DOCTOR OF PHILOSOPHY

Proteomic analysis of protein complexes and cell cycle regulation in Trypanosoma brucei

Crozier, Thomas William Monteiro

Award date:
2016

Awarding institution:
University of Dundee

Proteomic analysis of protein complexes and cell-cycle regulation in *Trypanosoma brucei*

Thomas William Monteiro Crozier

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

Division of Gene Regulation and Expression

School of Life Sciences

2016

Contents

Acknowledgements

First of all, I would like to thank my supervisors, Mike and Angus, for their support, freedom to develop my ideas, and the incredible laboratory facilities they made available to me during the course of my time in Dundee. Thanks also to the Wellcome Trust for the generous funding of the work presented here.

I would also like to thank Mark, Tony and Lucia who all supervised and guided me in different strands of my work on a day to day basis. Their advice on topics ranging from biochemistry, bioinformatics and *T. brucei* cell culture was invaluable in the development of the work for this thesis.

Thank you also to Michele, with whom I closely collaborated on many different projects involving bioinformatics. Without him this project would not have been possible.

To the current and past members of both the MAJF and AIL labs who made my experience in Dundee all the more enjoyable and fun.

To Rachel and Imogen, two unforgettable friends who generously welcomed me and introduced me to life in Dundee.

Thank you to both my parents, Ana and Alan, for their constant support, encouragement and belief in me to succeed.

Finally, thank you to Yasamin for being by my side and always reminding me of the positive things in life.

Declaration

Candidate

I declare that I am the sole author of this thesis and that all references cited have been consulted by me personally. The work, of which this thesis is a record, has been done by myself unless otherwise acknowledged. This work has not been previously submitted for a higher degree.

……

Thomas Crozier

Supervisors

I certify that the conditions of the relevant Ordinance and Regulations have been fulfilled.

……

Professor Michael A. J. Ferguson

……

Professor Angus I. Lamond

Abstract

Trypanosoma brucei is a unicellular trypanosomatid protozoan parasite and the etiological agent of sleeping sickness in sub-Saharan Africa. The trypanosomatid order also includes the parasites *Trypanosoma cruzi* (Chagas disease) and *Leishmania major* (Leishmaniasis). Sleeping sickness is estimated to cause ~10,000 deaths per year and current treatments are expensive, difficult to administer and toxic. Although genomic sequencing of all three parasites has identified the coding sequences of these organisms, much is still unknown about protein function, with 64% of identified genes annotated as "hypothetical", lacking obvious homology with proteins of known function. To further understand the unusual biology of this family of eukaryotes, this thesis aimed to provide evidence for protein function in *Trypanosoma brucei* in a high-throughput manner, utilising global proteomic analyses. This work has encompassed two main approaches: The global analysis of protein interactions and the analysis of proteome changes across the cell-cycle.

To enable these approaches, I developed protocols for proteome wide analysis of protein complexes in *Trypanosoma brucei*, combining multiple forms of chromatography on 'native' lysates of cells to produce a proteome wide map of core, soluble protein complexes in this organism. I further performed preliminary studies to optimise *in vivo* formaldehyde crosslinking in *T. brucei* in order to characterise membrane bound protein complexes.

I also developed methodologies to produce large populations of procyclic *T. brucei* cells highly enriched in different phases of the cell-cycle for proteomic analysis. In conjunction with the optimisation of methods for isobaric tag quantitation on Fusion mass spectrometers, I provide the first characterisation of protein regulation during cell division in *T. brucei* at an unparalleled proteomic depth.

Together, these datasets provide a wealth of information about the interaction and cell cycle regulation of many thousands of proteins in *T. brucei*, and contributes greatly to the understanding of protein function in trypanosomatid organisms. I highlight the ability of these methods to predict novel protein complexes, predict interactions between "hypothetical" proteins with proteins of known function, and to identify "hypothetical" cell-cycle regulated proteins that are essential for growth of the parasite, that are a potentially interesting source for novel drug targets. Data visualisation tools to browse the data in a user-friendly format will further allow the trypanosmatid research community to mine these datasets to understand function of proteins of interest and continue to extract functional information from these datasets to extend our understanding of trypanosomatid biology.

List of figures

List of tables

Abbreviations

AAT – Animal African Trypanosomiasis

ABK – aurora B kinase

AGC – automatic gain control

ANOVA – analysis of variance

AP-MS – affinity purification mass spectrometry

BSA – bovine serum albumin

BSF – bloodstream form

CBQCA – 3-(4-carboxybenzoyl) quinoline-2-carboxalehyde assay

CDK – cyclin dependent kinase

CID – collision induced dissociation

CPC – chromosomal passenger complex

CRK – cdc2 related kinase

DMSO –dimethyl sulfoxide

DRBD – double RNA binding domain

DSP – dithiobis(succinimidyl propionate)

EDTA – ethylenediametetraacetic acid

GD – gold standard dataset

GG-mito – glycerol gradient of mitochondrial extract

GG-WCL – glycerol gradient of whole cell lysate

gHAT – *T. b. gambiense* Human African Trypanosomiasis

GO – gene ontology

HAT – Human African Trypanosomiasis

HCD – higher energy collisional dissociation

HPLC – high performance liquid chromatography

iBAQ – intensity based absolute quantitation

IEX – ion exchange chromatography

IEX-cyto – ion exchange chromatography of cytosolic extract

IEX-mito – ion exchange chromatography of mitochondrial extract

KKT - kinetochore

LDS – lithium dodecyl sulphate

LFQ – label free quantitation

maxIT – maximum injection time

MS – mass spectrometry

NCC – normalised cross-correlation

NEM – N-ethylmaleimide

OMSSA – open mass spectrometry search algorithim

PBS – phosphate buffered saline

PCC – Pearson cross-correlation

PCF – procyclic form

PCP-MS – protein correlation profiling mass spectrometry

PGC – polycistronic gene cluster

PMSF – phenyl-methyl sulfonyl fluoride

psms – peptide spectral matches

PSP1 – polymerase suppressor 1 domain

PTM – post translational modification

PUF – pumilio protein domain

rHAT – *T. b. rhodisiense* Human African Trypanosomiasis

RITseq – RNA interference target sequencing

ROC – receiver operator curve

SAX – strong anion exchange chromatography

SDS – sodium dodecyl sulphate

SEC – size exclusion chromatography

SPS – synchronous precursor selection

SSR – strand swith regions

TCEP – tris(2-carboxyethyl)phosphine

TFA – trifluoroacetic acid

TLCK – 1-5-chloro-3-tosylamido-7-amino-2-heptone

TMT – tandem mass tag

UTR – untranslated region

VSG – variant surface glycoprotein

Chapter 1: Introduction

1.1 Trypanosoma brucei and sleeping sickness

Trypanosoma brucei is a unicellular trypanosomatid protozoan parasite, causing African sleeping sickness, or Human African Trypanosomiasis (HAT), a disease endemic to sub-Saharan Africa. Two sub-species of *T. brucei* (*T. b. gambiense* and *T. b. rhodisiense*) cause HAT, though they produce distinct pathologies and disease in different regions of Africa (Figure 1.1). *T. b. gambiense* is responsible for 98% of the detected cases of sleeping sickness in the past 10 years, causing a long-term chronic disease (gHAT) found in western and central Africa. *T. b. rhodisiense* is found in eastern and southern Africa, causing an acute form of sleeping sickness (rHAT), and is considered a zoonosis infecting mainly wild animals and livestock, only occasionally infecting humans (Franco et al., 2014). Infected tsetse flies of the *Glossina* genus act as the main vectors for transmission, releasing mammalian infective parasites in saliva when taking a blood-meal (Kennedy, 2008).

The disease is characterised by two clinical phases. In the first haemo-lymphatic stage, the parasite is found in the blood-stream and lymphatic system of the mammalian host. Symptoms include headache, intermittent fever, weakness and anaemia and can be easily confused for other diseases in the region. If untreated the disease progresses into the second meningo-encephalitic stage where parasites cross the blood-brain-barrier, accessing the central nervous system. Symptoms at this stage encompass a wide range of neurological disorders, including dysregulated sleeping patterns that give the name to the disease (Brun et al., 2010). If left untreated the disease is fatal in almost 100% of cases (Stuart et al., 2008). The acute rHAT can result in death within a period of

months, whereas chronic gHAT may remain in the first stage for many years before the second stage is observed (Brun et al., 2010).

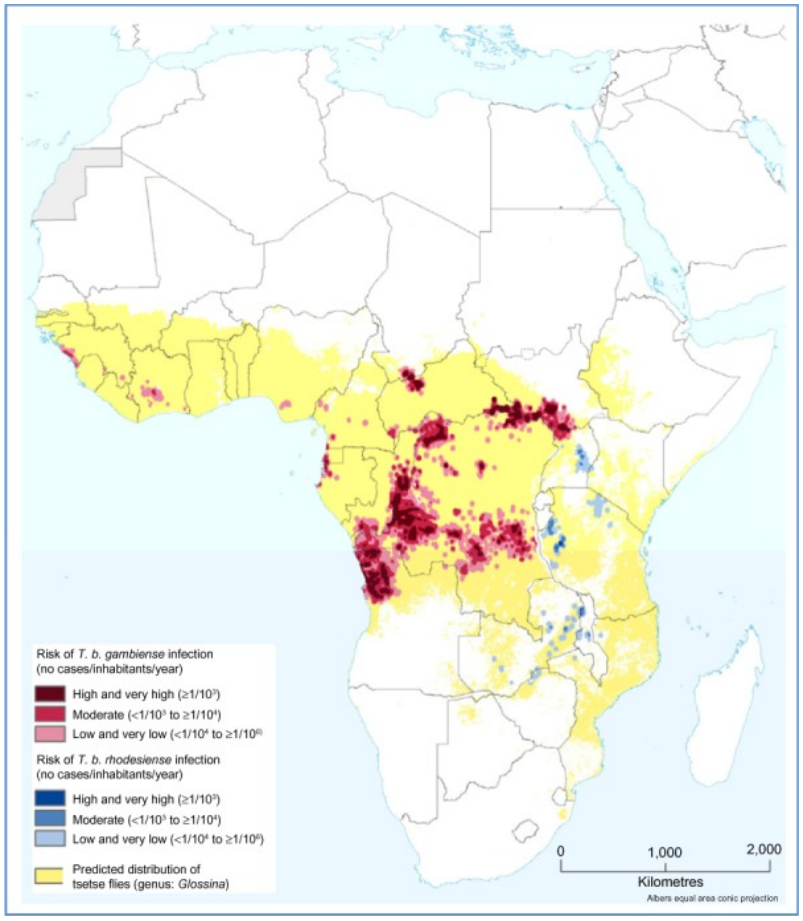


Figure 1.1: Distribution of gHAT and rHAT. Risk of *T. b. gambiense* infection in red, concentrated in Western Africa, and risk of *T. b. rhodisiense* infection in blue, concentrated in Eastern Africa. Yellow colour depicts distribution of the tsetse fly vectors. Reproduced from *(Franco et al., 2014)*.

HAT mainly affects a young working age population in poor and rural areas. Together with the chronic nature of gHAT this results in a dramatic economic impact on already impoverished communities, as added on to the cost of seeking diagnosis and medical treatment is the loss of income generators within families (Franco et al., 2014). Further economic damage results from Animal African Trypanosomiasis (AAT), more commonly known as nagana, which is caused by *T. b. brucei*, *T. congolense* and *T. vivax* (Fevre et al., 2006). AAT leads to a loss of productivity in livestock species and

can make this kind of farming impossible in endemic regions (Kristjanson et al., 1999; Swallow, 2000).

The World Health Organisation has led and co-ordinated joint efforts to reduce a growing HAT epidemic which peaked at an estimated 300,000 cases in the late 1990s (WHO, 1998), with the aim of eventually eliminating HAT transmission by 2030 (Simarro et al., 2015). Measures that increased surveillance and control of HAT implemented since 2000 have been highly successful, resulting in a 76% drop in case prevalance. Only 7,000 cases were detected between 2010-13, though due to under-reporting it is estimated there are about 20,000 cases per year (Franco et al., 2014).

There are only a handful of drugs used for treatment of HAT. gHAT is treated with pentamidine in the first stage of disease and with either, eflornitihine, or melarsoprol in the second stage. rHAT is treated with suramin during the first stage of disease and melarsoprol in the second stage (Brun et al., 2010; Steverding, 2010). These drugs can be expensive, have toxic side effects and require lengthy and complicated treatments. Although the recent development of combination therapy with nifurtimox and eflornithine has resulted in shorter treatment times and reduced toxicity, it still requires invasive intravenous administration (Priotto et al., 2009). It is therefore necessary to develop new drugs that can be taken orally and target the second-stage of disease, with fewer toxic side effects (Barrett et al., 2007).

Eflornithine is the only trypanocidal drug developed with a known mechanism of action, irreversibly inhibiting ornithine decarboxylase (Bacchi et al., 1980). The identification of dye-based compounds, arsenicals and the hypoglycaemic drug, synthalin, as trypanocides, led to the development of suramin, melarsoprol and pentamidine in the first half of the 20th century (Steverding, 2010). However, due to the

identification of these drugs through phenotypic screening, their mechanisms of action are poorly understood. Regulatory requirements now require a better understanding of drug efficacy and toxicity, which favours a target based approach to drug discovery. The identification of useful drug targets requires the understanding of the biological functions and roles of proteins in the parasite (Horn, 2014).

1.2 The life-cycle of T. brucei

Trypanosomes undergo a complicated life-cycle involving cellular differentiation, which is necessary for survival in the different environments of the insect vector and mammalian host (Figure 1.2). *T. brucei* survives in the bloodstream and lymphatic system of infected mammals as a long slender bloodstream form (BSF). This form of the parasite is proliferative, actively dividing within the mammalian host. The BSF of *T. brucei* is densely covered with a variant surface glycoprotein (VSG) coat (Cross, 1975), which is recognised by the immune system. A single VSG is expressed at any one time, and there is a large number of genes that can encode for them. Within an infecting population a subset of cells will switch their VSG encoding gene, allowing escape from the host immune response and expansion of a trypanosome population encoding a different VSG (Cross, 1979; Horn and Cross, 1997; Pays and Nolan, 1998). This process of antigen switching leads to the characteristic intermittent waves of parasitaemia observed in sleeping sickness (Doyle et al., 1980; MacGregor et al., 2012).

As the BSF population increases, a quorum sensing mechanism causes the long slender BSF parasites to differentiate into a short stumpy form that is cell-cycle arrested (Mony et al., 2014; Reuner et al., 1997; Vassella et al., 1997). This differentiation is thought to limit cell numbers, prolonging host survival, and prepare cells for the environment of the insect vector (MacGregor et al., 2012). Following a blood-meal, stumpy form cells further differentiate into replicative procyclic form (PCF) cells in the

tsetse fly midgut (Dyer et al., 2013). The mitochondrion of PCF cells is elaborated with the formation of cristae as proline becomes their major energy source instead of the glucose that is plentiful in mammalian blood (Milne et al., 1998; Vickerman, 1965; Vickerman et al., 1988).

PCF cells then migrate through to the proventriculus, differentiating into an epimastigote form. The epimastigote form divides asymmetrically into a long slender motile cell, like the parent cell, and a short cell with a short flagellum that attaches to the salivary gland (Van Den Abbeele et al., 1999). While attached to the salivary gland, the short epimastigote form cells differentiate into mammalian infective metacyclic cells, which are released from the endothelial wall into the salivary gland, ready to be transmitted in the next blood-meal (Vickerman et al., 1988). Metacyclic cells express one of around twenty-seven metacyclic specific VSGs as protection against the host immune system (Barry et al., 1998; Tetley et al., 1987).

Upon release into the dermal connective tissue of the host, the parasites migrate into the bloodstream via the lymphatic system and differentiate into BSF cells, completing the *T. brucei* life-cycle (Vickerman, 1985).

The parasite can also undergo the process of sexual recombination between different strains of *T. brucei*, a non-essential component of the parasite life-cycle that occurs in the salivary gland of the tse-tse fly. Prior to sexual recombination, parasites divide by meiosis to produce haploid gametes, which have a distinct morphology with a long free flagellum and pear-shaped body. Haploid gametes interact through intertwining their flagella prior to fusion of the two cells, though the mechanics of DNA exchange are currently unknown (Gibson, 2015).

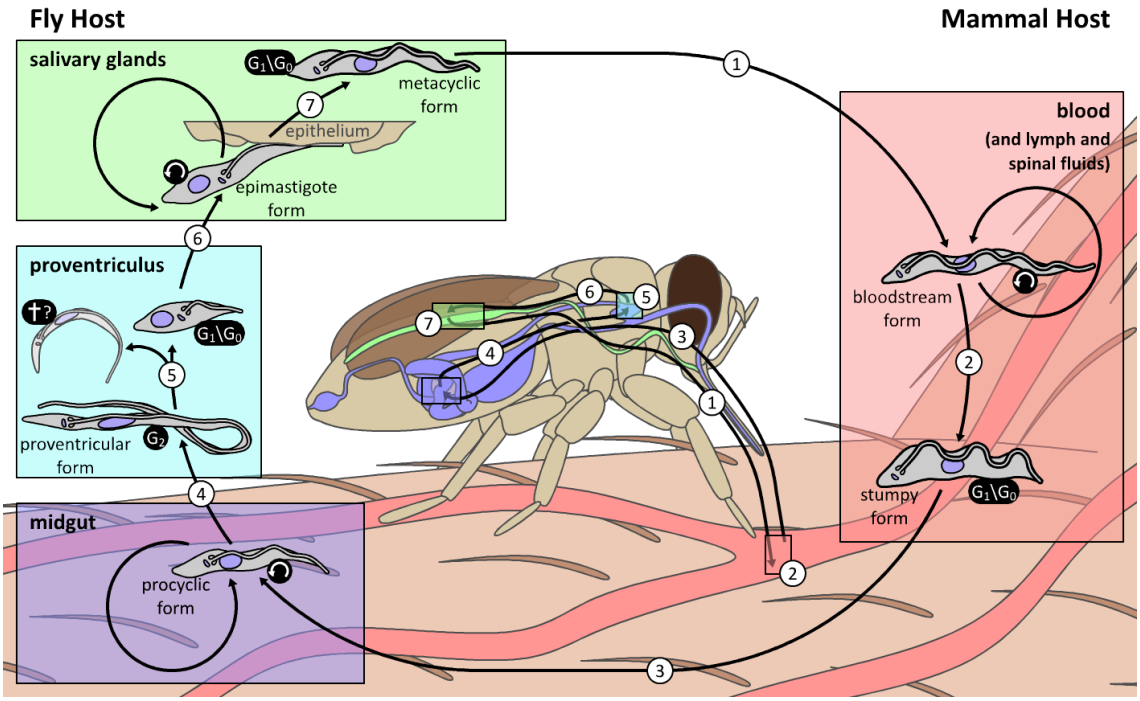


Figure 1.2: Life-cycle of *T. brucei*. Right-hand side depicts mammalian blood-stream stages of the parasite. Left-hand side depicts insect stages of the parasite. (1) Metacyclic cells differentiate into long-slender BSF as they are released into the mammalian blood-stream. (2) Long-slender BSF cells differentiate to cell-cycle arrested short stumpy form cells. (3) Following a tsetse fly blood-meal, short stumpy cells differentiate to PCF cells. (4-7) PCF cells migrate through the proventriculus to the salivary glands, eventually differentiating into epimastigote form cells, which upon release from the salivary glands differentiate into metacyclic form cells. Reproduced from www.richardwheeler.net.

