different from that of the *in vivo* parasites. To assess the significance of these differences and their importance in proteomics and phosphoproteomics studies, lesion-derived amastigotes were also included in the current study, although to a much lesser extent than their axenic peers. At the phosphoproteomics level, differences between these two types of parasites were apparent from the discovery experiments, encouraging quantitative analyses of lesion-derived amastigote proteins as well. The

quantitative analyses (iTRAQ and SRM), confirmed that differences were indeed present between these two types of parasites on the phosphorylation level. In most cases, the axenic amastigotes would show higher levels of phosphorylation than the lesion-derived amastigotes, although some phosphorylations were detected only in the latter type of parasites. Due to the relatively small number of lesion-derived amastigote samples assessed as well the quantitative analyses only being relative, it is difficult to evaluate just how significant the differences between axenic and lesion-derived amastigotes are. However, the findings confirm that axenic amastigotes are indeed just a model system, and findings will need assessment in *in vivo* parasites as well.

4.7 Future perspectives

Nothing is more typical, yet encouraging research-wise, than to discover ten new questions for every single one you manage to answer. *Leishmania* proteomics and phosphoproteomics, though gaining popularity fast, are still at a developmental stage, not least due to species-specific differences and the relative lack of proper databases. The current study has provided some improvement by establishing a stable proteomics/phosphoproteomics pipeline, providing the first library of *Leishmania mexicana* protein kinase and protein phosphatase phosphorylation sites, and discovering important sequence variations. There is still a lot more to gain from the acquired data, though. Extensive protein database fitting is still a critical necessity. The acquired data literally hold ten thousands of general *Leishmania mexicana* phosphorylation sites just waiting for evaluation. These sites as well as all those already validated should be examined for new as well as conserved phosphorylation site motifs, which in itself may require significant manual effort unless the available databases and motif tools are improved to fit alternatively annotated sequences. The benefits of extensive motif analysis can be further extended when additional analyses manage to characterise more of the numerous hypothetical proteins currently present in the *Leishmania* protein repositories.

The results of future motifs analyses should be evaluated in the light of qualitative as well as quantitative phosphoproteomics findings between the wild type parasites and the kinase deletion mutants, as this may suggest certain kinase-substrate relations. Extensive comparative sequence analyses to identify conserved phosphorylated sequences between different *Leishmania* species would allow prioritising of the identified phosphoproteins/peptides, as those sequences conserved between the species are likely to be of significance to certain parasite tasks. The conserved phosphopeptides could then be subjected to targeted quantitative analyses to identify any life stage-specific differences. Phosphopeptides displaying significant up-regulation in amastigotes will be the most interesting ones in terms of identifying new potential drug targets, especially if quantitation of the proteins from which these peptides derive, support the suggested regulation. Artificial peptides of the interesting sequences could then be synthesized for screening against potential kinases (kinase assays). Even more sophisticated, Breitkreutz *et al.* [213] recently demonstrated how this can be combined with other approaches to map interaction networks in yeast. By a combination of magnetic bead capture, on-bead protein digestion, and mass spectrometry, Breitkreutz *et al*. were able to characterise protein kinase and phosphatase complexes, and identify the complex components by epitope tags and expression systems [213].

The current study provides preliminary quantitative protein as well as phosphorylation results. Previously, quantitative proteomics/phosphoproteomics in *Leishmania* was confined to iTRAQ and 2D gel electrophoresis approaches. The current work has proven SRM and MS^E as additional valuable tools for quantitative characterisation of *Leishmania* proteins. These analyses should be further improved. The SRM-based phosphopeptide quantification will benefit from inclusion of heavy peptide standards, not only as positive controls, but more importantly to allow absolute quantification of the phosphopeptides in question. The procedure will also be valuable for future kinase-substrate analyses.

The MS^E analyses currently conductable in the PR group at SDU are not suited for modified proteins. While modified proteins, including phosphorylated ones, inevitably are analysed, the instrument-related software is liable of impaired registration of these proteins and peptides in the subsequent search results. Even if improvement of this may not be just around the corner, the *Leishmania* MS^E analyses can still be significantly advanced. Application of a 2D-LC-MS/MS set-up will significantly increase the number of identifications to be made, thus further improving not only the general knowledge of *Leishmania* proteins, but also the quantitative analyses. A completely standardised set-up of future MS^E experiments would also be desirable. In the current analyses, the samples from the different parasite probes (*L. mex.* Promastigotes and axenic amastigotes, and *L. major* promastigotes) were derived from widely different cell aliquots (5×10^8 - 1×10^9 cells/aliquot), and with the inherent difference in protein concentration between promastigotes and amastigotes (4:1), different amounts of trypsin had been applied in the preceding protein digestion procedure. In MS^E, quantification of trypsin autocleavage products can act as an internal quantitative standard. Thus, it would be advantageous to start out with identical amounts of protein, ensuring as homogeneous sample preparation as possible. Including lesion-derived amastigotes in future analyses would allow for a more general evaluation of the differences between these and axenic amastigotes, assisting possible improvement of the amastigote model system. Lesion-derived amastigotes analysed side-by-side with axenic amastigotes and promastigotes might further evaluate the indicated differences in ribosomal protein presence between the two life stages. This could provide additional information about the differentiation and adaptation of the parasites in the host (i.e. is the proposed down-regulation of ribosomal proteins an *in vitro* bias, do promastigotes and amastigotes employ different ribosomal proteins for similar tasks, how important is post-translational protein modification in the differentiation and adaptation processes, etc.).

CHAPTER 5: CONCLUSION

The current study led to the construction of a robust proteomics/phosphoproteomics pipeline for differential qualitative as well as quantitative analyses of *Leishmania* proteins. The pipeline was applied for general as well as specific proteomics and phosphoproteomics analysis in *Leishmania mexicana*. The general analyses resulted in the construction of a preliminary library of more than 2,000 *Leishmania mexicana* proteins, of which more than 60% are potential phosphoproteins. From a smaller subset of the acquired data, almost 2,000 phosphopeptides with more than 2,300 phosphorylation sites (not completely non-redundant) were validated by a semi-automatic approach. Specific analyses of protein kinases and protein phosphatases resulted in the identification of 107 different protein kinases and 36 different protein phosphatases, carrying a total of 424 non-redundant phosphorylation sites.