1.3 Evolutionary divergence of kinetoplastids

Trypanosome species belong to the class of kinetoplastid organisms, which are found in the Excavata super-group, one of five which divide the eukaryotic domain of life (Figure 1.3)(Adl et al., 2012; Lukeš et al., 2014). Kinetoplastids are defined through the presence of an organelle known as the kinetoplast – the dense DNA compartment of the single mitochondrion of the cell. This class of organism is thought to have diverged very early during eukaryotic evolution (Dacks et al., 2008; Lukeš et al., 2014) and demonstrates unusual biological processes in comparison to more commonly studied organisms such as yeast, mice and humans. For example, mitochondrial RNA

transcripts go through a complex process of RNA editing, glycolysis is compartmentalised into unique organelles and apart from two protein-coding genes, lack intronic DNA sequences (Simpson et al., 2006).

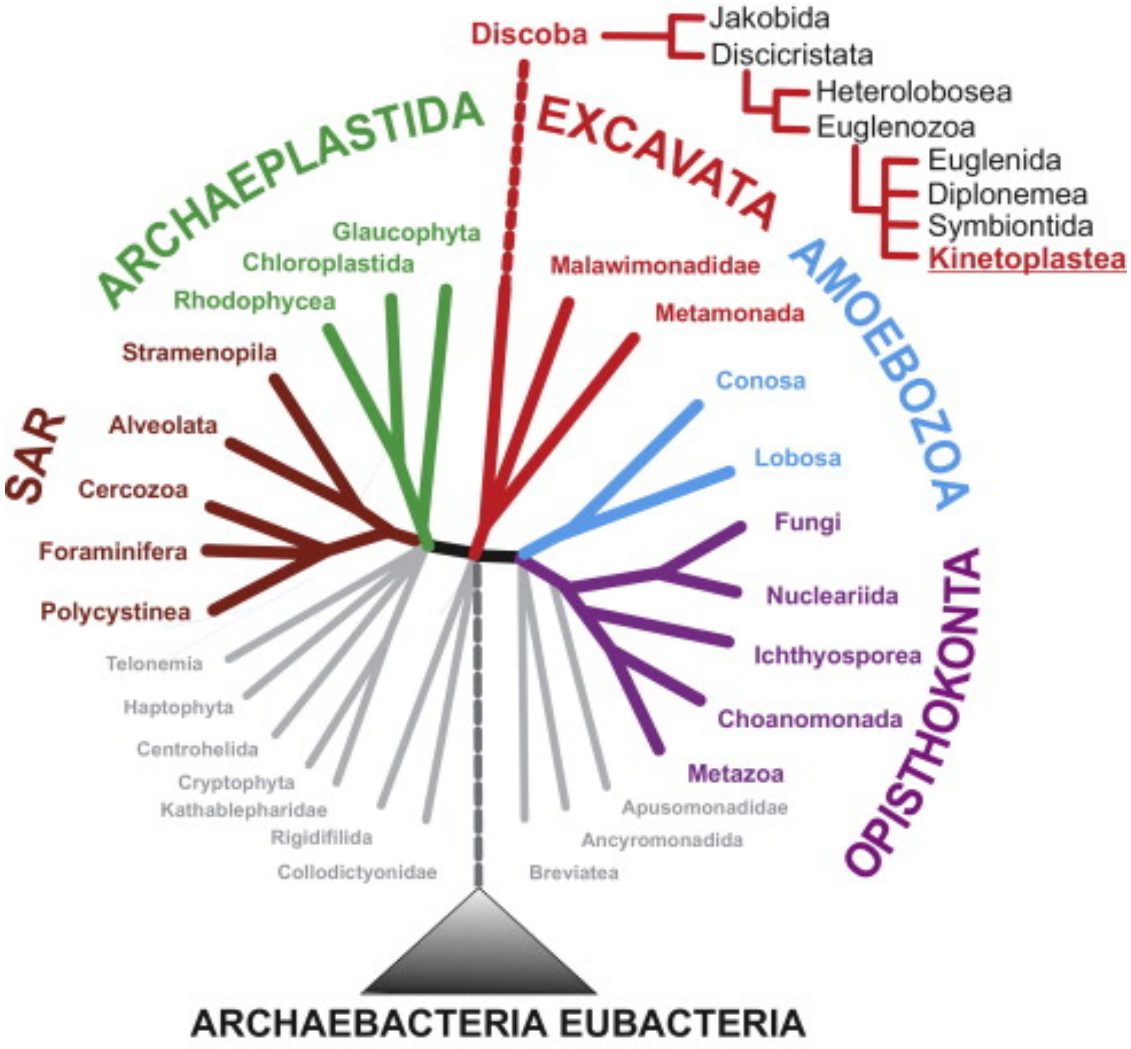


Figure 1.3: Phylogenetic tree of *Eukarya*. Eukaryotic species are divided into five super-groups, depicted as different colours of tree branches. SAR represents *Stramenopliles*, *Alvolates* and *Rhizaria*. Species in grey are not currently placed in any of the major super-groups. *T. brucei* is found in the *Excavata* super-group, within the *Kinetoplastea* class. Reproduced from *(Lukeš et al., 2014)*.

The genome organisation and mechanism of transcription in *T. brucei* further demonstrates the unusual biology of these parasites. The nuclear genome is composed of 11 diploid megabase-sized chromosomes (0.9-6Mbp), ~5 aneuploid intermediate-

sized chromosomes (150-900 Kbp) and ~100 mini-chromosomes (50-150 Kbp)(Berriman et al., 2005). The sub-telomeric region of megabase-sized chromosomes contain arrays of VSG genes and pseudogenes, though only one VSG is transcribed at any one time in the BSF of the parasite (Cross et al., 2014; Marcello and Barry, 2007). RNA polymerase I transcribes a single VSG gene, in conjunction with expression site associated genes, from a single bloodstream expression site that can be found on either megabase, or intermediate-sized chromosomes (Gunzl et al., 2003; Günzl et al., 2015). The intermediate-sized and minichromosomes also encode VSG genes that can be activated and expressed upon antigenic variation. VSG pseudogenes can also be involved in recombination events to produce new functional VSGs, providing a mechanism to create an endless repertoire of proteins crucial for immune evasion (Glover et al., 2013).

Most protein-coding genes in trypanosomatids are transcribed as long polycistronic units, with tens to hundreds of protein coding sequences produced in a single transcript (Imboden et al., 1987; Johnson et al., 1987). Genes within a polycistronic unit are arranged in a head to tail orientation, and distinct units are separated by strand switch regions, which can be oriented in a head to tail, convergent or divergent manner. Individual mRNAs are produced by co-transcriptional trans-splicing of a 39-nucleotide spliced leader mini-exon to the 5' untranslated region (UTR) of a gene, in combination with cleavage and polyadenylation of the 3' UTR (Figure 1.4)(Huang and Van der Ploeg, 1991; Ullu et al., 1993). Genes within a polycistronic unit do not necessarily share functional relationships and regulation of gene expression is thought to occur mainly at the post-transcriptional level. RNA binding proteins compose a major part of the trypanosomatid genome, and are known to play key roles in regulating the expression of mRNAs (Clayton, 2013; 2002). For example, a 3' UTR in VSG

transcripts promotes mRNA stability in BSF cells, while the same mRNA is rapidly degraded in PCF cells, ensuring VSGs are only expressed in the mammalian host (Berberof et al., 1995).

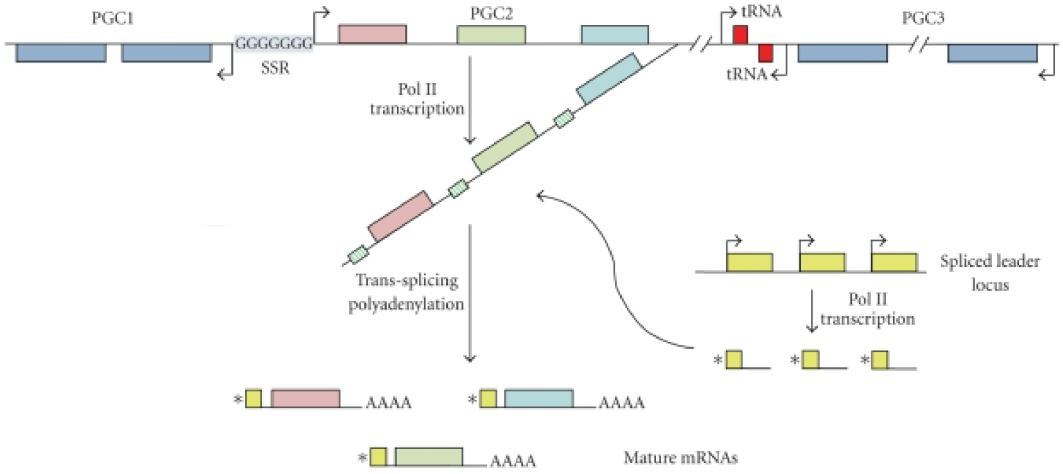


Figure 1.4: Overview of transcription in *T. brucei*. Diagrammatic representation of genes in *T. brucei*, transcribed from polycistronic gene clusters (PGCs). Strand switch regions (SSRs), characterized by long stretches of guanine (G) nucleotides, separate PGCs. Co-transcriptional trans-splicing, of a spliced leader sequence, and poly-adenylation resolves polycistronic mRNA transcripts into individual mRNA transcripts. Reproduced from *(Martínez-Calvillo et al., 2010)*.

Although a major cause of global disease, the large evolutionary divergence from intensively studied eukaryotic model organisms, means much of the unusual biology of trypanosomatids remains poorly characterised. In 2005 the genomic sequence of *T. brucei* was reported, identifying ~9,100 genes, providing a highly useful resource for the trypansomatid research community (Berriman et al., 2005). However, many of the identified genes lack classifiable homology to known proteins, hampering their functional classification. It has been previously estimated that 4,900 *T. brucei* genes have no reliable homologs in other organisms, and of the remaining 4,200 genes, 35% were annotated as “hypothetical”, lacking functional evidence in other organisms too

(Salavati and Najafabadi, 2010). In 2011 it was estimated that 64% of ~7,500 non-redundant genes were annotated as "hypothetical" (Alsford et al., 2011).

1.4 Global strategies to annotate gene function

To overcome the lack of functional annotation from genomic sequence, it is necessary to provide evidence for protein or gene function through a variety of high-throughput methods (Field et al., 2012).

It has been demonstrated that computational analysis of protein and gene sequences can allow for the prediction of function (Gazestani et al., 2014; Salavati and Najafabadi, 2010). For example, gene expression in *T. brucei* is heavily regulated at the post-transcriptional level through RNA binding proteins. The stability of mRNA is dependent on the sequence motifs to which these proteins bind, and genes of a functional class may share these regulatory sequences. It is possible to predict these regulatory elements within groups of proteins that are known to share a particular function, are involved in similar metabolic pathways, or display similar patterns of expression. Analysis of trypanosmatid genes within the same KEGG pathway predicted 15 and 21 function-specific motifs in 5' and 3' UTRs respectively (Mao et al., 2009). Together with analysis of gene codon-usage it was possible to predict components of metabolic pathways such as inositol phosphate metabolism, with a reasonable sensitivity (50%) and precision (>60%)(Salavati and Najafabadi, 2010). However, these methods are dependent on prior knowledge of function or co-regulation of a set of genes.

The availability of genomic data has also opened the door to transcriptomic and proteomic approaches to provide direct experimental evidence for the functional properties of trypanosomatid genes and proteins on a genome-wide scale. Omics

technologies are particularly useful in this scenario, as data is produced for many of the trypanosomatid specific "hypothetical" genes and proteins, which may begin to highlight their potential biological role, and direct hypothesis driven research to test potential functions.

Transcriptomic studies have been utilised to determine changes in mRNA abundance between life-cycle stages of the parasite and during time-courses of differentiation or cell-division (Archer et al., 2011; Brems et al., 2005; Jensen et al., 2009; Kabani et al., 2009; Koumandou et al., 2008; Siegel et al., 2010). These studies have led to the discovery of groups of transcripts which are up-regulated in different phases of the *T. brucei* life or cell cycle, together with the identification of RNA sequence motifs responsible for regulation of gene expression across cell-division (Archer et al., 2011). Transcriptomic datasets can produce very high coverage, with quantitative data from 70-90% of all annotated *T. brucei* genes in some experiments. However, the data are limited to information on mRNA abundance, whereas many other factors also account for gene function. Furthermore, as transcript regulation occurs mainly post-transcriptionally, and with the translation efficiency between individual transcripts varying by 100-fold, it is clear that mRNA abundance may not be representative of the final protein abundance in a cell (Jensen et al., 2014; Vasquez et al., 2014).

Proteomic analyses tend to have much lower coverage than transcriptomic studies, with the deepest data-sets reaching just over 4,000 proteins identified. However, a wider range of experimental methodologies can be employed to extract different functional information. Proteomic studies have also been used to perform comparisons of relative protein abundance between distinct life-cycle stages, or during a time-course of differentiation (Butter et al., 2013; Dejung et al., 2016; Gunasekera et al., 2012;

Urbaniak et al., 2012a), identifying novel factors such as a DEAD/H RNA helicase, and DOT1B, a histone deacetylase, necessary for cell division in BSF cells or differentiating BSF cells, respectively (Butter et al., 2013; Dejung et al., 2016).

In addition to quantifying the total abundance of proteins, proteomic methods can also identify and quantify post-translational modifications on proteins that further informs on function. Post-translational modifications (PTM) regulate protein function in a variety of ways that may either inhibit, or activate protein activity. PTMs are also crucial for signal transduction pathways that allow cells to respond and react to environmental conditions, without modifying total protein abundance. The simple identification of where a protein is post-translationally modified can define amino acids that may be critical for the regulation of protein function (Cohen, 2000; Nett et al., 2009). Comparative analysis of BSF and PCF parasites has identified phosphorylation sites that differ between the two biological states, and may be indicative of mechanisms that may establish and maintain these biological differences (Urbaniak et al., 2013).

Sub-cellular localisation is yet another factor that points towards the biological role of a protein. A large number of proteomic studies in *T. brucei* have focused on mapping proteins which localise to different organelles and structures within the cell, including the glycosome (Colasante et al., 2006; Güther et al., 2014; Parsons and Nielsen, 1990; Vertommen et al., 2008), flagellum (Broadhead et al., 2006; Hart et al., 2009; Lacomble et al., 2009; Oberholzer et al., 2011; Pullen et al., 2004; Subota et al., 2014), bi-lobed structure (Gheiratmand et al., 2013; Zhou et al., 2010), acidocalcisome (Huang et al., 2014), nucleus (DeGrasse et al., 2008; Rout and Field, 2001), mitochondrion (Panigrahi et al., 2009), and a range of lipid membranes (Acestor et al., 2009; Bridges et al., 2008; Niemann et al., 2012; Shimogawa et al., 2015).

Proteomics can also be utilised: to determine protein-protein interactions within complexes (section *1.5*); in combination with genetic approaches to determine proteome effects after knock-down or knock-out of a protein (Zoltner et al., 2015); or in artificial assays, identifying the mRNA (Lueong et al., 2016) or drug binding proteome (Urbaniak et al., 2012b) for example. The information produced for a protein from the wide range of proteomic experiments described here can be used to classify protein behaviour and group together proteins with similar behaviours and hence similar functions (Ohta et al., 2010).

Finally, the development of RNA interference target sequencing (RITseq) has produced a powerful genetic screening technology in *T. brucei* that can identify genes that are lethal, or promote survival under a particular treatment condition after knock-down (Alsford et al., 2011). Current publications have identified groups of genes which are essential for survival of *T. brucei* in culture (Alsford et al., 2011); have identified the targets of trypanocidal drugs (Alsford et al., 2012); or have identified components of the quorum sensing pathway in BSF *T. brucei* (Mony et al., 2014). The RIT-seq approach requires no *a priori* knowledge of proteins or genes that may be involved in a process of interest, so is another powerful tool to begin to annotate gene function on a genome wide scale.

With this in mind, the project plan for this thesis aimed to contribute to the unbiased genome wide annotation of gene function in *T. brucei*. I achieved this by providing experimental evidence for protein function using proteomic technologies to determine regulation of protein abundance across the cell-division cycle and to globally characterise protein-protein interactions.

1.5 Protein interactions

Many intracellular biological processes are dependent on the stable physical association between two or more proteins (Alberts, 1998). Indeed, many proteins require to form part of a complex to carry out their function, including many well characterised complexes such as the proteasome, ribosome and spliceosome. Thus, identifying which proteins are interacting can inform us of their likely function and associated biological process. The global characterisation of model organism interactomes has led to greater understanding of the proteome organisation and improved the functional annotation of uncharacterised proteins via guilt by association (Gavin et al., 2006; 2002; Ho et al., 2002; Hu et al., 2009; Krogan et al., 2006).

A number of proteomic methods have been developed to experimentally identify and characterise protein interactions in a large-scale, high throughput manner (Figure 1.5)(Gingras et al., 2007; Köcher and Superti-Furga, 2007; Mehta and Trinkle-Mulcahy, 2016; Smits and Vermeulen, 2016). Affinity purification in combination with mass spectrometry (AP-MS) is one of the most common methods for protein interaction mapping, at either a low or high-throughput scale. This method relies on high-quality antibodies to the endogenous protein being purified, or on genetic modification, to produce tagged proteins, that can be purified with a generic set of antibodies, or other probes, against the tag. A major strength of AP-MS is the ability to enrich for a protein of interest, allowing the characterisation of low abundance proteins and complexes. Furthermore, in genetically tractable systems it is possible to query the interactome in a high-throughput manner. Through the genetic tagging of thousands of proteins, studies

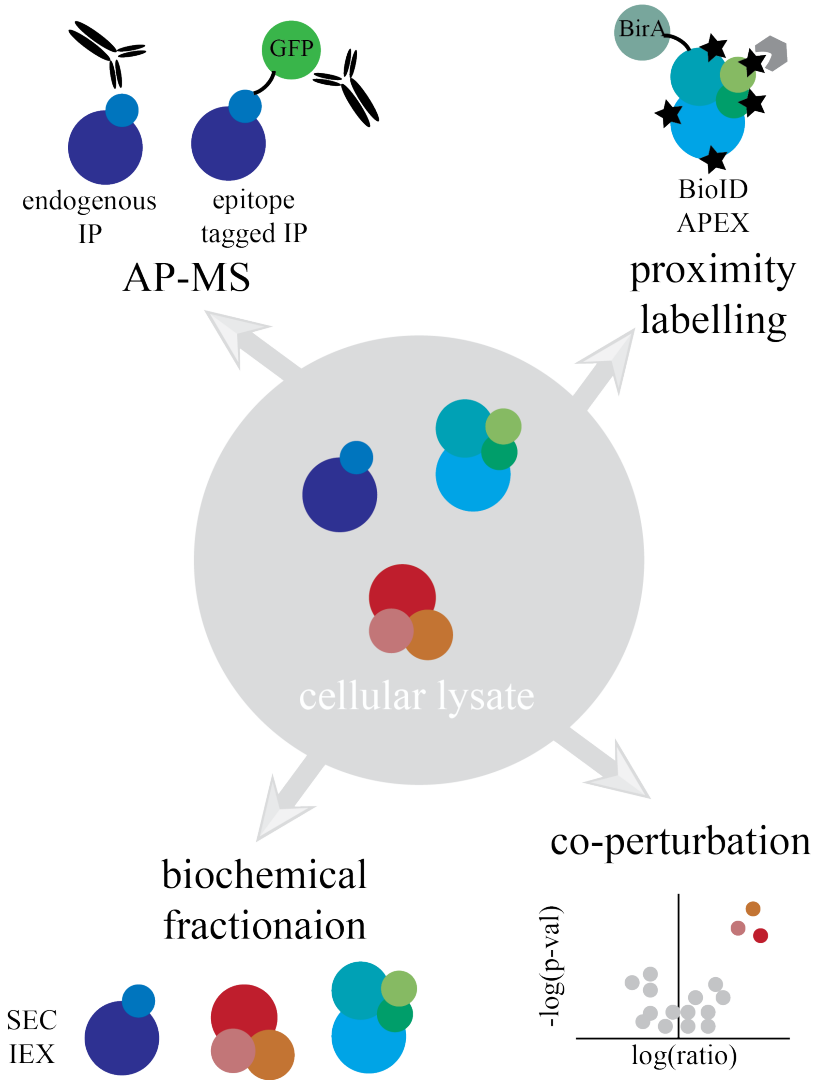


Figure 1.5: Methods for the characterisation of protein complexes. AP-MS can be used to purify endogenous or epitope tagged proteins using antibodies to identify components of protein complexes. Proximity labelling covalently modifies proteins within a 10-20 nm radius of a genetically tagged protein with biotin. Modified proteins can then be purified through streptavidin to identify interactors. Co-perturbation approaches rely on similar behaviour of proteins within a complex in proteomic experiments. Biochemical fractionation separates protein complexes based on properties such as charge and size/shape and attempts to predict protein interaction through co-fractionation of component proteins. Figure adapted from *(Smits and Vermeulen, 2016)*.

in yeast and human cell lines have been able to identify tens of thousands of protein interactions (Gavin et al., 2006; Hein et al., 2015; Huttlin et al., 2015; Krogan et al., 2006).

The lack of a library of high-quality antibodies against endogenous *T. brucei* proteins, and the high costs required to produce one, has meant that AP-MS studies have mainly utilised systems of genetic tagging to purify proteins and their associated binding partners. These methods have been a key tool for trypanosome biologists to identify and characterise individual protein complexes, including the spliceosome (Luz Ambrósio et al., 2009; Palfi et al., 2005), ATP synthase (Zikova et al., 2009) and mitochondrial ribosomes (Zikova et al., 2008), to name but a few.

Another method which has been used successfully in *T. brucei* is proximity labelling, which allowed the identification of proteins associated with the unique bi-lobe structure (Morriswood et al., 2013). This method genetically tags a protein with an enzyme that biotinylates neighbouring proteins within a 10 to 20 nm radius (Rhee et al., 2013; Roux et al., 2012). Biotinylated proteins can then be enriched using streptavidin. The *in vivo* labelling method may allow for the identification of transient or weak interacting proteins that may be disrupted upon cellular lysis. Proximity labelling also avoids the necessity of low-stringency purification strategies to minimise disruption of protein interactions required in AP-MS.

Alternatively, a number of studies have been able to identify proteins interacting within a complex through their co-behaviour across multiple experiments. This effect depends on the principle that if one member of a protein complex is disrupted or depleted, the other members of the complex will demonstrate a similar effect. Therefore, across multiple experiments the similar behaviour of a set of proteins

indicates a functional linkage between these proteins, and has been demonstrated to detect known protein complexes (Ohta et al., 2010).

Interactomes are highly dynamic and individual proteins may have completely different interaction partners in different biological conditions. Affinity purification based interactomes require the culturing of thousands of independent cell lines to profile protein interactions in one biological state, which is already costly and labour intensive. Furthermore, usually only one buffer condition is utilised, which is unlikely to be compatible with all protein-protein interactions. Affinity tagging can also lead to artefacts from either overexpression, or disruption of the endogenous function of the tagged protein (Banks et al., 2015; Boulon et al., 2010). AP-MS methods are therefore not amenable to the comparison of protein-protein interactions between different biological conditions on a proteome-wide scale.

The co-fractionation of assayable proteins (*i.e.* enzymes and proteins to which there are antibodies or other selective probes) by liquid chromatography and/or sedimentation methods has long been used to provide compositional data on sub-cellular organelles and protein complexes. Conventionally these methods have been used to isolate individual protein complexes or biological structures. For example, to define components of the proteasome (Wang et al., 2003) or sub-nuclear compartments (Rout and Field, 2001) in trypanosomes.

Applying modern shot-gun proteomics to such a chromatographic and/or density sedimentation fractions from cell lysates provides a more general protein detection and quantification mechanism. This, in turn, allows the simultaneous recording of the separation profiles of several thousands of proteins (independent of specific assays) and these data can be used to suggest physical associations between them. This method has

come to be termed PCP-MS (protein correlation profiling-MS) and was initially utilised to analyse sub-cellular organelles (Andersen et al., 2003; Dunkley et al., 2004; Foster et al., 2006). More recently PCP-MS has been used to analyse soluble protein complexes through fractionation of whole cell lysates by size-exclusion chromatography (SEC)(Kirkwood et al., 2013; Kristensen et al., 2012; Li and Giometti, 2007; Olinares et al., 2010) and ion-exchange chromatography (IEX)(Havugimana et al., 2012), identifying hundreds of protein complexes in single experiments.

One of the major benefits of PCP-MS methods is the applicability to any cultivatable biological system, without genetic manipulation, and the ability to compare different biological conditions in a simple manner. This makes PCP-MS an ideal method to take forward to carry out global protein interaction analysis in understudied organisms such as *T. brucei*, with the potential to produce a wealth of useful functional information on hundreds of uncharacterised proteins. Indeed, during the course of this thesis, work from Gazestani and colleagues was published, using glycerol gradient centrifugation and ion-exchange chromatography, to identify protein complexes in PCF *T.brucei* mitochondrial and cytoplasmic lysates (Gazestani et al., 2016).

1.6 Cell cycle

The eukaryotic cell cycle is an evolutionarily conserved process in which a cell duplicates and segregates newly synthesised cellular components to produce two daughter cells from a single mother cell. The activation and inactivation of key regulatory proteins plays an important role in controlling the temporal order of events that must occur for cell division to proceed correctly. The key events of cell division include DNA replication (S-phase) and segregation of replicated DNA (mitosis/M-phase), preceded and interceded by two ‘gap’ phases (G1 and G2-phases) where cells sense environmental conditions prior to commitment to cell division, or assess

completion of DNA replication prior to entry into mitosis, respectively. These events must occur in order and must occur only once per cell cycle (Elledge, 1996).

Although *T. brucei* shares this basic eukaryotic cell cycle regulatory scheme, it is clear there are some trypanosome-specific mechanisms of regulation. Trypanosomes contain a number of unique single copy organelles such as the basal body, the flagella, the mitochondrion and the kinetoplast (mitochondrial DNA network) that must be duplicated and segregated equally to produce viable daughter cells. In the initial stages of G1, trypanosomes contain a single kinetoplast and nucleus (a 1K1N cell). Duplication of the basal body is the first morphological event that can be observed during the cell cycle and begins the nucleation of a new flagellum. The basal body is closely associated with the kinetoplast and its segregation drives duplication and segregation of kinetoplast DNA, producing a 2K1N cell. Following completion of nuclear DNA replication, cells enter M-phase, segregating replicated chromosomes to daughter nuclei (2K2N), followed by cleavage furrow ingression longitudinally down the axis of the cell to produce two daughters (Figure 1.6)(Hammarton, 2007; Li, 2012; McKean, 2003).

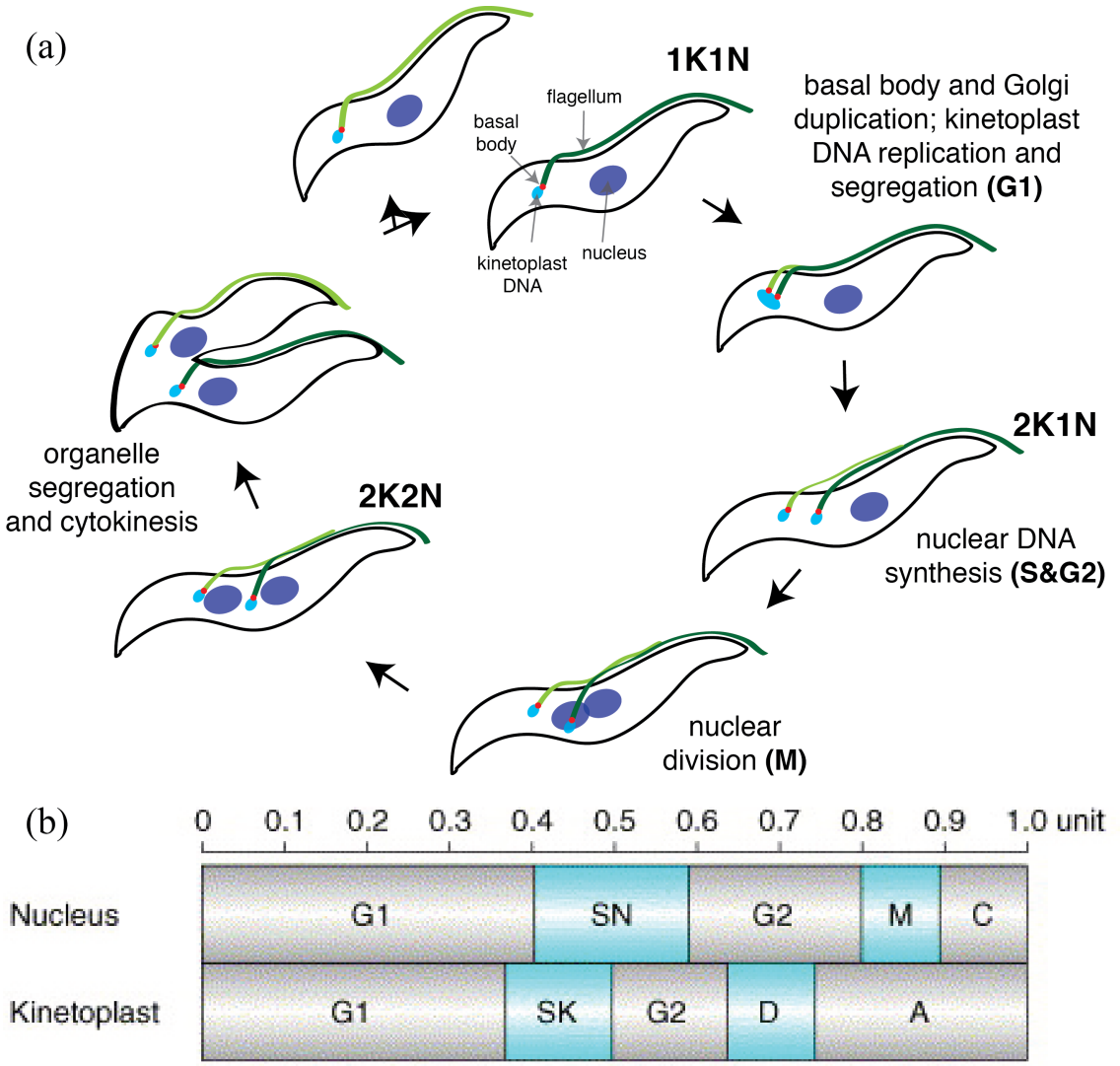


Figure 1.6: Overview of the procyclic *T. brucei* cell-division-cycle. (a) Cells begin with 1 kinetoplast and 1 nucleus (1K1N). Initial stages of division include basal body duplication, initiation of flagellar duplication and Golgi duplication. Segregation of the basal body is closely followed by replication and segregation of kinetoplast DNA, producing 2K1N cells. Following nuclear DNA replication and segregation in mitosis (2K2N) cells separate duplicated organelles and divide into two daughter cells. Images adapated from *(Akiyoshi and Gull, 2013)*. (b) Relative timing of the distinct but co-ordinated replication and segregation of the nuclear and kinetoplast DNA compartments. D stands for division of kinetoplast DNA and A for apportioning, where basal bodies continue to move apart. Reproduced from *(McKean, 2003)*.

The transition between distinct phases of the cell-cycle is controlled by regulatory proteins, several of which have been studied intensively in eukaryotic model organisms. Many of these regulatory proteins and their functions are highly conserved within eukaryotic species, including trypanosomes. Cyclin-dependent kinases (CDKs) play a major role in committing cells to entry into S-phase from a G1 state, and entry into M-phase from a G2-state. CDK activity is controlled through interaction with partner cyclin proteins, expressed in a cyclical manner during different phases of cell division, and through post-translational modification, such as phosphorylation. These mechanisms allow strict temporal control of cell division processes and allow for the input of signal transduction pathways to promote or inhibit progression of the cell cycle, depending on the internal and external cellular environment (Murray, 2004). The *T. brucei* genome contains 10 cyclins and 11 Cdc2-related kinases (CRKs) and experimental evidence indicates different pairs of cyclins and CRKs are necessary for transition between G1/S and G2/M-phases in trypanosomes, as found in other eukaryotes (Hammarton et al., 2004; Li and Wang, 2003; Tu and Wang, 2004; 2005).

Although some proteins are conserved, they may have trypanosome specific functions. Aurora B kinase (ABK) forms interactions with three evolutionarily conserved proteins (INCENP, Borealin and Survivin) to form the Chromosomal Passenger Complex (CPC) in yeast and metazoa (Carmena et al., 2012). The ABK interaction partners are involved in the dynamic localisation of the CPC during mitosis where it relocates from chromatin, to the mitotic spindle and mid-body of dividing cells. *T. brucei* contains a conserved ABK orthologue, though none of the partner CPC proteins have been identified. Instead, trypanosome ABK interacts with trypanosome specific proteins (CPC1 and CPC2) that show similar dynamic patterns of localisation

as seen in other eukaryotes from chromatin, to the mitotic spindle and the flagellar basal body (Li et al., 2008a).

Some of the major components of the cell division machinery in trypanosomes have only been discovered recently, or have yet to be discovered. The kinetochore is a macromolecular structure composed of hundreds of proteins that forms on the centromeres of chromosomes during mitosis, ensuring correct chromosome segregation to daughter cells (Akiyoshi and Gull, 2013; Santaguida and Musacchio, 2009; Wittmann et al., 2001). Core kinetochore proteins were expected to be conserved across all eukaryotes, although none had been identified in kinetoplastid species until work by Akiyoshi and Gull (Akiyoshi and Gull, 2014). Through fluorescent tagging of uncharacterised proteins in *T. brucei* that were upregulated in G2&M-phase in transcriptomic studies (Archer et al., 2011), a single protein was identified with localisation patterns indicative of a kinetochore component. Iterative immuno-precipitation experiments eventually identified nineteen novel trypanosome kinetochore proteins, conserved among kinetoplastids, but with insufficient homology to have been identified from BLAST searches to kinetochore proteins found in other organisms.

The use of cell cycle regulated transcriptomic data in the identification of a completely novel kinetochore complex demonstrates the power of large-scale datasets to provide information on the function of uncharacterised proteins. Previous transcriptomic analysis has comprehensively determined the cell cycle regulated patterns of mRNA abundance in trypanosomes. However, it is not necessarily the case that RNA abundance directly correlates to protein abundance. Discrepancies in this equation can occur due to differences in rates of translation of mRNA, or rates of degradation of protein. Furthermore, in an organism that controls gene expression

through RNA binding proteins, the transcriptome may not be an accurate representation of the proteome (Clayton, 2013; 2002; Jensen et al., 2014; Vasquez et al., 2014).