Simple in its set-up, the proteomics pipeline only involved a single sample pre-fractionation option, namely that of membrane versus cytosolic protein extraction. This leg of the pipeline was exploited for a few general proteomics/phosphoproteomics analyses, and otherwise reserved for phosphopeptide quantification by iTRAQ. The general phosphoproteomics analyses provided a list of 43 different membrane proteins with transporter functions, all being more or less heavily phosphorylated. Quantitative phosphopeptide analyses were conducted by iTRAQ on cytosolic proteins from wild type promastigotes, axenic amastigotes, and lesion-derived amastigotes, as well as PK4 and MKK kinase deletion mutants. In the wild type parasites, the results indicated some significant differences between the two life stages, axenic and lesion-derived amastigotes. Differences between the wild type and kinase deletion mutants were also abundant. Supplementary to iTRAQ, other quantitative analyses were carried out. MS^E analyses of wild type promastigotes and amastigotes as well as *L. major* promastigotes led to protein level quantification of 463 proteins as well as assisted in improving database annotation. Protein kinases and protein phosphatases were quantitatively evaluated by a targeted phosphoproteomics approach, which proved highly effective. Of

the variety of quantitative findings, the most intriguing were those of phosphorylation of specific MAP kinases being up-regulated in a life stage where the protein should be severely down-regulated, and the differences in ribosomal protein abundance between the life stages. For the kinase deletion mutants, the knock-out of MKK seemed to have the most significant effects, with extreme phosphorylation down-regulation of almost all of the identified MAP kinases. A few exceptions were noted, with phosphorylations being more abundant in the MKK mutant than the wild type and PK4 mutant, which could indicate that MKK also serves inhibitory purposes in the regulation of *Leishmania* signalling networks.

The current study is just the first of many steps towards mapping of the *Leishmania mexicana* proteome and phosphoproteome. Plenty of data derived from the study are awaiting further evaluation to expand the libraries of validated phosphopeptides and phosphorylation sites.

REFERENCES

1. LeishDOMUS LeishDOMUS. http://www.leishdomus.org/leishmaniasis.htm

2. Hotez, P. J.; Fenwick, A.; Savioli, L.; Molyneux, D. H., Rescuing the bottom billion through control of neglected tropical diseases. *Lancet* **2009,** 373, (9674), 1570-5.

3. WHO WHO | Leishmaniasis. http://www.who.int/leishmaniasis/en/

4. Ameen, M., Cutaneous leishmaniasis: advances in disease pathogenesis, diagnostics and therapeutics. *Clin Exp Dermatol* **2010,** 35, (7), 699-705.

5. Cobb, S. L.; Denny, P. W., Antimicrobial peptides for leishmaniasis. *Curr Opin Investig Drugs* **2010,** 11, (8), 868-75.

6. Croft, S. L.; Coombs, G. H., Leishmaniasis--current chemotherapy and recent advances in the search for novel drugs. *Trends Parasitol* **2003,** 19, (11), 502-8.

7. Naula, C.; Parsons, M.; Mottram, J. C., Protein kinases as drug targets in trypanosomes and Leishmania. *Biochim Biophys Acta* **2005,** 1754, (1-2), 151-9.

8. Sundar, S.; More, D. K.; Singh, M. K.; Singh, V. P.; Sharma, S.; Makharia, A.; Kumar, P. C.; Murray, H. W., Failure of pentavalent antimony in visceral leishmaniasis in India: report from the center of the Indian epidemic. *Clin Infect Dis* **2000,** 31, (4), 1104-7.

9. Xingi, E.; Smirlis, D.; Myrianthopoulos, V.; Magiatis, P.; Grant, K. M.; Meijer, L.; Mikros, E.; Skaltsounis, A. L.; Soteriadou, K., 6-Br-5methylindirubin-3'oxime (5-Me-6-BIO) targeting the leishmanial glycogen synthase kinase-3 (GSK-3) short form affects cell-cycle progression and induces apoptosis-like death: exploitation of GSK-3 for treating leishmaniasis. *Int J Parasitol* **2009,** 39, (12), 1289-303.

10. Tekirian, T. L.; Thomas, S. N.; Yang, A., Advancing signaling networks through proteomics. *Expert Rev Proteomics* **2007,** 4, (4), 573-83.

11. Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* **2005,** 4, (7), 873-86.

12. Thingholm, T. E.; Jensen, O. N.; Robinson, P. J.; Larsen, M. R., SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol Cell Proteomics* **2008,** 7, (4), 661-71.

13. Haile, S.; Papadopoulou, B., Developmental regulation of gene expression in trypanosomatid parasitic protozoa. *Curr Opin Microbiol* **2007,** 10, (6), 569-77.

14. Kamhawi, S., Phlebotomine sand flies and Leishmania parasites: friends or foes? *Trends Parasitol* **2006,** 22, (9), 439-45.

15. Bengs, F.; Scholz, A.; Kuhn, D.; Wiese, M., LmxMPK9, a mitogen-activated protein kinase homologue affects flagellar length in Leishmania mexicana. *Mol Microbiol* **2005,** 55, (5), 1606-15.

16. Wiese, M., A mitogen-activated protein (MAP) kinase homologue of Leishmania mexicana is essential for parasite survival in the infected host. *Embo J* **1998,** 17, (9), 2619-28.

17. Wang, Q.; Melzer, I. M.; Kruse, M.; Sander-Juelch, C.; Wiese, M., LmxMPK4, a mitogen-activated protein (MAP) kinase homologue essential for promastigotes and amastigotes of Leishmania mexicana. *Kinetoplastid Biol Dis* **2005,** 4, 6.

18. DPDx DPDx Center for Disease Control (CDC): Leishmaniasis. http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/Leishmaniasis_il.htm

19. Piscopo, T. V.; Mallia, A. C., Leishmaniasis. *Postgrad Med J* **2006,** 82, (972), 649-57.

20. Tripathi, P.; Singh, V.; Naik, S., Immune response to leishmania: paradox rather than paradigm. *FEMS Immunol Med Microbiol* **2007,** 51, (2), 229-42.

21. Handman-Laboratory The Walther and Eliza Hall Institute of Medical Research. http://www.wehi.edu.au/research/divisions/inf/labs/handman/leishmaniasis.html

22. Mahajan, V. K.; Sharma, N. L., Therapeutic options for cutaneous leishmaniasis. *J Dermatolog Treat* **2007,** 18, (2), 97-104.

23. Drummelsmith, J.; Brochu, V.; Girard, I.; Messier, N.; Ouellette, M., Proteome mapping of the protozoan parasite Leishmania and application to the study of drug targets and resistance mechanisms. *Mol Cell Proteomics* **2003,** 2, (3), 146-55.

24. Ouellette, M.; Drummelsmith, J.; Papadopoulou, B., Leishmaniasis: drugs in the clinic, resistance and new developments. *Drug Resist Updat* **2004,** 7, (4-5), 257-66.

25. Parsons, M.; Ruben, L., Pathways involved in environmental sensing in trypanosomatids. *Parasitol Today* **2000,** 16, (2), 56-62.

26. Wiese, M., Leishmania MAP kinases--familiar proteins in an unusual context. *Int J Parasitol* **2007,** 37, (10), 1053-62.

27. Campbell, D. A.; Thomas, S.; Sturm, N. R., Transcription in kinetoplastid protozoa: why be normal? *Microbes Infect* **2003,** 5, (13), 1231-40.

28. Clayton, C. E., Life without transcriptional control? From fly to man and back again. *Embo J* **2002,** 21, (8), 1881-8.

29. Stiles, J. K.; Hicock, P. I.; Shah, P. H.; Meade, J. C., Genomic organization, transcription, splicing and gene regulation in Leishmania. *Ann Trop Med Parasitol* **1999,** 93, (8), 781-807.

30. Parsons, M.; Worthey, E. A.; Ward, P. N.; Mottram, J. C., Comparative analysis of the kinomes of three pathogenic trypanosomatids: Leishmania major, Trypanosoma brucei and Trypanosoma cruzi. *BMC Genomics* **2005,** 6, 127.

31. Temporini, C.; Calleri, E.; Massolini, G.; Caccialanza, G., Integrated analytical strategies for the study of phosphorylation and glycosylation in proteins. *Mass Spectrom Rev* **2008,** 27, (3), 207-36.

32. Stryer, L., *Biochemistry*. 4 ed.; Freeman.

33. Dunn, D. A., Mining the human "kinome". *Drug Discov Today* **2002,** 7, (22), 1121-3.

34. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S., The protein kinase complement of the human genome. *Science* **2002,** 298, (5600), 1912-34.

35. Arena, S.; Benvenuti, S.; Bardelli, A., Genetic analysis of the kinome and phosphatome in cancer. *Cell Mol Life Sci* **2005,** 62, (18), 2092-9.

36. Wang, Z.; Shen, D.; Parsons, D. W.; Bardelli, A.; Sager, J.; Szabo, S.; Ptak, J.; Silliman, N.; Peters, B. A.; van der Heijden, M. S.; Parmigiani, G.; Yan, H.; Wang, T. L.; Riggins, G.; Powell, S. M.; Willson, J. K.; Markowitz, S.; Kinzler, K. W.; Vogelstein, B.; Velculescu, V. E., Mutational analysis of the tyrosine phosphatome in colorectal cancers. *Science* **2004,** 304, (5674), 1164-6.

37. Brenchley, R.; Tariq, H.; McElhinney, H.; Szoor, B.; Huxley-Jones, J.; Stevens, R.; Matthews, K.; Tabernero, L., The TriTryp phosphatome: analysis of the protein phosphatase catalytic domains. *BMC Genomics* **2007,** 8, 434.

38. Parsons, M.; Worthy, E. A.; Ward, P. N.; Mottram, J. C., Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics* **2005,** 6, 127.

39. 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.; Myler, P. J., The genome of the kinetoplastid parasite, Leishmania major. *Science* **2005,** 309, (5733), 436-42.

40. Chang, L.; Karin, M., Mammalian MAP kinase signalling cascades. *Nature* **2001,** 410, (6824), 37-40.

41. Hammarton, T. C., Cell cycle regulation in Trypanosoma brucei. *Mol Biochem Parasitol* **2007,** 153, (1), 1-8.

42. Kuhn, D.; Wiese, M., LmxPK4, a mitogen-activated protein kinase kinase homologue of *Leishmania mexicana* with a potential role in parasite differentiation. *Molecular Microbiology* **2005,** 56, (5), 1169-1182.

43. Wiese, M.; Wang, Q.; Görcke, I., Identification of mitogen-activated protein kinase homologues from *Leishmania mexicana*. *International Journal for Parasitology* **2003,** 33, (14), 1577-1587.

44. Erdmann, M.; Scholz, A.; Melzer, I. M.; Schmetz, C.; Wiese, M., Interacting protein kinases involved in the regulation of flagellar length. *Mol Biol Cell* **2006,** 17, (4), 2035-45.

45. Wiese, M.; Kuhn, D.; Grünfelder, C. G., Protein kinase involved in flagellar-length control. *Eukaryotic Cell* **2003,** 2, (4), 769-777.

46. Dell, K. R.; Engel, J. N., Stage-specific regulation of protein phosphorylation in Leishmania major. *Mol Biochem Parasitol* **1994,** 64, (2), 283-92.

47. Salotra, P.; Ralhan, R.; Sreenivas, G., Heat-stress induced modulation of protein phosphorylation in virulent promastigotes of Leishmania donovani. *Int J Biochem Cell Biol* **2000,** 32, (3), 309-16.

48. Bakalara, N.; Seyfang, A.; Davis, C.; Baltz, T., Characterization of a life-cycle-stage-regulated membrane protein tyrosine phosphatase in Trypanosoma brucei. *Eur J Biochem* **1995,** 234, (3), 871-7.

49. Szoor, B., Trypanosomatid protein phosphatases. *Mol Biochem Parasitol* **2010,** 173, (2), 53-63.

50. von Freyend, S. J.; Rosenqvist, H.; Fink, A.; Melzer, I. M.; Clos, J.; Jensen, O. N.; Wiese, M., LmxMPK4, an essential mitogen-activated protein kinase of Leishmania mexicana is phosphorylated and activated by the STE7-like protein kinase LmxMKK5. *Int J Parasitol* **2010**, 40, (8), 969-78.

51. Shah, K.; Liu, Y.; Deirmengian, C.; Shokat, K. M., Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc Natl Acad Sci U S A* **1997,** 94, (8), 3565-70.

52. Morandell, S.; Grosstessner-Hain, K.; Roitinger, E.; Hudecz, O.; Lindhorst, T.; Teis, D.; Wrulich, O. A.; Mazanek, M.; Taus, T.; Ueberall, F.; Mechtler, K.; Huber, L. A., QIKS--Quantitative identification of kinase substrates. *Proteomics* **2010**, 10, (10), 2015-25.

53. de Hoog, C. L.; Mann, M., Proteomics. *Annu Rev Genomics Hum Genet* **2004,** 5, 267-93.

54. Yates, J. R., 3rd, Mass spectrometry. From genomics to proteomics. *Trends Genet* **2000,** 16, (1), 5-8.

55. Kalume, D. E.; Molina, H.; Pandey, A., Tackling the phosphoproteome: tools and strategies. *Curr Opin Chem Biol* **2003,** 7, (1), 64-9.

56. Jensen, O. N., Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* **2004,** 8, (1), 33-41.

57. Jensen, O. N., Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol* **2006,** 7, (6), 391-403.

58. Larsen, M. R.; Roepstorff, P., Mass spectrometric identification of proteins and characterization of their post-translational modifications in proteome analysis. *Fresenius J Anal Chem* **2000,** 366, (6-7), 677-90.

59. Mann, M.; Jensen, O. N., Proteomic analysis of post-translational modifications. *Nat Biotechnol* **2003,** 21, (3), 255-61.