I therefore decided to carry out a proteomic analysis of cell division in PCF trypanosomes. Our aim was to identify cell-cycle regulated proteins, characterise the temporal regulation of thousands of uncharacterised proteins and to identify key proteins which may be useful drug targets involved in key aspects of trypanosome biology. The data produced in this thesis will provide a strong platform to further understand the biology of cell-division in this unusual eukaryotic parasite.

Chapter 2: Materials and methods

2.1 SDM-79 media preparation

Powdered SDM-79 media was hydrated with 5 L of Milli-Q water, and supplemented with haemin to 7.5 mg/L (from a 10 mg/mL stock in 0.1 M NaOH) and 2 g/L of sodium bicarbonate. The pH was adjusted to 7.3 with NaOH, and sterile filtered using Stericups 500. Under sterile conditions, heat inactivated and non-dialysed fetal bovine serum (PAA) was added to final 15% (v/v) and Glutamax I to 2 mM final concentration. The antibiotics, G418 and hygromycin, were added to 15 μg/mL and 50 μg/mL respectively.

2.2 Cell culture

Procyclic trypanosomes (clone 29.13.6) were cultured in SDM-79 media at 28°C, without CO_2, in fully capped culture flasks.

2.3 Size-exclusion chromatography (SEC)

Procyclic trypanosomes ($3x10^9$ cells/replicate) were washed three times in 50 mL of phosphate-buffered saline (PBS) and lysed using a Bioruptor Pico (Diagenode) water bath sonicator for 10 cycles of 30 seconds on/off, in 0.75 mL of PBS containing 0.1 μM 1-5-chloro-3-tosylamido-7-amino-2-heptone (TLCK), 1 mM phenyl-methyl sulfonyl fluoride (PMSF), 1 μg/mL leupeptin, 1 μg/mL pepstatin and 5 mM ethylenediametetraacetic acid (EDTA). Lysates were spun at 17,000 g for 10 min, and the supernatant filtered through a 0.45 μm filter unit, all at 4°C. Bradford assays were performed on the filtrates for protein quantitation.

Filtered lysates (~1.2 mg in 200 μL, ~$8x10^8$ cell equivalents) were injected onto a BioBasic SEC 300 or a BioBasic SEC 1000 column (30 nm or 100 nm pore size respectively), using a Dionex Ultimate 3000 high-performance liquid chromatography

(HPLC) system, and collected into 48 fractions of 120 µL. Columns were equilibrated with PBS and run at 0.3 mL/min at 4°C. Protein concentrations of collected fractions were determined using Bradford assays.

Each fraction was made up to 0.1 M Tris-HCl (pH 8.0), 1 M urea and 5 mM dithiothreitol and incubated for 2 h at 37°C, followed by addition of iodoacetamide at a final concentration of 25 mM at room temperature for 1 h. Trypsin and LysC were added, each at a ratio of 1:100 to total average protein per fraction and incubated overnight at 37°C. Each fraction was made up to 1% (v/v) trifluoroacetic acid (TFA) and desalted using Sep-Pak tC18 plates, with peptides eluted in 50% acetonitrile, 0.1% TFA. Peptides were dried using a GeneVac evaporator and resuspended in 5% formic acid and quantified using a 3-(4-carboxybenzoyl) quinoline-2-carboxaldehyde assay (CBQCA). Five biological replicates were performed in total for each chromatography column.

For a highly denaturing lysis, cells were lysed in 4% sodium dodecyl sulphate (SDS), 10 mM sodium phosphate (pH 6.0) and 100 mM NaCl, sonicated and filtered as above, followed by separation in a running buffer containing 0.2% SDS on a BioBasic SEC300 column.

2.4 Strong anion exchange (SAX)

Procyclic trypanosome cells were prepared in a similar manner as described for SEC analysis, with lysis in 1mL 20 mM ethanolamine (pH 9.0) containing 0.1 µM TLCK, 1 mM PMSF, 1 µg/mL leupeptin, 1 µg/mL pepstatin and 5 mM EDTA. Lysates were spun and filtered as described previously.

Filtered lysate was injected onto a Protein-Pak Hi Res Q, 5 µm, 4.6 x 100 mm, column (Waters), equilibrated in 20 mM ethanolamine (pH 9.0). Proteins were resolved

over a gradient of 0 to 100 % 0.5 M NaCl in 20 mM ethanolamine (pH9.0) over the course of 26 min, at a flow rate of 0.3 mL/min at 5°C. 96 fractions, of 105 µL, were collected from 1.5 to 35 min.

Collected fractions were made up to 4% SDS and 25 mM tris(2-carboxyethyl)phosphine (TCEP), then heated to 65°C for 30 min. Once samples had cooled to room temperature, N-ethylmaleimide (NEM) was added to a final concentration of 50 mM, and incubated for 1 h.

The denatured, reduced and alkylated proteins in each fraction were prepared for digestion utilising a Kingfisher Flex Purification System (ThermoFisher Scientific) in combination with magnetic SP3 beads. 20 µL of a 1:1 mixture of hydrophobic and hydrophilic, carboxylate modified, Sera-Mag SpeedBead magnetic particles (20 mg/mL in H_2O, GE) was added to each fraction, followed by the addition of 500 µL of acetonitrile and 15 µL of 10% formic acid. In a 96-well plate format, the Kingfisher Flex System, was then utilised to wash the magnetic beads (protein bound) for each collected fraction, twice in 1 mL 70% ethanol, once in 1 mL 100% acetonitrile, and then released into a pre-cooled plate, containing 50 µL 0.1 M Tris-HCl (pH 8.0), 0.1% SDS, 1 mM $CaCl_2$ and trypsin and LysC at a 1:100 ratio of protease to estimated protein per fraction. The plate was incubated overnight at 37°C at 500 rpm in a ThermoMixer (Eppendorf). Following overnight digestion, the 96-well plate was thoroughly vortexed to ensure resuspension of SeraMag beads, and 950 µL of acetonitrile added. Peptides bound to the magnetic beads were washed in 1 mL of acetonitrile, eluted in 40 µL of 2% dimethylsulfoxide (DMSO) and beads removed from the sample again on the Kingfisher System. Formic acid was added to each sample to a final concentration of 5%.

2.5 SDS-PAGE and Western blotting

Aliquots of fractions from SEC runs were pooled into groups of three, made up to 1x Lithium Dodecyl Sulfate (LDS) Sample Loading Buffer and 25 mM TCEP, heated to 95°C for 10 min, and resolved on 4-12% SDS-PAGE. Gels were run in MES buffer for 45 min at 200 V, and were either stained for total protein using SYPRO Ruby, as per manufacturers protocol, or transferred onto nitrocellulose membranes at 35 V for 90 min.

Membranes were blocked in 3% skimmed milk and 0.1% Tween20 in PBS for 1 h at room temperature and subsequently probed with primary antibody overnight at 4°C in Stericup filtered 5% bovine serum albumin (BSA), 0.1% Tween20 and 0.02% sodum azide in PBS. The primary antibodies were utilised at the following dilutions: rabbit anti-enolase, 1:3000; rat anti-tryparedoxin peroxidase, 1:500; mouse anti-beta-tubulin, 1:2000. Membranes were then washed three times in 0.1% Tween20 in PBS at room temperature. Secondary anti-rat, rabbit and mouse antibodies, conjutgated to 800 and/or 680 IRDyes, were used at 1: 15,000 dilutions in 3% skimmed milk and 0.1% Tween20 in PBS and incubated with membranes for 1 h at room temperature. Membranes were washed five times in 0.1% Tween20 in PBS and three times in PBS prior to imaging on a LiCOR Odyssey CLx imager.

2.6 LC-MS/MS and analysis of spectra

For each biological replicate of 48 SEC or 96 SAX fractions, 1µg of peptide was injected from the most concentrated fraction, and the equivalent volume injected for the remaining fractions. Peptides in 5% formic acid were injected onto a C18 nano-trap column using a Thermo Scientific Ultimate 3000 nanoHPLC system. Peptides were washed with 2% acetonitrile, 0.1% formic acid and resolved on a 150 mm x 75 µm C18 reverse phase analytical column over a gradient from 2% to 28% acetonitrile over 120

min at a flow rate of 200 nL/min. Peptides were ionised by nano-electrospray ionisation at 2.5 kV. Tandem mass spectrometry analysis was carried out on a QExactive+ mass spectrometer, using HCD fragmentation of precursor peptides. A data-dependent method was utilised, acquiring MS/MS spectra for the top 15 most abundant precursor ions.

SEC RAW data files were analysed using MaxQuant version 1.5.1.3, supplied with the *T. brucei brucei* 927 annotated protein database from TriTrypsDB release 8.1, containing 11,567 entries. The mass tolerance was set to 4.5 ppm for precursor ions and MS/MS mass tolerance was set at 20 ppm. The enzyme was set to trypsin and endopeptidase LysC, allowing up to 2 missed cleavages. Carbamidomethyl on cysteine was set as a fixed modification. Acetylation of protein N-termini, deamidation of asparagine and glutamine, pyro-glutamate (with N-terminal glutamine), oxidation of methionine and phosphorylation of serine, threonine and tyrosine were set as variable modifications. Match between runs was enabled, allowing transfer of peptide identifications of sequenced peptides from one LC-MS run to non-sequenced ions, with the same mass and retention time, in another run. A 20-min time window was set for alignment of separate LC-MS runs and a 30-second time window for matching of identifications. The false-discovery rate for protein and peptide level identifications was set at 1%, using a target-decoy based strategy. Each individual SEC fraction was set as an individual experiment in MaxQuant parameters, to output IBAQ data for protein groups in every fraction, and only unique peptides were utilised for quantitation.

SAX RAW data files were analysed using MaxQuant version 1.5.3.30, supplied with the *T. brucei brucei* 927 annotated protein database from TriTrypDB release 26.0, also containing 11,567 entries. All other settings were identical, apart from the fixed modification on cysteine, which was set to NEM.

2.7 Data analysis of protein elution profiles

Data analysis was performed using custom scripts in Python language using numpy, scikit-learn, pandas and matplotlib libraries. Elution profiles for individual proteins were created using the following MaxQuant abundance metrics: label free quantitation (LFQ) intensity, intensity based absolute quantitation (IBAQ), MS/MS count and unique peptide count. Profiles produced from each metric were normalised by the maximum value detected, and used to determine the Pearson correlation of proteins between biological replicates. All further data analysis utilised maximum normalised LFQ intensities, using the mean of four or three biological replicates from SEC300 and SEC1000 experiments. From SEC experiments, proteins were required to be detected with at least one unique peptide found in two biological replicates and with Pearson correlation coefficients between elution profiles >0.6.

2.8 Hierarchical clustering

The mean LFQ profiles for each protein were hierarchically clustered, separately for each experiment type (SEC300, SEC1000 and SAX), using the Euclidean distance measurement and Ward's agglomeration method. The Gene Ontology (GO) term enrichment was computed for each cluster obtained by cutting the dendrogram tree at predetermined distances. Cutting distances from 0 to *n* were evaluated, where *n* was the cutting distance producing only two clusters. GO term enrichment p-values were computed with a Fisher test. The Bonferroni correction was applied and only the GO-terms with a p-value <0.05 were accepted. The cutting distance producing the highest number of enriched GO terms was taken to produce the final clusters for each experiment type.

2.9 Peak picking

For the analysis of the observed against the expected molecular weight, the LFQ elution profiles of each protein detected in the SEC300 experiment were split into individual peaks by adapting a Ricker wavelet encompassing 2 to 4 fractions. The two minimum points of the wavelet were used to define the peak range, and any other LFQ values outside this range were set to 0. Several filters were applied before considering peaks further. Firstly, individual peaks where the maximum intensity was below 20% of the most intense peak across the entire protein profile were filtered. Secondly, peaks lying in a region of the fractionation profile corresponding to, or less than, the protein dimer molecular weight (as annotated in TriTrypDB) or present in the void (fractions 1 to 6) were discarded. Observed molecular weights were calculated from a linear regression analysis of the elution of molecular weight standards. GO term enrichment analysis was performed for each category as described in section *2.8*, and plots were produced using ReviGO (Supek et al., 2011).

2.10 Machine learning

A pipeline similar to that applied previously for PCP analysis was utilised (Havugimana et al., 2012) to predict protein complexes using data from all three experiment types. The protein elution profiles were used to train a random forest predictor implemented with the scikit-learn python package. Protein pairs were scored according to four features, namely: the co-apex score, Normalized Cross Correlation (NCC), Pearson Correlation Coefficient (PCC) and STRING scores. The first three features are based purely from the protein elution profiles.

The co-apex score was used by Havugimana and colleagues (Havugimana et al., 2012), and is based on the number of biological replicates in which the protein pairs showed maximum abundance in the same fraction. The co-apex score was derived for

the SEC300 and SEC1000 experiments with four and three biological replicates respectively by using the scipy package. For the SEC1000, with three biological replicates, the possible co-apex scores were: 1 (3 of 3 replicates), 0.6 (2 of 3 replicates), 0.3 (1 of 3 replicates) and 0 (none of the replicates). Similarly, For the SEC300, with four biological replicates, the possible co-apex scores were 1, 0.75, 0.5, 0.25 and 0.

The NCC was derived in two steps. First, the maximum cross correlation between the two protein profile pairs $P_{1\text{-}2}CC$ was computed. Then the maximum self-cross-correlation of the first protein profile (P_1CC) and the maximum self-cross-correlation of the second protein profile (P_2CC) was determined. The NCC was finally derived as $P_{1\text{-}2}CC$ / max(P_1CC, P_2CC).

The PCC was computed as the Pearson correlation score between the two elution profiles.

The PCC and NCC was calculated for the SEC300, SEC1000 and SAX experiments described here, and for experiments produced in (Gazestani et al., 2016). This includes ion exchange of mitochondrial extracts (IEX-mito) and cytoplasmic extracts (IEX-cyto) and glycerol gradient fractionation of whole cell lysates (GG-WCL) and mitochondrial extracts (GG-mito). The STRING features (Neighborhood, Fusion, Cooccurence, Coexpression, Experimental, Database, Text Mining) are derived from version 10 of the STRING database (Szklarczyk et al., 2015). The STRING IDs were mapped to the TriTrypDB IDs, and the values were normalized from 0 to 1. The aforementioned scoring features were calculated for all the possible permutations of protein pairs that showed a NCC value greater than 0.15 in at least one of the SEC300, SEC1000 or SAX experiments, creating a matrix of 609,100 protein pairs with 23 features.

For the machine learning, a dataset of 'gold standard' true positive peak pairs (GD) was manually assembled. Thirty-one theoretical protein complexes were derived from data deposited in CORUM (Ruepp et al., 2007), together with manual addition of protein complexes derived from literature information, producing 293 unique true positive pairs of interacting proteins. A negative dataset was extracted by random sampling of the 293 proteins annotated in different complexes. As it would be possible to introduce false negative interactions in this step, the random sampling was repeated 100 times. Finally, using these true positive and true negative test pairs 100 Random Forest classifiers were assembled based on the same true positive pairs, but with each using a different negative set. Two predictor sets were developed, one set of 100 predictors based on the features derived only from experiments performed in this thesis (SEC300, SEC1000 and SAX) and a second set of predictors that implemented all the available features (including STRING, IEX-mito, IEX-cyto, GG-mito and GG-WCL).

All the classifiers were inspected to determine the area under the curve values of the receiver operator curves in ten-fold cross validation. The median values of the probability score outputs of the two 100 classifier sets were used as the final interaction prediction score for the protein pairs.

An interaction prediction score cut-off of 0.75 was selected, and separately fed the protein pairs from the two predictor sets to the ClusterONE algorithm. A search matrix was created for the ClusterONE program with the parameters 'd' (0.1 to 1, step 0.1), 'haircut' (0.1 to 1, step 0.1) and 's' fixed to 2. The outputs were parsed to derive the parameters that were optimal to obtain the maximum number of GD true positive pairs grouped together. The complexes predicted by ClusterONE using the outputs of the two different predictor sets were merged together, joining predicted complexes that shared two or more proteins.

2.11 In vivo chemical crosslinking

Procyclic *T. brucei* cells were used ($6x10^7$ cells/condition) to assess the optimal conditions for *in vivo* chemical crosslinking with either formaldehyde or dithiobis(succinimidyl propionate) (DSP). Cells were washed twice in 15 mL of PBS prior to resuspension in PBS containing the appropriate crosslinker concentration.

DSP was prepared as a 20 mM stock in DMSO and diluted in PBS to working concentrations between 0.125-2 mM. Samples were crosslinked for 30 min at room temperature with gentle agitation on a roller. Crosslinking was quenched by the addition of Tris-HCl (pH 7.5) to 20 mM at room temperature for 15 min Cells were pelleted, washed with PBS and lysed in 4% SDS, 10 mM sodium phosphate (pH 6.0) and 100 mM NaCl. Lysates were heated to 95°C for 10 min and sonicated for 10 cycles of 30 seconds on/off in a Bioruptor Pico (Diagenode) water bath sonicator. Lysates were split into two samples each, and mixed with 4x LDS sample buffer, one with the addition of 5 mM TCEP, and heated to 95°C for ten min prior to SDS-PAGE and/or Western blotting as described previously.

Methanol-free aliquots of 16% (w/v) formaldehyde solution were used to prepare PBS ranging from 0.2-6% formaldehyde. Samples were crosslinked as above, and quenched by pelleting cells and resuspending in 250 mM L-glycine for 15 min at room temperature. Cells were washed and lysed as above, though only heated to 37°C for 30 min prior to sonication. Samples were mixed with 4x LDS sample buffer and split in half and either heated to 65°C or 95°C for ten min prior to SDS-PAGE and/or Western blotting.

Sonicated lysates were filtered through a 0.45 μm filter unit before loading onto a BioBasic SEC 1000 column under denaturing conditions already described. Fractions were not collected.