60. Mayya, V.; Han, D. K., Phosphoproteomics by mass spectrometry: insights, implications, applications and limitations. *Expert Rev Proteomics* **2009,** 6, (6), 605-18.

61. Ozlu, N.; Akten, B.; Timm, W.; Haseley, N.; Steen, H.; Steen, J. A., Phosphoproteomics. *Wiley Interdiscip Rev Syst Biol Med* 2, (3), 255-76.

62. Rosenqvist, H.; Yuanying, Y.; Jensen, O. N., Analytical strategies in mass spectrometry based phosphoproteomics. *Methods in Molecular Biology* **2011**.

63. Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P., Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* **2004,** 101, (33), 12130-5.

64. Cao, P.; Stults, J. T., Phosphopeptide analysis by on-line immobilized metal-ion affinity chromatography-capillary electrophoresis-electrospray ionization mass spectrometry. *J Chromatogr A* **1999,** 853, (1-2), 225-35.

65. Nuhse, T. S.; Stensballe, A.; Jensen, O. N.; Peck, S. C., Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol Cell Proteomics* **2003,** 2, (11), 1234-43.

66. Posewitz, M. C.; Tempst, P., Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem* **1999,** 71, (14), 2883-92.

67. Stensballe, A.; Andersen, S.; Jensen, O. N., Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics* **2001,** 1, (2), 207-22.

68. Feng, S.; Ye, M.; Zhou, H.; Jiang, X.; Jiang, X.; Zou, H.; Gong, B., Immobilized zirconium ion affinity chromatography for specific enrichment of phosphopeptides in phosphoproteome analysis. *Mol Cell Proteomics* **2007,** 6, (9), 1656-65.

69. Andersson, L.; Porath, J., Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. *Anal Biochem* **1986,** 154, (1), 250-4.

70. Anguenot, R.; Yelle, S.; Nguyen-Quoc, B., Purification of tomato sucrose synthase phosphorylated isoforms by Fe(III)-immobilized metal affinity chromatography. *Arch Biochem Biophys* **1999,** 365, (1), 163-9.

71. Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M., Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* **2002,** 20, (3), 301-5.

72. Ficarro, S. B.; Parikh, J. R.; Blank, N. C.; Marto, J. A., Niobium(V) oxide (Nb2O5): application to phosphoproteomics. *Anal Chem* **2008,** 80, (12), 4606-13.

73. Leitner, A.; Sturm, M.; Smatt, J. H.; Jarn, M.; Linden, M.; Mechtler, K.; Lindner, W., Optimizing the performance of tin dioxide microspheres for phosphopeptide enrichment. *Anal Chim Acta* **2009,** 638, (1), 51-7.

74. Pinkse, M. W.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J., Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* **2004,** 76, (14), 3935-43.

75. Sano, A.; Nakamura, H., Evaluation of titanium and titanium oxides as chemo-affinity sorbents for the selective enrichment of organic phosphates. *Anal Sci* **2007,** 23, (11), 1285-9.

76. Wolschin, F.; Wienkoop, S.; Weckwerth, W., Enrichment of phosphorylated proteins and peptides from complex mixtures using metal oxide/hydroxide affinity chromatography (MOAC). *Proteomics* **2005,** 5, (17), 4389-97.

77. Zhou, H.; Tian, R.; Ye, M.; Xu, S.; Feng, S.; Pan, C.; Jiang, X.; Li, X.; Zou, H., Highly specific enrichment of phosphopeptides by zirconium dioxide nanoparticles for phosphoproteome analysis. *Electrophoresis* **2007,** 28, (13), 2201-15.

78. Ikeguchi, Y.; Nakamua, H., Determination of organic phosphates by column-switching high performance anion-exchange chromatography using on-line preconcentration on titania. *Analytical Sciences* **1997,** 13, (3), 479-483.

79. Cantin, G. T.; Shock, T. R.; Park, S. K.; Madhani, H. D.; Yates, J. R., 3rd, Optimizing TiO2-based phosphopeptide enrichment for automated multidimensional liquid chromatography coupled to tandem mass spectrometry. *Anal Chem* **2007,** 79, (12), 4666-73.

80. Yu, L. R.; Zhu, Z.; Chan, K. C.; Issaq, H. J.; Dimitrov, D. S.; Veenstra, T. D., Improved titanium dioxide enrichment of phosphopeptides from HeLa cells and high confident phosphopeptide identification by cross-validation of MS/MS and MS/MS/MS spectra. *J Proteome Res* **2007,** 6, (11), 4150-62.

81. Simon, E. S.; Young, M.; Chan, A.; Bao, Z. Q.; Andrews, P. C., Improved enrichment strategies for phosphorylated peptides on titanium dioxide using methyl esterification and pH gradient elution. *Anal Biochem* **2008,** 377, (2), 234-42.

82. Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006,** 127, (3), 635-48.

83. Villen, J.; Beausoleil, S. A.; Gerber, S. A.; Gygi, S. P., Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci U S A* **2007,** 104, (5), 1488-93.

84. Zhang, X.; Ye, J.; Jensen, O. N.; Roepstorff, P., Highly Efficient Phosphopeptide Enrichment by Calcium Phosphate Precipitation Combined with Subsequent IMAC Enrichment. *Mol Cell Proteomics* **2007,** 6, (11), 2032-42.

85. de Hoffmann, E.; Stroobant, V., *Mass spectrometry, principles and applications*. 2 ed.; John Wiley & Sons, Ltd.: 2002.

86. Bakhtiar, R.; Nelson, R. W., Mass spectrometry of the proteome. *Mol Pharmacol* **2001,** 60, (3), 405-15.

87. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** 246, (4926), 64-71.

88. Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F., Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion processes* **1987,** 78, 53-66.

89. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **1988,** 60, (20), 2299-301.

90. Zhou, M.; Veenstra, T., Mass spectrometry: m/z 1983-2008. *Biotechniques* **2008,** 44, (5), 667-8, 670.

91. Han, X.; Aslanian, A.; Yates, J. R., 3rd, Mass spectrometry for proteomics. *Curr Opin Chem Biol* **2008,** 12, (5), 483-90.

92. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** 422, (6928), 198-207.

93. Borch, J.; Jorgensen, T. J.; Roepstorff, P., Mass spectrometric analysis of protein interactions. *Curr Opin Chem Biol* **2005,** 9, (5), 509-16.

94. Canas, B.; Lopez-Ferrer, D.; Ramos-Fernandez, A.; Camafeita, E.; Calvo, E., Mass spectrometry technologies for proteomics. *Brief Funct Genomic Proteomic* **2006,** 4, (4), 295-320.

95. Oh, H.; Breuker, K.; Sze, S. K.; Ge, Y.; Carpenter, B. K.; McLafferty, F. W., Secondary and tertiary structures of gaseous protein ions characterized by electron capture dissociation mass spectrometry and photofragment spectroscopy. *Proc Natl Acad Sci U S A* **2002,** 99, (25), 15863-8.

96. Ong, S. E.; Foster, L. J.; Mann, M., Mass spectrometric-based approaches in quantitative proteomics. *Methods* **2003,** 29, (2), 124-30.

97. Rosenqvist, H. Analysis of covalent cross-linking between E. coli single-stranded DNA-binding protein and ssDNA by UV-cross-linking and MALDI TOF MS - The importance of the DNA sequence for the formation of DNA-protein complexes and the efficiency of cross-linking. Bachelor, University of Southern Denmark, Odense, 2002.

98. Rosenqvist, H. Determination of DNA-peptide heteroconjugate structures by tandem MS. Individual Theoretical and Experimental Course, University of Southern Denmark, Odense, 2003.

99. Rosenqvist, H. Identification and quantification of secreted proteins from human mesenchymal stem cells. Master, University of Southern Denmark, Odense, 2005.

100. Rosenzweig, D.; Smith, D.; Myler, P. J.; Olafson, R. W.; Zilberstein, D., Post-translational modification of cellular proteins during Leishmania donovani differentiation. *Proteomics* **2008,** 8, (9), 1843-50.

101. Tsutsui, Y.; Wintrode, P. L., Hydrogen/deuterium exchange-mass spectrometry: a powerful tool for probing protein structure, dynamics and interactions. *Curr Med Chem* **2007,** 14, (22), 2344-58.

102. Hop, C. E. C. A.; Bakhtiar, R., An introduction to electrospray ionization and matrix-assited laser desorption/ionization mass spectrometry: essential tools in a modern biotechnology environment. *Biospectroscopy* **1997,** 3, (4), 259-280.

103. Chen, H. S.; Rejtar, T.; Andreev, V.; Moskovets, E.; Karger, B. L., High-speed, high-resolution monolithic capillary LC-MALDI MS using an off-line continuous deposition interface for proteomic analysis. *Anal Chem* **2005,** 77, (8), 2323-31.

104. Frohlich, T.; Arnold, G. J., Proteome research based on modern liquid chromatography--tandem mass spectrometry: separation, identification and quantification. *J Neural Transm* **2006,** 113, (8), 973-94.

105. Marvin, L. F.; Roberts, M. A.; Fay, L. B., Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry. *Clin Chim Acta* **2003,** 337, (1-2), 11-21.

106. Cech, N. B.; Enke, C. G., Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev* **2001,** 20, (6), 362-87.

107. Glish, G. L.; Vachet, R. W., The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov* **2003,** 2, (2), 140-50.

108. Yost, R. A.; Enke, C. G., Selected ion fragmentation with a tandem quadrupole mass spectrometer. *Journal of the American Chemical Society* **1978,** 100, (7), 2274-75.

109. Cox, D. M.; Zhong, F.; Du, M.; Duchoslav, E.; Sakuma, T.; McDermott, J. C., Multiple reaction monitoring as a method for identifying protein posttranslational modifications. *J Biomol Tech* **2005,** 16, (2), 83-90.

110. Unwin, R. D.; Griffiths, J. R.; Leverentz, M. K.; Grallert, A.; Hagan, I. M.; Whetton, A. D., Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. *Mol Cell Proteomics* **2005,** 4, (8), 1134-44.

111. Yocum, A. K.; Chinnaiyan, A. M., Current affairs in quantitative targeted proteomics: multiple reaction monitoring-mass spectrometry. *Brief Funct Genomic Proteomic* **2009,** 8, (2), 145-57.

112. Hughes, N.; Winnik, W.; Dunyach, J.; Amad, M.; Splendore, M.; Paul, G., High-sensitivity quantitation of cabergoline and pergolide using a triple-quadrupole mass spectrometer with enhanced mass-resolution capabilities. *Journal of Mass Spectrometry* **2003,** 38, 743-751.

113. Thermo TSQ Vantage Triple Stage Quadrupole Mass Spectrometer. http://thermoscientific.com/wps/portal/ts/products/detail?navigationId=LA11126__10962&categoryId=87169&productId=11962004

114. Thermo TSQ series hardware manual.

115. de Hoffmann, E.; Stroobant, V., *Mass spectrometry, principles and applications*. 3rd ed.; Wiley: 2007.

116. Jensen, O. N., BM21 lecture: Mass analyzers. In University of Southern Denmark: 2002.

117. Matamaros Fernandéz, L. E., BM21: Quadrupole and ion trap analyzers. In University of Southern Denmark: 2002.

118. Scigelova, M.; Makarov, A., Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* **2006,** 6 Suppl 2, 16-21.

119. Hardman, M.; Makarov, A. A., Interfacing the Orbitrap mass analyzer to an electrospray ion source. *Analytical Chemistry* **2003,** 75, (7), 1699-1705.

120. Makarov, A. A., Electrostatic axially harmonic orbital trapping: a high-performancce technique of mass analysis. *Analytical Chemistry* **2000,** 72, (6), 1156-62.

121. Perry, R. H.; Cooks, R. G.; Noll, R. J., Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom Rev* **2008,** 27, (6), 661-99.

122. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R., The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **2005,** 40, (4), 430-43.

123. Kocher, T.; Engstrom, A.; Zubarev, R. A., Fragmentation of peptides in MALDI in-source decay mediated by hydrogen radicals. *Anal Chem* **2005,** 77, (1), 172-7.

124. Purcell, A. W.; Gorman, J. J., The use of post-source decay in matrix-assisted laser desorption/ionisation mass spectrometry to delineate T cell determinants. *J Immunol Methods* **2001,** 249, (1-2), 17-31.

125. Sleno, L.; Volmer, D. A., Ion activation methods for tandem mass spectrometry. *J Mass Spectrom* **2004,** 39, (10), 1091-112.

126. Waters, Q-TOF Premier operator's guide. In.

127. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **2007,** 4, (9), 709-12.

128. Bakhtiar, R.; Guan, Z., Electron capture dissociation mass spectrometry in characterization of post-translational modifications. *Biochem Biophys Res Commun* **2005,** 334, (1), 1-8.

129. Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F., The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta* **2006,** 1764, (12), 1811-22.

130. Thermo Using Multistage Activation in an Ion Trap Mass Spectometer. http://www.thermo.com/eThermo/CMA/PDFs/Articles/articlesFile_30220.pdf

131. Segu, Z. M.; Mechref, Y., Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Commun Mass Spectrom* **2010**, 24, (9), 1217-25.

132. Zhang, Y.; Ficarro, S. B.; Li, S.; Marto, J. A., Optimized Orbitrap HCD for quantitative analysis of phosphopeptides. *J Am Soc Mass Spectrom* **2009,** 20, (8), 1425-34.