2.12 Direct elutriation

Procyclic cells ($2.7x10^9$) were harvested from 100 mL of a log-phase culture by centrifugation and resuspended in 10 mL of elutriation buffer (a 1:4 dilution of SDM-79). Cells were passed two times through a 20-gauge needle and injected into a Sanderson loading chamber of an Avanti J-26 XP elutriation centrifuge equipped with JE5.0 rotor at a temperature of 28°C. Cells were loaded at a flow rate of 10 mL/min. The rotor was kept at a constant speed of 5,000 rpm. Fractions of 50 mL were collected at each flow rate of 10, 15, 18, 20, 23, 25, 27, 28, 29, 31, 32, 33 and 35 mL/min. The final fraction was collected at 35 mL/min with the rotor turned off. Aliquots were taken from each collected fraction for flow cytometry analysis, and the remaining cell pellets were snap frozen.

2.13 Single and double-cut elutriation

Cells were prepared in a similar manner as described in section *2.12*. Instead of collecting multiple fractions over a range of flow rates, cells were collected at two flow rates – 15 mL/min (small cells) and 32 mL/min (large cells). Both collected cell populations were pelleted by centrifugation and resuspended in SDM-79 at a concentration of $3x10^7$ cells/mL and recultured.

The small cell population was used for the single-cut time-course and aliquots were taken for flow cytometry at 0.5, 3, 4, 5, 6, 7, 8, 9, 10 and 11 h after elutriation.

The large cell population was cultured for 1 h and elutriated again, splitting the cells into the newly divided small cells, that were recultured in SDM-79, and large cell that

were discarded. This second small cell population was used for the double-cut time-course and aliquots were taken at the same intervals as the single-cut samples.

2.14 Flow cytometry

Cells were washed three times in 5 mL of PBS, fixed in 1 mL 70% ice-cold ethanol and stored at -20°C until ready for DNA staining for flow cytometry. Fixed cells were washed with 1 mL of PBS and resuspended in staining solution composed of 50 µg/mL propidium iodide, 100 µg/mL ribonuclease A, 0.5% (w/v) Triton-X100 and 0.5% bovine serum albumin in PBS. Cells were incubated in the dark at room temperature for a minimum of 20 min. Propidium iodide fluorescence was detected from 10,000 cells per sample on an LSR Fortessa.

2.15 Tandem Mass Tag (TMT) Labelling

For TMT labelling, ~100 µg of peptide was solubilised in 100 µL of 50 mM HEPES (pH 8.5) and mixed with the appropriate TMT labelling reagent. Aliquots (0.8 mg) of each reagent were solubilised in 41 µL of anhydrous acetonitrile. Peptide and label were incubated together for 2 h at room temperature. The reaction was quenched through the addition of 8 µL of 5% hydroxylamine for 15 min at room temperature. This reaction was scaled up or down dependent on the amount of peptide being labelled.

TMT 6-plex and 10-plex labelled peptides were mixed at this stage. Samples were acidified through the addition of TFA to a concentration of 1% and loaded onto 200 mg SepPak cartridges (Waters) that had been wetted with 100% acetonitrile and equilibrated with 0.1% TFA. Labelled peptides were washed with 0.1% TFA, eluted in 1 mL of 80% acetonitrile and 0.1% TFA and dried using a GeneVac evaporator. Peptides were solubilised in 5% formic acid for injection onto a mass spectrometer or in

2% acetonitrile in 10 mM ammonium formate (pH 9.0) for high-pH reverse phase chromatography.

2.16 High-pH reverse phase chromatography

Tryptic procyclic *T. brucei* peptides (2 mg) were labelled with TMT-zero labelling reagent and solubilised in 1 mL of 2% acetonitrile in 10 mM ammonium formate (pH 9.0). TMT labelled or non-TMT labelled peptides (200 µg) were injected onto an Xbridge BEH C18 column (130 Å, 3.5 µm, 4.6 x 150 mm), using a Dionex Ultimate 3000 HPLC system. Buffer A was composed of 2% acetonitrile in 10 mM ammonium formate (pH 9.0) and buffer B of 80% acetonitrile in 10 mM ammonium formate (pH 9.0). Columns were run at 1 mL/min at 30°C and equilibrated with ten column volumes to the composition of buffer A and B from the start of each respective run. A number of different gradients were used, starting at either 0, 10, 25 or 35% buffer B, rising to 60% B over the course of a linear gradient from 0 to 11 min. There was a spike to 100% buffer B from 11 to 12 min followed by a drop back to the starting percentage B from 12 to 13 min until the end of the run at 20 min.

2.17 Optimisation of TMT quantitation

Peptides from human HEK293 cells and procyclic *T. brucei* cells were labelled with 6-plex TMT reagents. Human peptides were labelled with all six reagents (126-131) and mixed in a ratio of 50:20:5:5:20:50 µg for the respective TMT labels. Trypanosome peptides were only labelled with the 126, 127 and 128 reagents and mixed with the human peptides in a ratio of 50:50:50 µg. After desalting and dehydrating the mixed sample, peptides were resuspended in 5% formic acid.

Unless otherwise specified, 1 µg of human-trypanosome peptide mix was injected onto QExative+ or Fusion mass spectrometers for each run. The chromatography methods employed were identical to those described in section *2.6*.

The QExactive+ survey scan was performed across the range from 300-1400 *m/z* at a resolution of 70,000. An Automatic Gain Control (AGC) target of $1x10^6$ and a maximum injection time (maxIT) of 20 ms was set for MS1 survey scans. The top 10 most intense ions with an isotopic envelope and a charge state between 2-7 were selected for fragmentation using Higher-Energy Collisional Dissociation (HCD) using a stepped normalised collision energy of 28, 35 and 42%. MS2 scans were acquired from 120-1050 *m/z* at a resolution of 70,000. The AGC target was set to $2x10^5$, with a maxIT of 250 ms. The isolation width for selection of precursor ions for fragmentation was varied between 0.4 – 4.0 *m/z*. Ions selected for fragmentation were excluded from further selection for 40 s.

Survey scans carried out on the Fusion mass spectrometer were carried out on the Orbitrap at a resolution of 120,000 over a range of 350-1400 *m/z* with an AGC target of $2x10^5$ and a maxIT of 50 ms. Monoisotopic ion precursor selection was turned on, and only ions with a charge state between 2-7 and a minimum intensity of $5x10^3$ were selected for fragmentation. Ions selected were excluded from further selection for 40 s. The default starting method ('Thermo') utilised a 0.8 *m/z* isolation width to select ions from the MS1 survey scan for Collision Induced Dissociation (CID) fragmentation at a normalised collision energy of 30%. Scans of fragment ions were acquired on the ion trap in Rapid Scan mode with an AGC target of $1x10^4$ and a 70 ms maxIT. Fragment ions were selected for further fragmentation using Synchronous Precursor Selection. Fragment ions were selected from 400-1,200 *m/z* and excluded ions 20 *m/z* below or 5 *m/z* above the precursor ion mass, and *m/z* ratios correlating to the loss of TMT from the

precursor ion. The top 10 most intense fragment ions, unless otherwise specified, were selected for HCD fragmentation with a 55% normalised collision energy and an isolation width of 2 *m/z*. The default method ('Thermo') MS3 scans were acquired on the Orbitrap at a resolution of 60,000 from 100-500 *m/z*, an AGC target of $1x10^5$ and a maxIT of 150 ms. A cycle time between MS1 survey scans was set to 2 seconds and a Top Speed method to select precursor ions for MS2 and MS3 analysis was used, allowing software to determine the optimal number of ions to fragment in this period.

The 'Thermo' method was sequentially modified to increase the maxIT to 300 ms ('300 ms scan'), then increasing the isolation width for precursor ions for fragmentation for the MS2 scan to 1.6 *m/z* ('1.6 *m/z*').

A second batch of peptides, using mouse and trypanosome samples, was prepared as described above, with mouse peptides labelled using all six TMT reporter labels and trypanosome in three. This was utilised to determine the effects of: modifying the number of 'notches' (number of fragment precursor ions) in the SPS method for the MS3 scan; titrating the maxIT for the MS3 scan; changing the isolation width for the MS1 to MS2 scan leading to the MS3 scan; and titrating the AGC target at either 10 or 2 notches. The output from these experiments is detailed in the Appendix.

RAW data files were analysed using COMPASS version 1.4.5295.18481, supplied with a concatenated database of *Homo sapiens* or *Mus musculus* from UniProt and *T. brucei brucei* proteins from TriTrypDB. The COMPASS Database Maker was used to create a peptide database containing concatenated target and reversed decoy peptides. DTA files were produced from RAW files using the DTA Generator and peptides matched using the Open Mass Spectrometry Search Algorithm (OMSSA). The mass tolerance was set to 20 ppm for precursor ions and MS/MS tolerance was set to 0.35 Da.

The enzyme was set to trypsin, allowing for up to 3 missed cleavages. NEM on cysteine was set as a fixed modification. Acetylation of protein N-termini, deamidation of asparagine and glutamine, pyro-glutamate (with N-terminal glutamine), oxidation of methionine and TMT 6-plex on lysine and N-termini of peptides were set as variable modifications. A maximum e-value of 100 was set and peptide matches required at least one match to the top ten most intense fragment peaks. Peptide spectral matches were then filtered using a maximum false discovery rate of 1% and a maximum error of 20 ppm. Resulting files of filtered peptide spectral matches (psms) were then quantified using the Tag Quant application in COMPASS. MS3 quantitation was turned on for quantitation of data acquired on the Fusion.

Quantified psms were filtered to ensure they matched human peptides, were identified with a TMT modification, were identified in all six TMT channels and the sum of the reporter intensities were above 100,000. Ratios were calculated between TMT reporter channels 126:128, 126:127, 127:128, 131:129, 131:130 and 130:129.

2.18 TMT labelling of samples from single-cut elutriation

Three biological replicates of single-cut elutriation were performed as described in section *2.13*, with time-points collected at 0.5, 3, 5, 6, 7, 8, 9, 10 and 11 h. For each biological replicate cells were harvested during log-phase growth at a density ~$3x10^7$ cells/mL. Between 4.6-$5.7x10^9$ cells were harvested in each experiment, producing between 1.35-$1.65x10^9$ small cells. At each time-point cells were washed once in 15 mL of PBS and three times in 2 mL PBS at 4°C prior to lysis in 200 µL of 4% SDS, 10 mM sodium phosphate (pH 6), 100 mM NaCl, 25 mM TCEP and 50 mM NEM. Lysates were sonicated in a Bioruptor Pico (Diagenode) water bath sonicator for 10 min and heated to 65°C for 10 min prior to chloroform-methanol precipitation.

For chloroform-methanol precipitation one volume of lysate (200 µL) was mixed with four volumes of methanol, one volume of chloroform and three volumes of Milli-Q H_2O and vortexed for 1 min. Samples were spun at 9,000 g for 5 min at room temperature in a bench-top centrifuge. The upper phase was removed, carefully avoiding the inter-phase of precipitated protein. Three volumes of methanol were added and the sample centrifuged again, followed by removal of all remaining supernatant. Protein pellets were air-dried and resuspended in one volume of 8 M urea, 1 mM $CaCl_2$ in 0.1 M Tris-HCl (pH 8).

Protein concentrations were determined by Bradford assay for each time-point and LysC added at a 1:100 ratio of protein to protease and digested overnight at 37°C. Samples were diluted to 1 M urea with 0.1 M Tris-HCl (pH 8) and 1 mM $CaCl_2$ and trypsin added at the same ratio. Digestion proceeded for 6 h prior to acidification of samples with TFA to 1%. Each time-point was loaded onto a 500 mg SepPak cartridge to desalt the samples as described in section *2.13*, although peptides were eluted using 50% acetonitrile and 0.1% TFA. Peptides were dried using a GeneVac evaporator and resuspended in 50 mM HEPES (pH 8.5) with 123 µg from each sample used for TMT labelling.

TMT ten-plex reagents were used to label one biological replicate each. The labels from 127N-131 were used to label the nine time-points collected in each biological replicate, and the first reagent (126) was used to label a peptide mix derived from only one biological replicate. This mix was produced by equally mixing aliquots of peptide from each time-point, and following TMT labelling, this was equally added to each set of biological replicates.

Each biological replicate was fractionated using high-pH reverse phase chromatography. The concentration of buffer B at the start of the chromatography run was 35%, and proceeded as described in section *2.12*. Fractions were collected from 2 to 16 min with 8.75 seconds per fraction, producing 96 fractions. Fractions were collected into 24 samples, for example the 1st, 25th, 49th and 73rd fractions were collected in the same well of a 96-well plate. The 24 samples per biological replicate were dried again using a GeneVac evaporator and solubilised in 5% formic acid.

2.19 LC-MultiNotch-MS3 and analysis of spectra

Data was acquired for each sample in triplicate producing 216 RAW files (3 runs of 24 samples from 3 biological replicates). Chromatography was performed as described in section *2.6* and data was acquired on a Fusion mass spectrometer as described in section *2.17*, with a 300 ms maxIT and a 1.6 *m/z* isolation width.

RAW data files were analysed using MaxQuant version 1.5.3.8, supplied with the *T. brucei brucei* 927 annotated protein database from TriTrypDB release 9.0. The mass tolerance was set to 4.5 ppm for precursor ions and MS/MS mass tolerance was set at 20 ppm. The enzyme was set to trypsin and endopeptidase LysC, allowing up to 2 missed cleavages. NEM on cysteine was set as a fixed modification. Acetylation of protein N-termini, deamidation of asparagine and glutamine, pyro-glutamate (with N-terminal glutamine), oxidation of methionine and phosphorylation of serine, threonine and tyrosine were set as variable modifications. The false-discovery rate for protein and peptide level identifications was set at 1%, using a target-decoy based strategy. Only unique peptides were utilised for quantitation.

2.20 Data analysis of cell cycle

Each TMT reporter channel was normalised by dividing by the median value of the intensity distribution. The cell-cycle time-course was then scaled for each individual protein quantified in each technical replicate by dividing all values by the maximum detected value. Mean values for the biological replicates were calculated from the technical triplicates, and maximum fold-changes were calculated by dividing the maximum detected value by the minimum. Proteins detected in all three biological replicates were tested by an analysis of variance (ANOVA) to determine if there was more variance between time-points or among the same time-points between biological replicates. Maximum normalised proteins with a p-value <0.05 from the ANOVA, were hierarchically clustered based on Euclidean distance measurement and a 'complete' agglomeration method. GO enrichment analysis of each cluster was performed using the GO enrichment tool for sets of genes on TriTrypDB.

Gene classifier information was obtained from multiple sources and mapped to proteins in this study using assigned TriTrypDB gene IDs. TriTrypDB was used to filter for proteins annotated with GO terms associated with the cell-cycle; genes specified in (Alsford et al., 2011) were classified as 'essential'; and genes identified as upregulated in different cell cycle phases in (Archer et al., 2011) were classified with the appropriate phase (early or late G1, S or G2&M phase). TriTrypDB was used to determine the overlap in identified proteins/genes, and those classified as cell-cycle regulated, between proteomic and transcriptomic data.

Chapter 3: Mapping core protein complexes of *Trypanosoma brucei*

3.1 Introduction

In this chapter, I present the application of protein-correlation-profiling mass spectrometry (PCP-MS) methodologies to map protein complexes in *T. brucei*. Our aim was to produce information that, through 'guilt-by-association', would provide functional suggestions for many uncharacterised proteins, aiding the annotation of the *T. brucei* proteome. PCP-MS methods are particularly useful for the study of protein complexes in trypanosomes, because they obviate the need for genetic tagging of thousands of individual proteins, along with the associated financial and labour costs of maintaining cell lines and performing thousands of immunoprecipitations. Rather, PCP-MS methods allow us to study protein complexes in a 'native' state, directly from cellular lysates, and provide a platform that can be easily adapted for future comparative studies of *T. brucei* interactomes in different biological conditions.

To achieve this, I fractionated 'native' lysates from procyclic form (PCF) *T. brucei*, followed by mass spectrometry to identify and quantify elution profiles of 5,854 protein groups. I have utilised three different chromatography columns, encompassing two biochemically orthogonal modes of separation, size exclusion chromatography (SEC) and strong anion exchange (SAX), and distinct molecular weight separation ranges for SEC. We demonstrate that detected protein elution profiles are highly reproducible across biological replicates, and that proteins in the same complex consistently co-elute. We utilise this large dataset to assess co-elution patterns between all proteins detected by hierarchical clustering and, further, use machine

learning to produce a high-confidence global interaction network, which identifies 805 protein groups in 234 complexes. These predicted protein complexes encompass many known and novel protein complexes and protein-protein interactions. This is the largest and most comprehensive analysis of soluble cytoplasmic complexes in PCF *T. brucei* to date.

3.2 Aims and hypotheses

Our hypotheses were that:

- We could separate physiologically relevant protein complexes from PCF *Trypanosoma brucei*.
- Orthogonal modes of chromatography could provide complementary and reinforcing evidence for protein co-elution and interaction.
- Through the identification of co-eluting proteins, we could propose the functions of many 'hypothetical' proteins.

Our aims were to:

- Optimise methods to separate physiologically relevant protein complexes from cellular lysates.
- Advance methods of data analysis to reconstitute protein complexes *in silico*, from chromatographic elution profiles of individual proteins.
- Identify novel interactions in *T. brucei* from our predictions to improve genome and proteome annotation.

3.3 Results

3.3.1 SEC and SAX of Trypanosoma brucei lysates

PCF *Trypanosoma brucei brucei,* strain 427, were prepared for native protein complex analysis by resuspension in either ice-cold PBS (for SEC), or 20 mM ethanolamine (for SAX), containing protease inhibitors, followed by sonication lysis. The resulting lysates were centrifuged, filtered, and fractionated either using BioBasic SEC300 or SEC1000 columns, separating protein complexes based on their size and shape; or a Protein-Pak HiRes SAX column, separating protein complexes based on their charge. The proteins in the fractions from each type of chromatography were reduced, S-alkylated and digested to peptides with trypsin and endopeptidase LysC. After desalting, the resulting peptides were analysed by LC-MS/MS (Figure 3.1).