133. Huang, C. Y.; Chang, C. P.; Huang, C. L.; Ferrell, J. E., Jr., M phase phosphorylation of cytoplasmic dynein intermediate chain and p150(Glued). *J Biol Chem* **1999,** 274, (20), 14262-9.

134. Morales, M. A.; Watanabe, R.; Laurent, C.; Lenormand, P.; Rousselle, J. C.; Namane, A.; Spath, G. F., Phosphoproteomic analysis of Leishmania donovani pro- and amastigote stages. *Proteomics* **2008,** 8, (2), 350-63.

135. El Benna, J.; Han, J.; Park, J. W.; Schmid, E.; Ulevitch, R. J.; Babior, B. M., Activation of p38 in stimulated human neutrophils: phosphorylation of the oxidase component p47phox by p38 and ERK but not by JNK. *Arch Biochem Biophys* **1996,** 334, (2), 395-400.

136. Lee, H. J.; Na, K.; Kwon, M. S.; Kim, H.; Kim, K. S.; Paik, Y. K., Quantitative analysis of phosphopeptides in search of the disease biomarker from the hepatocellular carcinoma specimen. *Proteomics* **2009,** 9, (12), 3395-408.

137. Kawakami, H.; Ohtsuki, S.; Kamiie, J.; Suzuki, T.; Abe, T.; Terasaki, T., Simultaneous absolute quantification of 11 cytochrome P450 isoforms in human liver microsomes by liquid chromatography tandem mass spectrometry with In silico target peptide selection. *J Pharm Sci*. **2011**, 100, (1), 341-52

138. Silva, J. C.; Gorenstein, M. V.; Li, G. Z.; Vissers, J. P.; Geromanos, S. J., Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **2006,** 5, (1), 144-56.

139. Manteca, A.; Sanchez, J.; Jung, H. R.; Schwammle, V.; Jensen, O. N., Quantitative proteomics analysis of Streptomyces coelicolor development demonstrates that onset of secondary metabolism coincides with hypha differentiation. *Mol Cell Proteomics* **2010**, 9, (7), 1423-36.

140. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **2007,** 389, (4), 1017-31.

141. Tedford, N. C.; Hall, A. B.; Graham, J. R.; Murphy, C. E.; Gordon, N. F.; Radding, J. A., Quantitative analysis of cell signaling and drug action via mass

spectrometry-based systems level phosphoproteomics. *Proteomics* **2009,** 9, (6), 1469-87.

142. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002,** 1, (5), 376-86.

143. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004,** 3, (12), 1154-69.

144. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C., Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003,** 75, (8), 1895-904.

145. Zhang, H.; Brown, R. N.; Qian, W. J.; Monroe, M. E.; Purvine, S. O.; Moore, R. J.; Gritsenko, M. A.; Shi, L.; Romine, M. F.; Fredrickson, J. K.; Pasa-Tolic, L.; Smith, R. D.; Lipton, M. S., Quantitative analysis of cell surface membrane proteins using membrane-impermeable chemical probe coupled with 18O labeling. *J Proteome Res* **2010**, 9, (5), 2160-9.

146. Unwin, R. D., Quantification of proteins by iTRAQ. *Methods Mol Biol* **2010**, 658, 205-15.

147. Boja, E. S.; Phillips, D.; French, S. A.; Harris, R. A.; Balaban, R. S., Quantitative mitochondrial phosphoproteomics using iTRAQ on an LTQ-Orbitrap with high energy collision dissociation. *J Proteome Res* **2009,** 8, (10), 4665-75.

148. Sachon, E.; Mohammed, S.; Bache, N.; Jensen, O. N., Phosphopeptide quantitation using amine-reactive isobaric tagging reagents and tandem mass spectrometry: application to proteins isolated by gel electrophoresis. *Rapid Commun Mass Spectrom* **2006,** 20, (7), 1127-34.

149. Zhang, Y.; Wolf-Yadlin, A.; Ross, P. L.; Pappin, D. J.; Rush, J.; Lauffenburger, D. A.; White, F. M., Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* **2005,** 4, (9), 1240-50.

150. Thingholm, T. E.; Palmisano, G.; Kjeldsen, F.; Larsen, M. R., Undesirable charge-enhancement of isobaric tagged phosphopeptides leads to reduced identification efficiency. *J Proteome Res* **2010,** 9, (8), 4045-52.

151. Wu, J.; Warren, P.; Shakey, Q.; Sousa, E.; Hill, A.; Ryan, T. E.; He, T., Integrating titania enrichment, iTRAQ labeling, and Orbitrap CID-HCD for global identification and quantitative analysis of phosphopeptides. *Proteomics* **2010** 10, (11), 2224-34.

152. Bantscheff, M.; Boesche, M.; Eberhard, D.; Matthieson, T.; Sweetman, G.; Kuster, B., Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Mol Cell Proteomics* **2008,** 7, (9), 1702-13.

153. Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I.; Wright, P. C., iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J Proteome Res* **2009,** 8, (11), 5347-55.

154. Treumann, A.; Thiede, B., Isobaric protein and peptide quantification: perspectives and issues. *Expert Rev Proteomics* **2010**, 7, (5), 647-53.

155. Kjeldsen, F.; Giessing, A. M.; Ingrell, C. R.; Jensen, O. N., Peptide sequencing and characterization of post-translational modifications by enhanced ion-charging and liquid chromatography electron-transfer dissociation tandem mass spectrometry. *Anal Chem* **2007,** 79, (24), 9243-52.

156. Phanstiel, D.; Zhang, Y.; Marto, J. A.; Coon, J. J., Peptide and protein quantification using iTRAQ with electron transfer dissociation. *J Am Soc Mass Spectrom* **2008,** 19, (9), 1255-62.

157. Zhu, W.; Smith, J. W.; Huang, C. M., Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* **2010**, 840518.

158. Asara, J. M.; Christofk, H. R.; Freimark, L. M.; Cantley, L. C., A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* **2008,** 8, (5), 994-9.

159. Cirulli, C.; Chiappetta, G.; Marino, G.; Mauri, P.; Amoresano, A., Identification of free phosphopeptides in different biological fluids by a mass spectrometry approach. *Anal Bioanal Chem* **2008,** 392, (1-2), 147-59.

160. Glinski, M.; Weckwerth, W., Differential multisite phosphorylation of the trehalose-6-phosphate synthase gene family in Arabidopsis thaliana: a mass spectrometry-based process for multiparallel peptide library phosphorylation analysis. *Mol Cell Proteomics* **2005,** 4, (10), 1614-25.

161. Johnson, R. P.; El-Yazbi, A. F.; Hughes, M. F.; Schriemer, D. C.; Walsh, E. J.; Walsh, M. P.; Cole, W. C., Identification and functional characterization of protein kinase A-catalyzed phosphorylation of potassium channel Kv1.2 at serine 449. *J Biol Chem* **2009,** 284, (24), 16562-74.