Protein molecular weight standards were utilised to characterise the separation ranges of the BioBasic SEC300 (Figure 3.2) and SEC1000 (Figure 3.3) columns, indicating that the SEC300 column has an effective separation range from 8 kDa to 1.2 MDa, while the SEC1000 column separates material above 1.2 MDa. The retention times of each standard were used to generate a linear regression model, allowing us to calculate apparent molecular weights for the proteins and protein complexes found in our dataset (Figure 3.4). A separate set of protein standards were used to characterise the resolution and separation of the SAX column (Figure 3.5).

To assess the monomeric molecular weights of proteins eluting across the SEC300 fractionation range, fractions were pooled in groups of three, run on SDS-PAGE under reducing conditions, and stained for total protein. Most proteins eluted at a higher apparent molecular weight by native SEC than expected from their monomeric, denatured and reduced molecular weights (indicated by SDS-PAGE). This is consistent

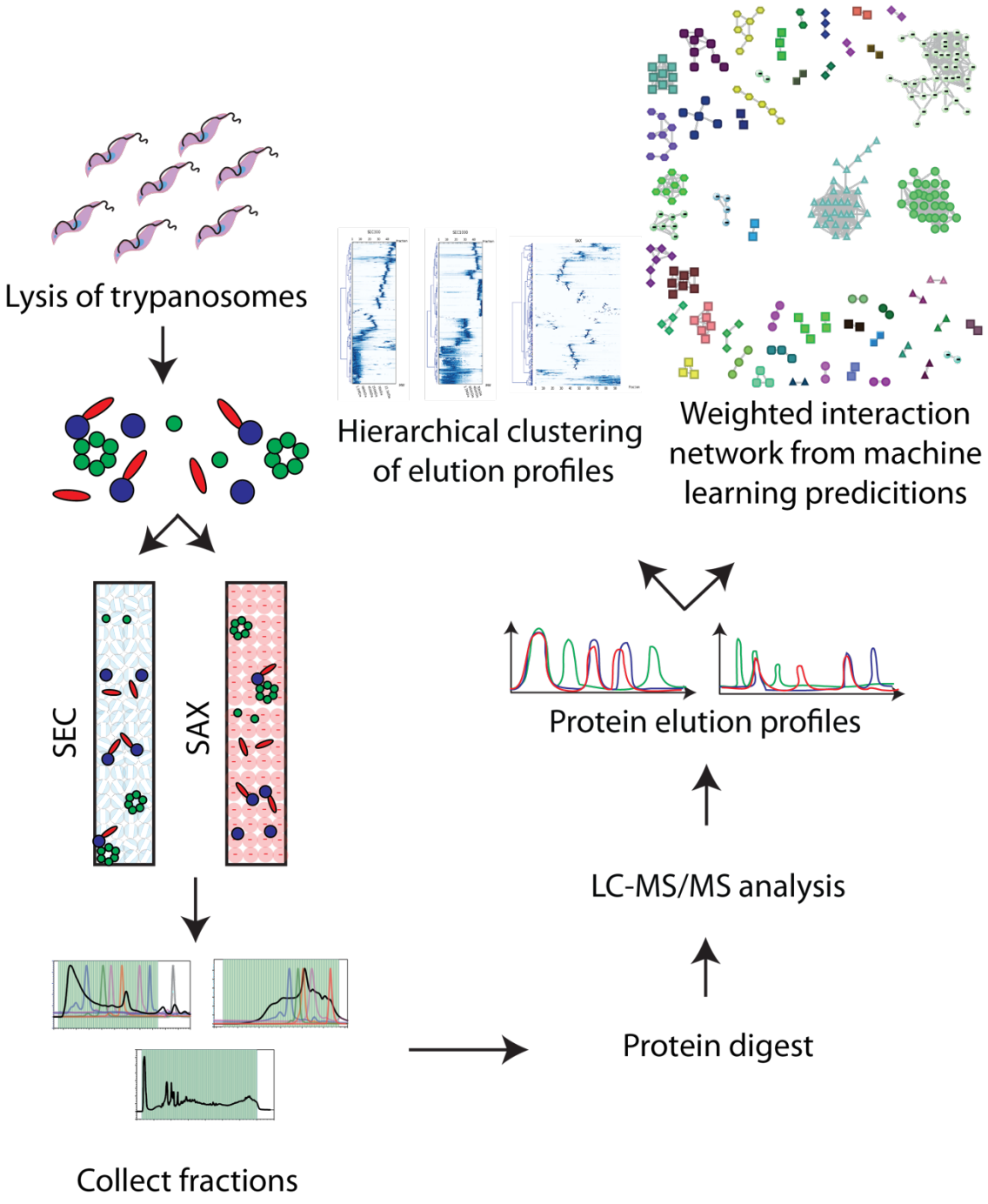


Figure 3.1: Overview of workflow for protein complex prediction. Lysates are produced containing a mixture of protein complexes, which are separated by either size-exclusion (300 and 1000 Å pore size) or strong anion exchange chromatography. The proteins in each fraction are digested and identified by LC-MS/MS from which protein elution profiles can be deduced. Putative protein-protein interactions are predicted via both hierarchical clustering of similar elution profiles and through machine learning analysis. Figure adapted from *(Kirkwood et al., 2013)*.

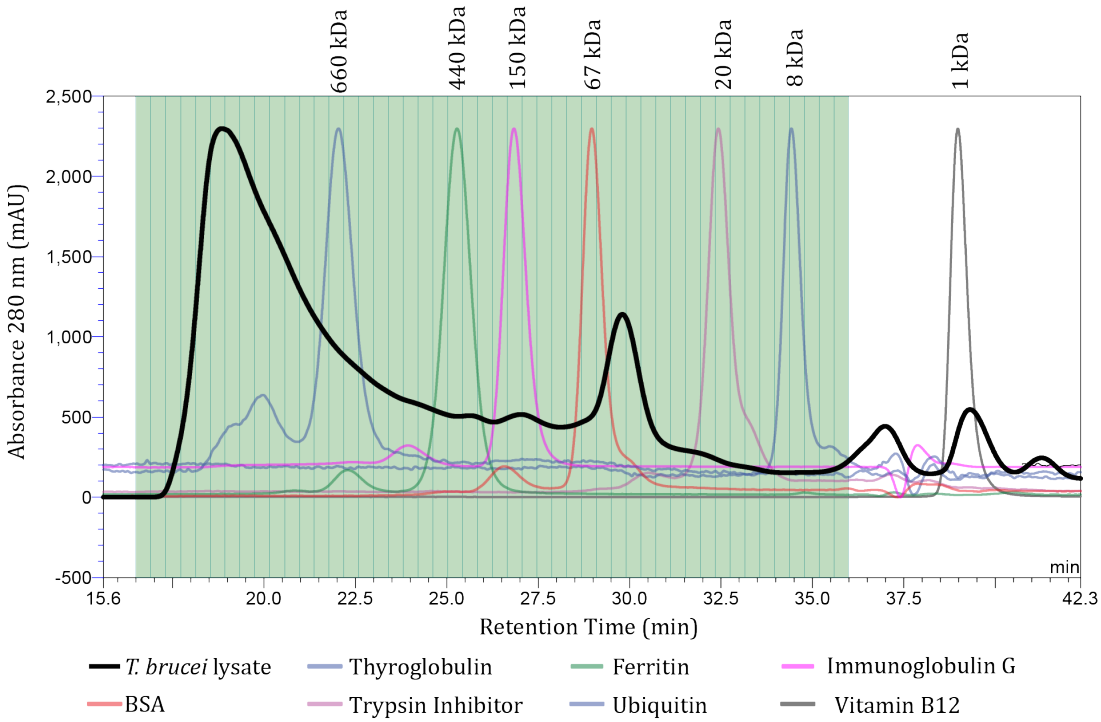


Figure 3.2: SEC300 chromatography. Chromatogram of absorbance (mAU) at 280 nm of a PBS based *T. brucei* lysate (black) and multiple molecular weight markers (colour) injected onto an SEC300 column.

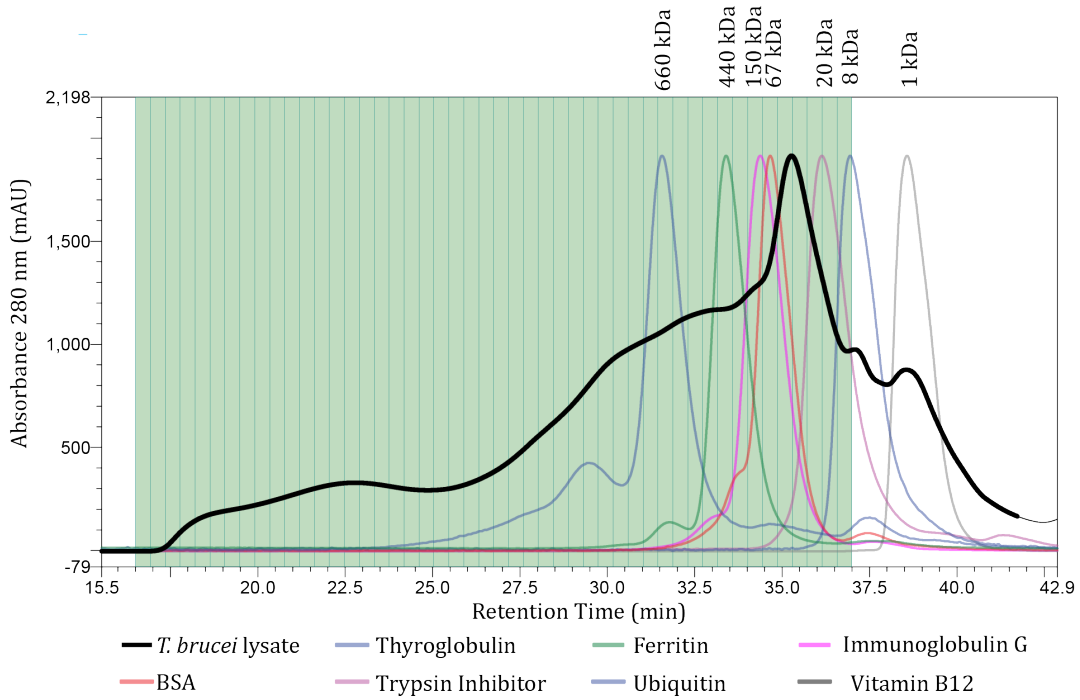


Figure 3.3: SEC1000 chromatography. Chromatogram of absorbance (mAU) at 280 nm of a PBS based *T. brucei* lysate (black) and multiple molecular weight markers (colour) injected onto an SEC1000 column.

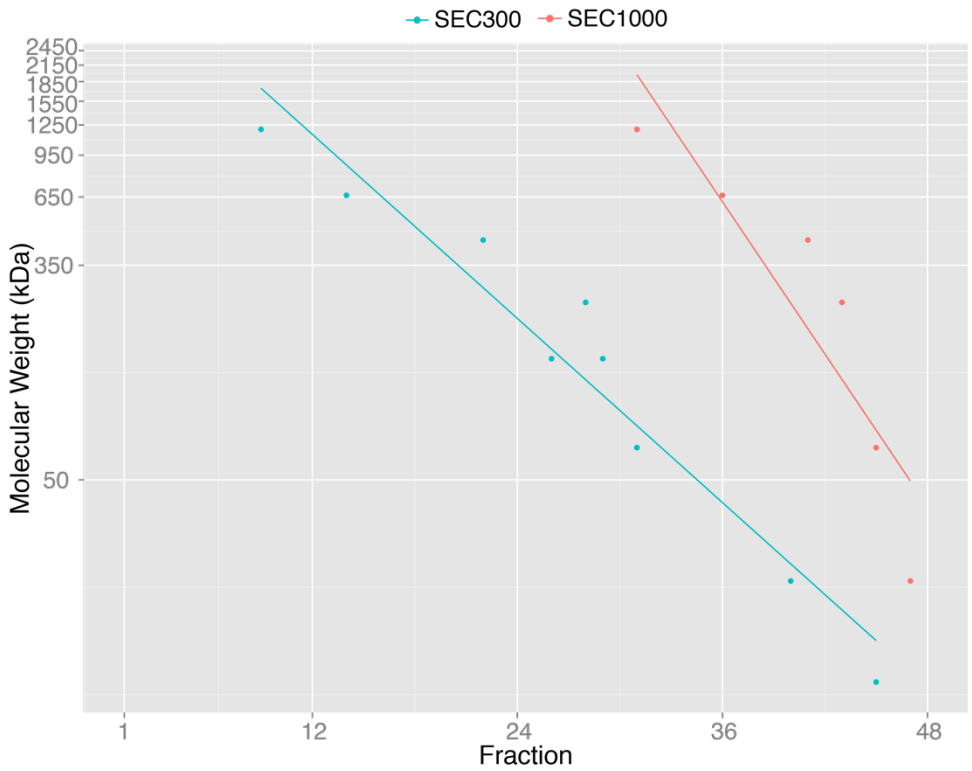


Figure 3.4: Linear regression model for M_w. Approximate molecular weight of complexes/proteins in each fraction as determined from elution of molecular weight markers on SEC300 column (blue) and SEC1000 column (red).

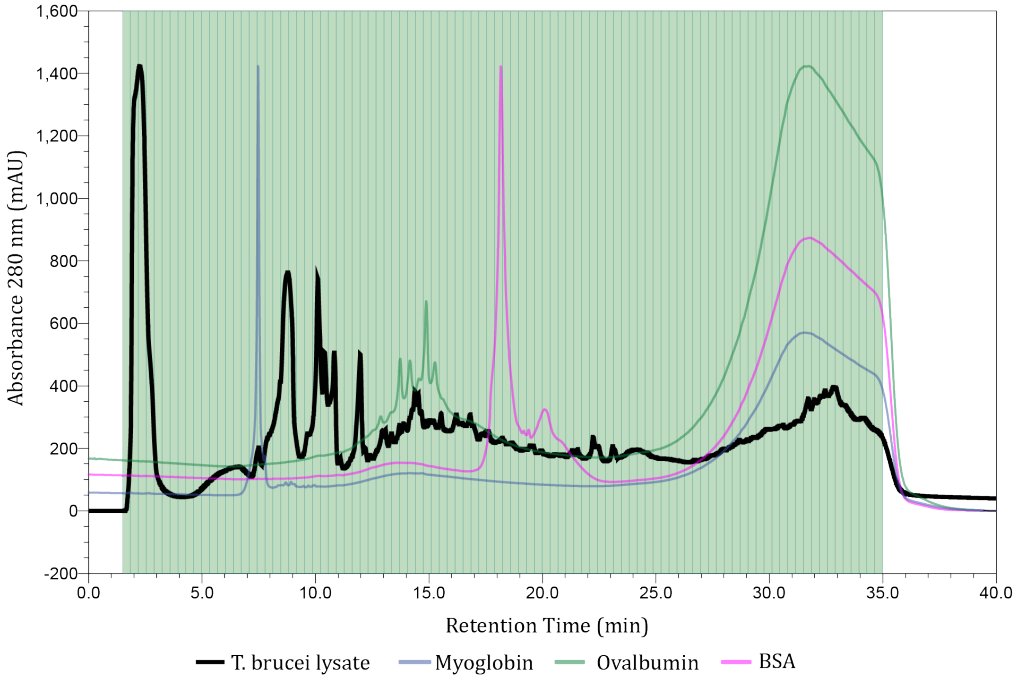


Figure 3.5: SAX chromatography Chromatogram of absorbance (mAU) at 280 nm of a *T. brucei* lysate, in black, and multiple molecular weight markers, in colour, injected onto a SAX column.

with many of the individual proteins participating as components of larger complexes. By contrast, when cells were lysed in a highly denaturing buffer, containing 4% SDS, and the SEC was carried out in the presence 0.2% SDS, I could see a direct correlation between SEC and SDS-PAGE apparent molecular weights (Figure 3.6).

I found by Western blotting that enolase, thought to be a monomeric protein, eluted mostly at the correct predicted molecular weight by SEC, whereas tryparedoxin peroxidase, known to form 50 kDa dimers and 250 kDa decamers in its recombinant form (Alphey et al., 2000; Piñeyro et al., 2005), eluted at the void volume (> 1 MDa) by SEC300 (Figure 3.7). The latter result may suggest that tryparedoxin peroxidase forms higher order oligomers *in vivo* than *in vitro*.

In addition, a peak-picking analysis of the SEC300 dataset was performed, where individual protein elution profiles were broken up into their constituent peaks, extracting multiple 'peak' profiles from a single 'protein' profile (Figure 3.8). Peaks were then classified based on their retention time from SEC. Proteins eluting in the first ten fractions were classified as 'void volume' proteins, i.e., present in material that is larger than the 30 nm pores of the SEC column, hence eluting first without any separation. The second classification was based on the expected molecular weight of the protein; if an individual protein eluted at the expected molecular weight, or two times the expected molecular weight, the protein was classified as a monomer or a dimer. Above this range, the protein was classified as being in a protein complex (Figure 3.9). Of 3,326 detected protein elution peaks, 1,480 (44%) are classified in the monomer or dimer range, 658 (20%) in the protein complex range, and 1,188 (36%) in the void volume of the SEC300 column.