162. Palmisano, G.; Thingholm, T. E., Strategies for quantitation of phosphoproteomic data. *Expert Rev Proteomics* **2010**, 7, (3), 439-56.

163. Zappacosta, F.; Collingwood, T. S.; Huddleston, M. J.; Annan, R. S., A quantitative results-driven approach to analyzing multisite protein phosphorylation: the phosphate-dependent phosphorylation profile of the transcription factor Pho4. *Mol Cell Proteomics* **2006,** 5, (11), 2019-30.

164. Levin, Y.; Bahn, S., Quantification of proteins by label-free LC-MS/MS. *Methods Mol Biol* **2010**, 658, 217-31.

165. Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G. Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P.; Geromanos, S., Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* **2005,** 77, (7), 2187-200.

166. Vissers, J. P.; Pons, S.; Hulin, A.; Tissier, R.; Berdeaux, A.; Connolly, J. B.; Langridge, J. I.; Geromanos, S. J.; Ghaleh, B., The use of proteome similarity for the qualitative and quantitative profiling of reperfused myocardium. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009,** 877, (13), 1317-26.

167. ABSciex MRMPilot™. http://www.absciex.com/Products/Software/MRMPilot-Software

168. MacCoss Lab Software, D. o. G. S., University of Washington School of Medicine Skyline Targeted Proteomics Environment. https://brendanx-uw1.gs.washington.edu/labkey/wiki/home/software/Skyline/page.view?name=default

169. Thermo Pinpoint Software. http://www.thermoscientific.com/wps/portal/ts/products/detail?productId=12784473

170. Cuervo, P.; de Jesus, J. B.; Junqueira, M.; Mendonca-Lima, L.; Gonzalez, L. J.; Betancourt, L.; Grimaldi, G., Jr.; Domont, G. B.; Fernandes, O.; Cupolillo, E., Proteome analysis of Leishmania (Viannia) braziliensis by two-dimensional gel electrophoresis and mass spectrometry. *Mol Biochem Parasitol* **2007,** 154, (1), 6-21.

171. Dea-Ayuela, M. A.; Rama-Iniguez, S.; Bolas-Fernandez, F., Proteomic analysis of antigens from Leishmania infantum promastigotes. *Proteomics* **2006,** 6, (14), 4187-94.

172. Leifso, K.; Cohen-Freue, G.; Dogra, N.; Murray, A.; McMaster, W. R., Genomic and proteomic expression analysis of Leishmania promastigote and amastigote

life stages: the Leishmania genome is constitutively expressed. *Mol Biochem Parasitol* **2007,** 152, (1), 35-46.

173. Nugent, P. G.; Karsani, S. A.; Wait, R.; Tempero, J.; Smith, D. F., Proteomic analysis of Leishmania mexicana differentiation. *Mol Biochem Parasitol* **2004,** 136, (1), 51-62.

174. Hem, S.; Gherardini, P. F.; Osorio y Fortea, J.; Hourdel, V.; Morales, M. A.; Watanabe, R.; Pescher, P.; Kuzyk, M. A.; Smith, D.; Borchers, C. H.; Zilberstein, D.; Helmer-Citterich, M.; Namane, A.; Spath, G. F., Identification of Leishmania-specific protein phosphorylation sites by LC-ESI-MS/MS and comparative genomics analyses. *Proteomics* **2010**, 10, (21), 3868-83.

175. Morales, M. A.; Watanabe, R.; Dacher, M.; Chafey, P.; Osorio y Fortea, J.; Scott, D. A.; Beverley, S. M.; Ommen, G.; Clos, J.; Hem, S.; Lenormand, P.; Rousselle, J. C.; Namane, A.; Spath, G. F., Phosphoproteome dynamics reveal heat-shock protein complexes specific to the Leishmania donovani infectious stage. *Proc Natl Acad Sci U S A* **2010**, 107, (18), 8381-6.

176. Paape, D.; Barrios-Llerena, M. E.; Le Bihan, T.; Mackay, L.; Aebischer, T., Gel free analysis of the proteome of intracellular Leishmania mexicana. *Mol Biochem Parasitol* **2010**, 169, (2), 108-14.

177. Kuhn, D.; Wiese, M., LmxPK4, a mitogen-activated protein kinase kinase homologue of Leishmania mexicana with a potential role in parasite differentiation. *Mol Microbiol* **2005,** 56, (5), 1169-82.

178. Wiese, M.; Kuhn, D.; Grunfelder, C. G., Protein kinase involved in flagellar-length control. *Eukaryot Cell* **2003,** 2, (4), 769-77.

179. Wiese, M.; Wang, Q.; Gorcke, I., Identification of mitogen-activated protein kinase homologues from Leishmania mexicana. *Int J Parasitol* **2003,** 33, (14), 1577-87.

180. Rappsilber, J.; Ryder, U.; Lamond, A. I.; Mann, M., Large-scale proteomic analysis of the human spliceosome. *Genome Res* **2002,** 12, (8), 1231-45.

181. Li, G. Z.; Vissers, J. P.; Silva, J. C.; Golick, D.; Gorenstein, M. V.; Geromanos, S. J., Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **2009,** 9, (6), 1696-719.

182. GeneDB GeneDB. http://www.genedb.org/Homepage/Lmexicana

183. Hertz-Fowler, C., Lmexicana6frametrans_v1assembly_Jun08. In The Welcome Trust Sanger Institute: 2008.

184. Welcome Trust Sanger Institute, p. s. u., GeneDB. In.

185. Højrup, P. *GPMAW*, 8.00sr1; 2008.

186. Beausoleil, S. A.; Villen, J.; Gerber, S. A.; Rush, J.; Gygi, S. P., A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* **2006,** 24, (10), 1285-92.

187. Rosenqvist, H. *PRG 8 month report: Analysis of the Leishmania mexicana proteome and phosphoproteome*; University of Strathclyde: 2008.

188. Rosenqvist, H. *PRG 20 month report: Analyses of the Leishmania mexicana phosphoproteome*; University of Strathclyde: 2009.

189. Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd, An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **2001,** 73, (23), 5683-90.

190. Liska, A. J.; Shevchenko, A., Combining mass spectrometry with database interrogation strategies in proteomics. *Trac-Trends in Analytical Chemistry* **2003,** 22, (5), 291-+.

191. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** 20, (18), 3551-67.

192. Savitski, M. M.; Lemeer, S.; Boesche, M.; Lang, M.; Mathieson, T.; Bantscheff, M.; Kuster, B., Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics* **2011**, 10, (2), M110.003830.

193. *Minitab 16*, Minitab Ltd.: 2010.