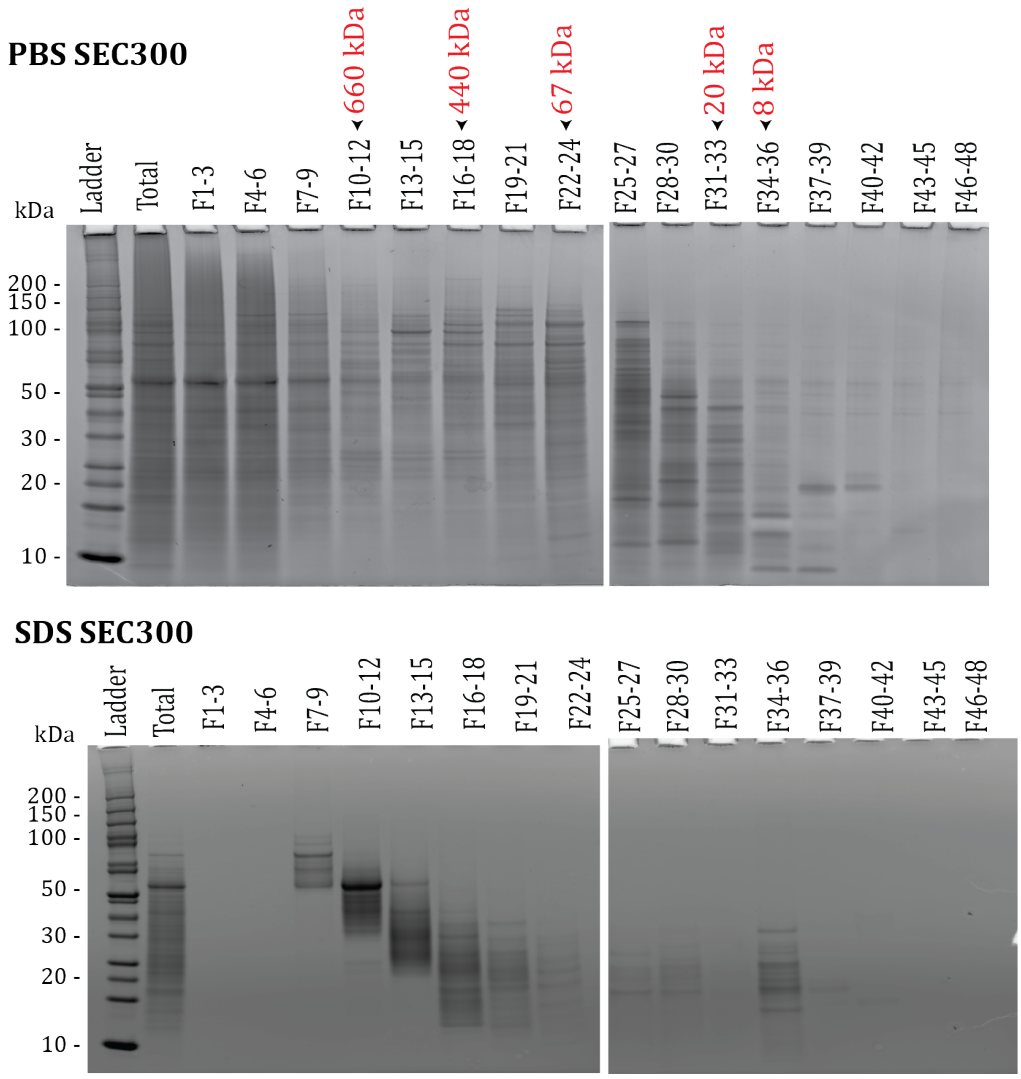


Figure 3.6: SyPro Ruby staining of SDS-PAGE gels. Staining of consecutive fractions collected from SEC300 fractionation of PBS based lysate (top) and SDS based lysate with denaturing chromatography (bottom).

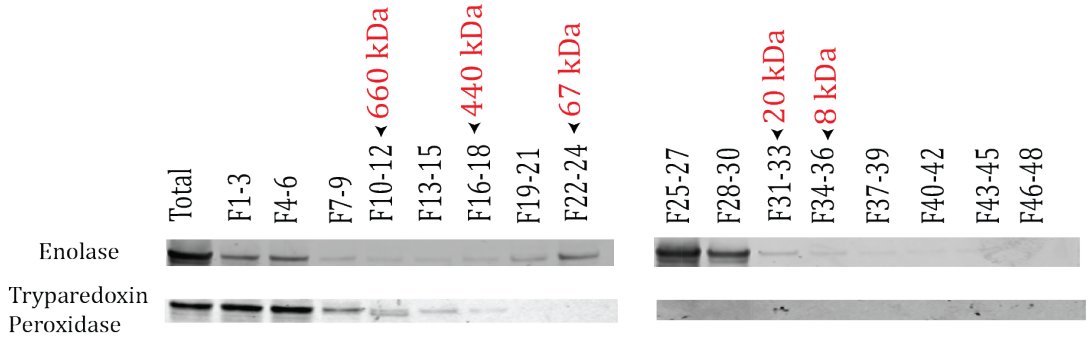


Figure 3.7: Western blotting for enolase and tryparedoxin peroxidase. Western blotting of consecutive fractions collected from SEC300 fractionation of PBS based lysate.

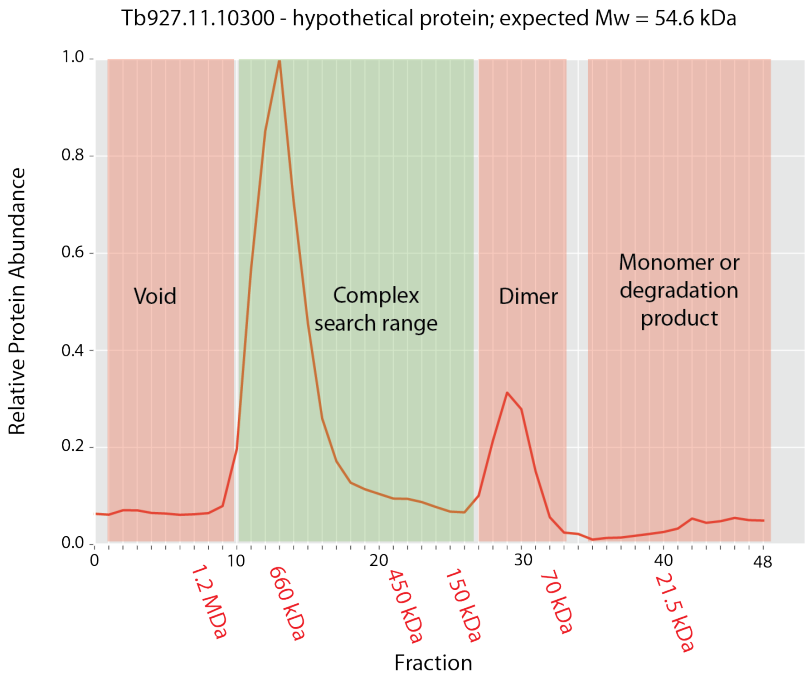


Figure 3.8: Peak picking analysis from SEC300 fractionation. If a protein contains a peak in the first ten fractions it is classified as 'void' material. Proteins with peaks in their expected monomeric or dimer molecular weight are classified as 'dimer/monomer or degradation product'. Remaining peaks are in the optimal molecular weight range for an SEC300 column.

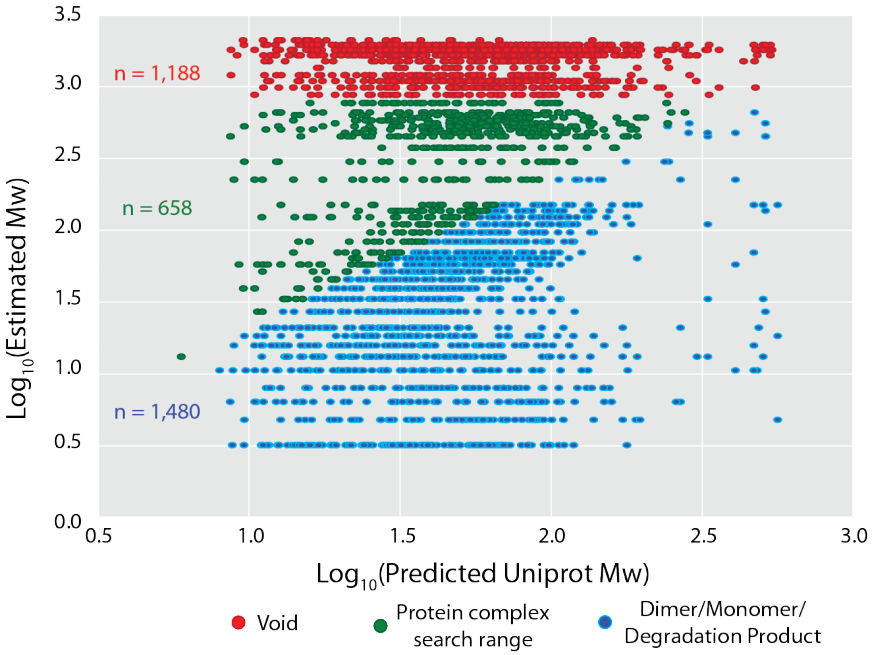


Figure 3.9: Comparison of Uniprot predicted and SEC300 estimated molecular weight. Different colours depict the different categories classified from the peak picking analysis, and n shows number of protein peaks detected in each category.

Gene ontology analysis of these three different categories highlights the enrichment of integral membrane and organellar proteins in the void volume, a number of distinct protein complexes (ARP2/3, chaperonin, proteasome) in the complex range category and more general cytosolic terms in the monomer and dimer category (Figure 3.10).

From five biological replicates of SEC300 fractionation experiments, 65,967 peptides were identifed, corresponding to 5,929 protein groups, detected by at least one unique peptide. From five biological replicates of SEC1000 chromatography, 58,069 peptides were identified, corresponding to 5,628 protein groups, and from a single SAX fractionation experiment, 38,804 peptides were identified, corresponding to 3,213 protein groups.

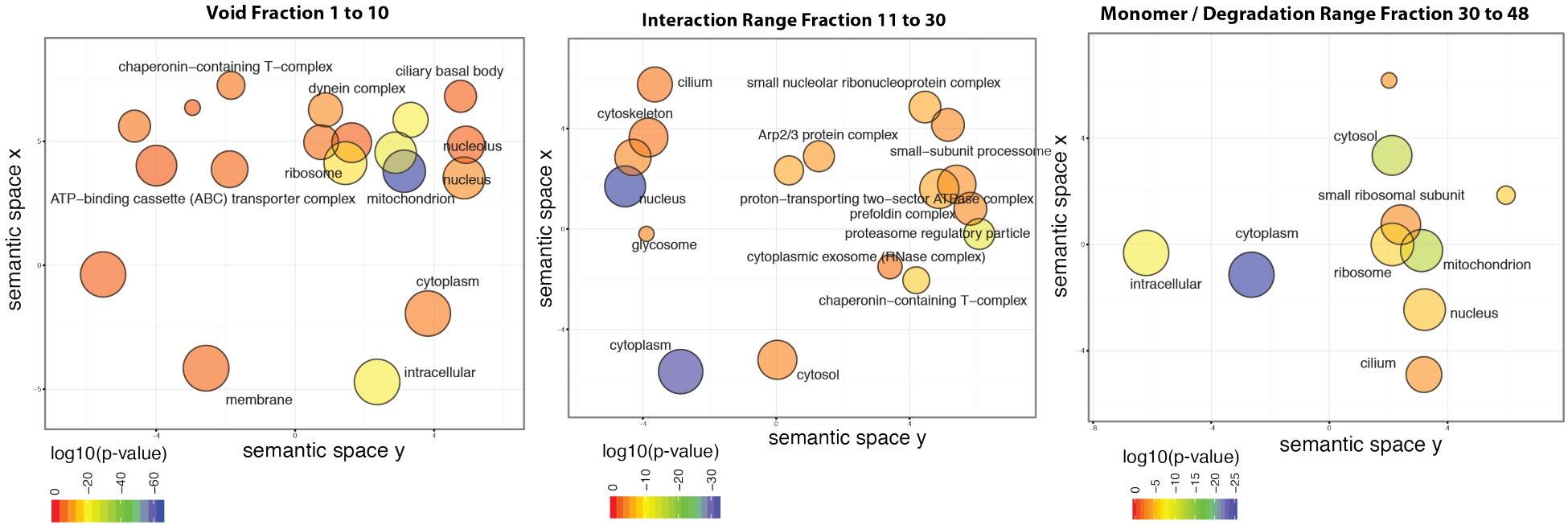


Figure 3.10: Gene ontology enrichment analysis of proteins detected in each category from peak picking analysis. The left hand side panel shows the GO enrichment of void fractions; the middle panels shows the protein complex range; and the right hand side panel shows the dimer/monomer/degradation range. The circle size represents the frequency of the GO term in the underlying GO annotation database.

3.3.2 Reproducibility of mass spectrometry based elution profiles

The reproducibility of individual biological replicates was assessed for both the SEC300 and SEC1000 experiments. Figure 3.11 shows the median Pearson correlation coefficient in each fraction from each biological replicate compared to the four other replicates, for both SEC300 and SEC1000 fractionation experiments. Most biological replicates within SEC300 or SEC1000 data showed high reproducibility, with median Pearson correlation coefficients of each fraction typically >0.75. However, for one experiment on SEC300 (E3025) and two experiments on SEC1000 (E3022 and E3026) the datasets deviated from the other replicates. Comparison of the elution profiles of individual proteins between replicates demonstrated that reproducible protein elution patterns were lost in E3025, E3022 and E3026 due to compromised chromatographic resolution (Figure 3.12), and these replicates were therefore excluded from further analyses.

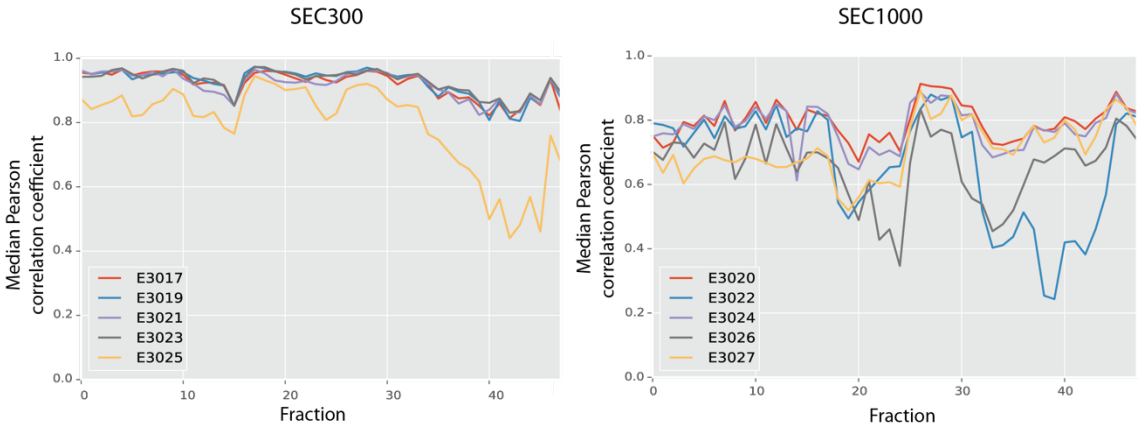


Figure 3.11: Median Pearson correlation within fractions. Median Pearson correlation coefficient of each fraction in each biological replicate from SEC300 (left) and SEC1000 (right). E3025 (SEC300) and E3022 and E3026 (SEC1000) deviate from correlation pattern seen for other biological replicates.

Pearson correlation coefficients of elution profiles were also calculated for each protein detected across remaining biological replicates, using unique and total peptide counts (MS), Intensity Based Absolute Quantitation (IBAQ) and Label Free

Quantitation (LFQ) as metrics of protein abundance (Figure 3.13). These data are described on density plots, demonstrating the high reproducibility of elution profiles across biological replicates for all metrics, with LFQ values showing the highest reproducibility. LFQ values were therefore for further data analysis.

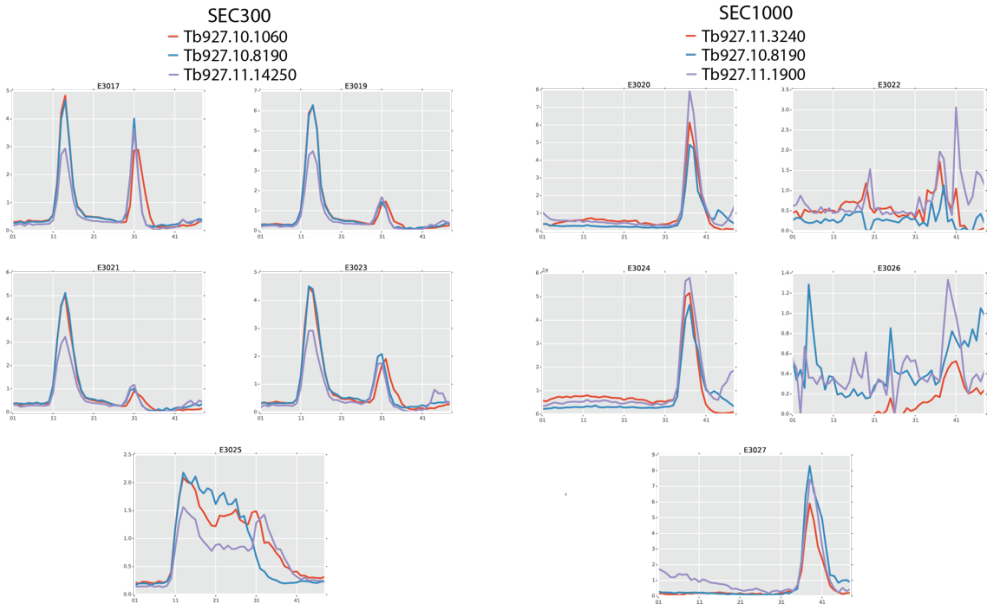


Figure 3.12: Reproducibility of replicates. Elution profiles for three proteins in either SEC300 (left) and SEC1000 (right) in each biological replicate displaying LFQ intensity over time. Reproducible elution patterns are seen in four of five SEC300 experiments and three of five SEC1000 experiments.

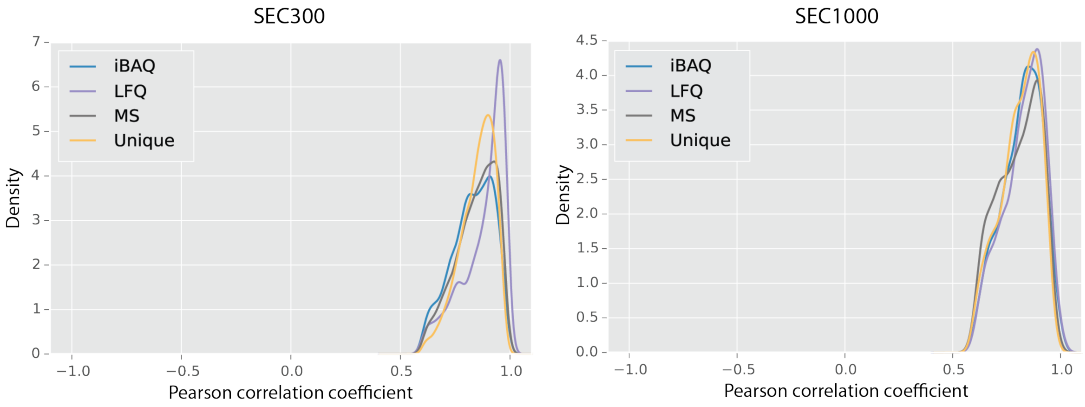


Figure 3.13: Pearson correlation of elution profiles. Density plot of the median Pearson correlation for elution profiles of each protein detected across biological replicates in SEC300 (left) and SEC1000 (right). The reproducibility of four different quantitative metrics of protein abundance are compared: iBAQ, LFQ, MS counts and unique peptide counts.