194. Morales, M. A.; Renaud, O.; Faigle, W.; Shorte, S. L.; Spath, G. F., Over-expression of Leishmania major MAP kinases reveals stage-specific induction of phosphotransferase activity. *Int J Parasitol* **2007,** 37, (11), 1187-99.

195. Nett, I. R.; Martin, D. M.; Miranda-Saavedra, D.; Lamont, D.; Barber, J. D.; Mehlert, A.; Ferguson, M. A., The phosphoproteome of bloodstream form Trypanosoma brucei, causative agent of African sleeping sickness. *Mol Cell Proteomics* **2009,** 8, (7), 1527-38.

196. Kyono, Y.; Sugiyama, N.; Imami, K.; Tomita, M.; Ishihama, Y., Successive and selective release of phosphorylated peptides captured by hydroxy acid-modified metal oxide chromatography. *J Proteome Res* **2008,** 7, (10), 4585-93.

197. Silverman, J. M.; Chan, S. K.; Robinson, D. P.; Dwyer, D. M.; Nandan, D.; Foster, L. J.; Reiner, N. E., Proteomic analysis of the secretome of Leishmania donovani. *Genome Biol* **2008,** 9, (2), R35.

198. McNulty, D. E.; Annan, R. S., Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol Cell Proteomics* **2008,** 7, (5), 971-80.

199. Dai, J.; Jin, W. H.; Sheng, Q. H.; Shieh, C. H.; Wu, J. R.; Zeng, R., Protein phosphorylation and expression profiling by Yin-yang multidimensional liquid chromatography (Yin-yang MDLC) mass spectrometry. *J Proteome Res* **2007,** 6, (1), 250-62.

200. Dai, J.; Wang, L. S.; Wu, Y. B.; Sheng, Q. H.; Wu, J. R.; Shieh, C. H.; Zeng, R., Fully automatic separation and identification of phosphopeptides by continuous pH-gradient anion exchange online coupled with reversed-phase liquid chromatography mass spectrometry. *J Proteome Res* **2009,** 8, (1), 133-41.

201. Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N., Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* **2005,** 4, (3), 310-27.

202. Christensen, G. L.; Kelstrup, C. D.; Lyngso, C.; Sarwar, U.; Bogebo, R.; Sheikh, S. P.; Gammeltoft, S.; Olsen, J. V.; Hansen, J. L., Quantitative phosphoproteomics dissection of seven-transmembrane receptor signaling using full and biased agonists. *Mol Cell Proteomics* **2010**, 9, (7), 1540-53.

203. Mok, J.; Kim, P. M.; Lam, H. Y.; Piccirillo, S.; Zhou, X.; Jeschke, G. R.; Sheridan, D. L.; Parker, S. A.; Desai, V.; Jwa, M.; Cameroni, E.; Niu, H.; Good, M.; Remenyi, A.; Ma, J. L.; Sheu, Y. J.; Sassi, H. E.; Sopko, R.; Chan, C. S.; De Virgilio, C.; Hollingsworth, N. M.; Lim, W. A.; Stern, D. F.; Stillman, B.; Andrews, B. J.; Gerstein, M. B.; Snyder, M.; Turk, B. E., Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. *Sci Signal* **2010**, 3, (109), ra12.

204. Bandeira, N., Protein identification by spectral networks analysis. *Methods Mol Biol* **2011**, 694, 151-68.

205. Hjerno, K.; Alm, R.; Canback, B.; Matthiesen, R.; Trajkovski, K.; Bjork, L.; Roepstorff, P.; Emanuelsson, C., Down-regulation of the strawberry Bet v 1-homologous allergen in concert with the flavonoid biosynthesis pathway in colorless strawberry mutant. *Proteomics* **2006,** 6, (5), 1574-87.

206. Lucker, J.; Laszczak, M.; Smith, D.; Lund, S. T., Generation of a predicted protein database from EST data and application to iTRAQ analyses in grape

(Vitis vinifera cv. Cabernet Sauvignon) berries at ripening initiation. *BMC Genomics* **2009,** 10, 50.

207. Ishii, A.; Dutta, R.; Wark, G. M.; Hwang, S. I.; Han, D. K.; Trapp, B. D.; Pfeiffer, S. E.; Bansal, R., Human myelin proteome and comparative analysis with mouse myelin. *Proc Natl Acad Sci U S A* **2009,** 106, (34), 14605-10.

208. ProteomeSoftware Frequently Asked Questions about Scaffold. http://proteome-software.wikispaces.com/FAQ+-+Statistics

209. Chavez-Fumagalli, M. A.; Costa, M. A.; Oliveira, D. M.; Ramirez, L.; Costa, L. E.; Duarte, M. C.; Martins, V. T.; Oliveira, J. S.; Olortegi, C. C.; Bonay, P.; Alonso, C.; Tavares, C. A.; Soto, M.; Coelho, E. A., Vaccination with the Leishmania infantum ribosomal proteins induces protection in BALB/c mice against Leishmania chagasi and Leishmania amazonensis challenge. *Microbes Infect* **2010**, 12, (12-13), 967-77.

210. Iborra, S.; Parody, N.; Abanades, D. R.; Bonay, P.; Prates, D.; Novais, F. O.; Barral-Netto, M.; Alonso, C.; Soto, M., Vaccination with the Leishmania major ribosomal proteins plus CpG oligodeoxynucleotides induces protection against experimental cutaneous leishmaniasis in mice. *Microbes Infect* **2008,** 10, (10-11), 1133-41.

211. Poot, J.; Janssen, L. H.; van Kasteren-Westerneng, T. J.; van der Heijden-Liefkens, K. H.; Schijns, V. E.; Heckeroth, A., Vaccination of dogs with six different candidate leishmaniasis vaccines composed of a chimerical recombinant protein containing ribosomal and histone protein epitopes in combination with different adjuvants. *Vaccine* **2009,** 27, (33), 4439-46.

212. Ramirez, L.; Iborra, S.; Cortes, J.; Bonay, P.; Alonso, C.; Barral-Netto, M.; Soto, M., BALB/c mice vaccinated with Leishmania major ribosomal proteins extracts combined with CpG oligodeoxynucleotides become resistant to disease caused by a secondary parasite challenge. *J Biomed Biotechnol* **2010**, 181690.

213. Breitkreutz, A.; Choi, H.; Sharom, J. R.; Boucher, L.; Neduva, V.; Larsen, B.; Lin, Z. Y.; Breitkreutz, B. J.; Stark, C.; Liu, G.; Ahn, J.; Dewar-Darch, D.; Reguly, T.; Tang, X.; Almeida, R.; Qin, Z. S.; Pawson, T.; Gingras, A. C.; Nesvizhskii, A. I.; Tyers, M., A global protein kinase and phosphatase interaction network in yeast. *Science* **2010**, 328, (5981), 1043-6.