3.3.3 Hierarchical clustering of protein elution profiles

Hierarchical clustering of all detected protein elution profiles was performed for each dataset separately (Figure 3.14). By cutting the resulting dendrograms of hierarchically clustered proteins, we can define groups of proteins that have similar elution profiles, and hence that potentially interact. We simulated a range of cutting distances and observed the within-cluster Gene Ontology (GO) term enrichment and the mean Pearson correlation coefficients of elution profiles (Figure 3.14). As we reduced the number of clusters, a sharp increase in the number of enriched GO terms was observed in all datasets, as functionally associated proteins were grouped together within a cluster. As clusters became larger and unrelated proteins were grouped together, the number of enriched GO terms decreased. We therefore chose the cutting distance producing the highest GO term enrichment within clusters across the dataset for each form of fractionation. Thus, for the SEC300, SEC1000 and SAX experiments, cutting distances of 1.53, 1.28 and 1.92 were chosen, producing 440, 365 and 529 clusters of proteins, respectively. These data suggest that we are grouping together proteins with similar annotated functions, indicative of proteins interacting in a complex, and supporting the idea that hierarchical clustering of elution profiles can predict protein-protein interactions. As a control, we randomly shuffled the order of proteins within the dendrograms of each dataset and then performed the same analyses; under these conditions we did not see a similar increase in GO term enrichment across cutting distance (Figure 3.14).

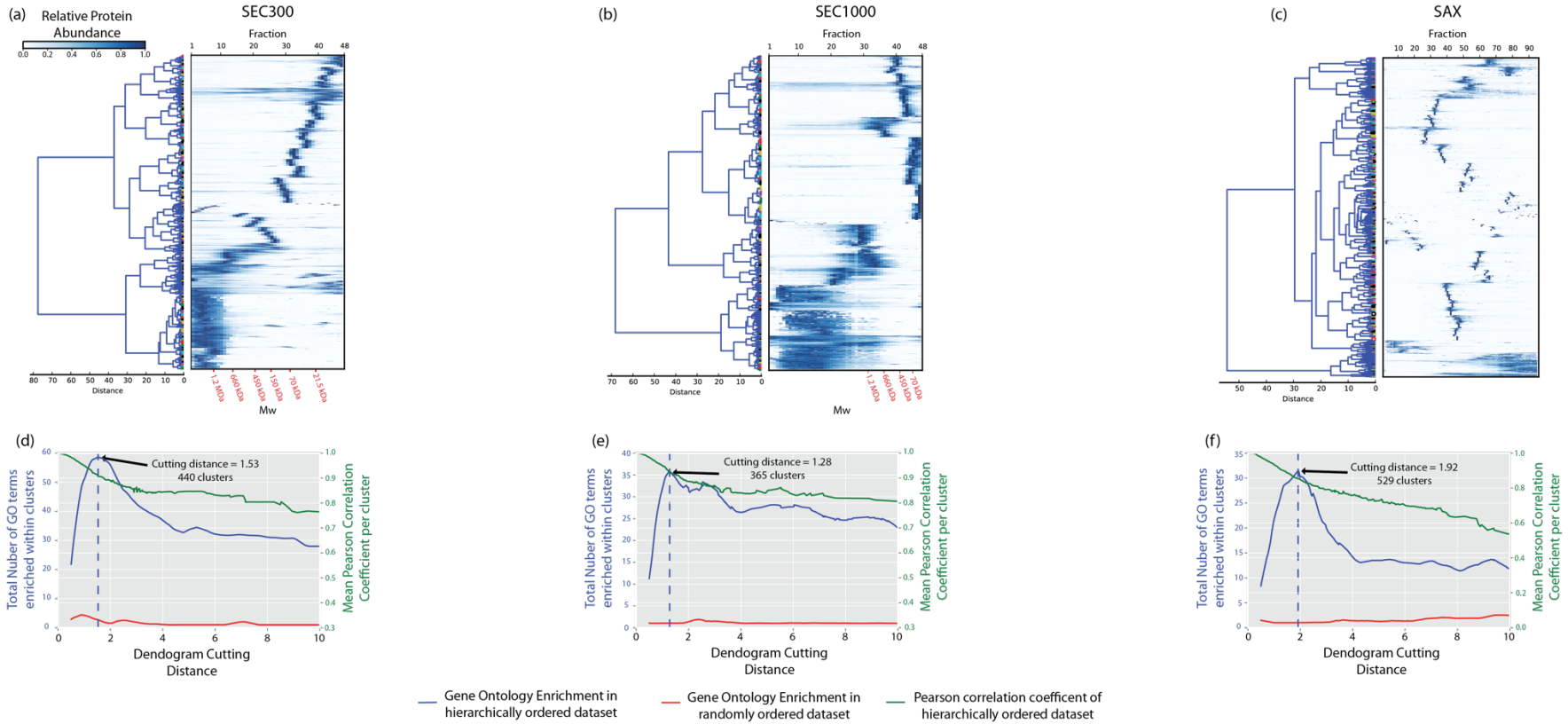


Figure 3.14: Hierarchical clustering of protein elution profiles. Heat-maps of hierarchically clustered protein elution profiles from lysates separated using either SEC, with either (a) 300 or (b) 1000 Å pore size, and (c) SAX chromatography. Plots below heat-maps demonstrate the effect of varying the dendrogram cutting distance on the mean Pearson correlation coefficient of proteins (green line), and the total number of gene ontology terms enriched (blue line) within clusters in (d) SEC300, (e) SEC1000 and (f) SAX. The red line depicts the number of GO terms enriched within clusters with a random ordering of proteins each dataset. Proteins within clusters can be viewed in the Cluster Explorer web-tool described later in the text.

3.3.4 Characterisation of known complexes

To benchmark how well each form of chromatography performed in maintaining and separating protein complexes, we inspected the elution profiles of proteins expected to be present as stable complexes, either from information from the *T. brucei* literature, or by analogy with complexes known to be highly conserved in other organisms. The elution profiles of proteins thought to form the chaperonin T-complex, the core proteasome subunit, the proteasome regulatory cap, ATP synthase, the chaperone prefoldin complex and the ARP2/3 complex across SEC300, SEC1000 and SAX chromatography are shown in (Figure 3.15).

The proteasome core and regulatory subunits have quite distinct elution profiles on the SEC300 column, with the regulatory cap eluting at ~770 kDa and the core complex at ~660 kDa. This is consistent with previous reports showing that the proteasome core and cap dissociate in PBS based *T. brucei* lysates (Li et al., 2002). The elution profiles of the proteasome core and regulatory cap components by an orthogonal chromatographic method, ion exchange on a SAX column, further demonstrates the stability of these complexes, although the core subunit appears more stable to SAX chromatography than the regulatory subunit.

Unlike the trypanosome proteasome, the prefoldin complex has not previously been characterised in *T. brucei*. Here, I show the co-elution of many predicted prefoldin subunits in SEC and SAX chromatography, providing the first supporting evidence that these proteins come together as a functional prefoldin chaperone complex in *T. brucei*. Together with the elution profiles of other known or suggested *T. brucei* protein complexes, we can see that high-resolution chromatography, fractionation, mass spectrometric identification and quantification can produce distinct and consistent

co-chromatographic elution profiles for the protein subunits belonging to unique complexes.

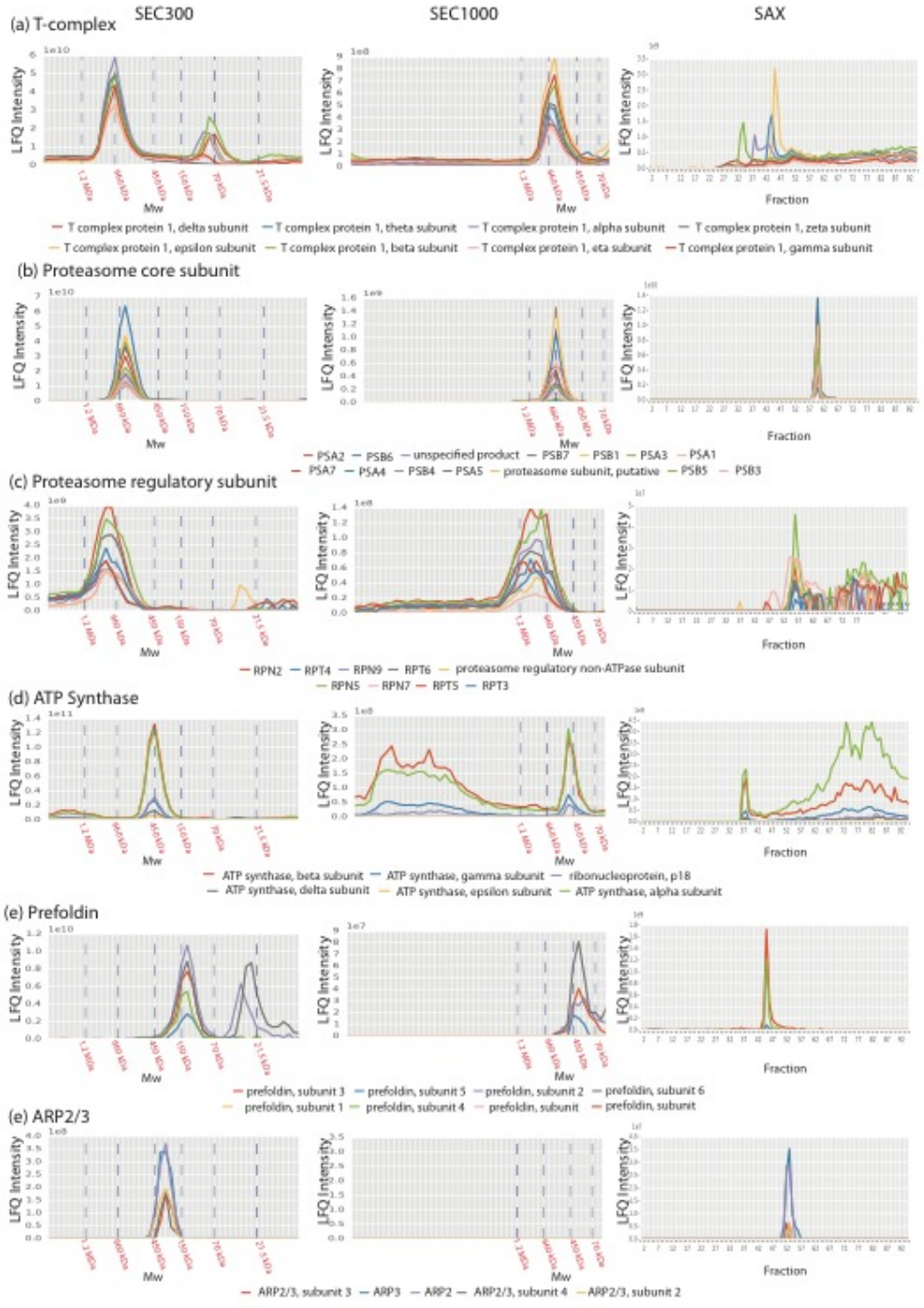


Figure 3.15: Elution of components of known core protein complexes across multiple forms of chromatography. LFQ intensity for each protein plotted across the fractionation range collected for mass spectrometry analysis. Molecular weight marker retention times on SEC300 and SEC1000 columns marked below the plot.

3.3.5 Machine learning analysis to predict protein complexes

To predict the likelihood of binary interactions between all pairs of co-eluting proteins detected in our datasets, a scoring methodology was designed to quantify the similarity of elution patterns. Two random forest predictors were implemented. The first predictor was trained with features extracted from data produced in this project. The second predictor was trained by combining the first set of features with features extracted from a recently published interactome study in *T. brucei* (Gazestani et al., 2016). Features retrieved from the version 10 of the STRING interaction database were also added to the second predictor (Szklarczyk et al., 2015) with the intention of promoting interacting protein pairs with orthogonal literature evidence for interaction (Figure 3.16).

Both predictors were trained using 31 protein complexes, comprising 293 true positive interaction pairs (Supplementary Table 1). True positive pairs had an average interaction prediction score >0.75 (Figure 3.17), hence this was used as the threshold for positive interaction across the whole dataset. Receiver Operator Curves (ROC) also demonstrate the high precision and recall of the machine learning method (Figure 3.18)

Analysis of how often the random forest predictors utilised each feature to classify positive and negative interactions showed that SEC300 co-apex score, cross-correlation and Pearson correlation and SEC1000 co-apex features had the highest predictive power (Figure 3.19).

We performed an analysis of the Pearson correlation coefficient distribution of all pairwise combinations of elution profiles detected within our SAX and SEC300 and compared them to IEX fractionation of mitochondrial extracts and glycerol gradient fractionation of whole cell lysates (Gazestani et al., 2016), respectively.

The outputs of the two random forest predictors were fed to ClusterOne to derive two sets of predicted protein complexes. These two complex predictions were merged to assemble 234 predicted protein complexes encompassing 805 proteins, with complexes ranging from 2-18 proteins. I have ascribed a putative function, or name, to the complexes when they contain proteins of known or suggested biological function in the TriTrypDB genome database (Figure 3.21, Supplementary Tables 2 and 3)(Aslett et al., 2010).

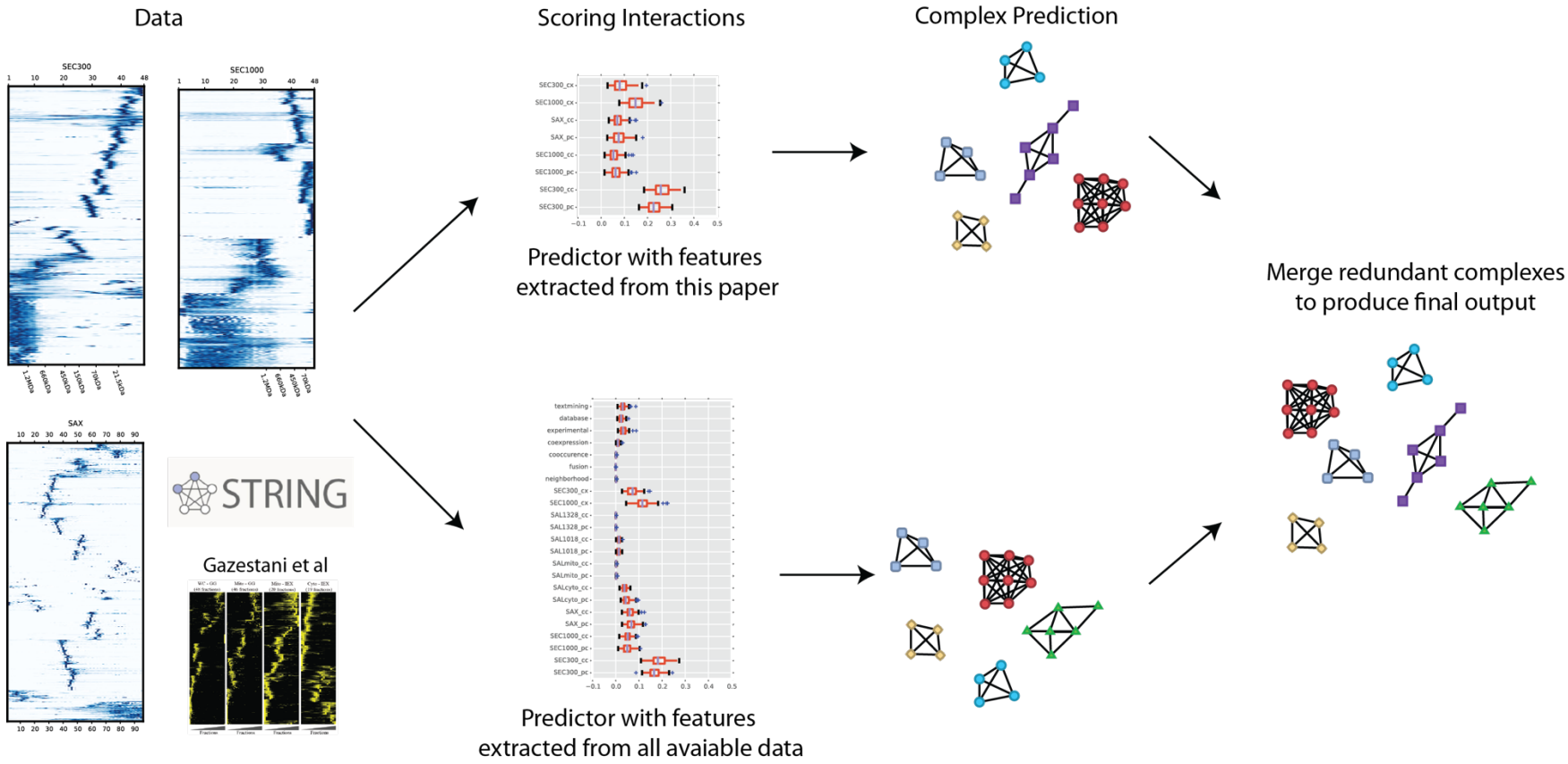


Figure 3.16: Overview of machine learning protein complex prediction pipeline. Data produced either solely in this project, or including data from STRING and a recent *T. brucei* interactome paper *(Gazestani et al., 2016)*, were used to train two sets of random forest predictors to score binary protein-protein interactions. The resulting scores were used to predict protein complexes separately for each set, then merged to join redundant complexes.

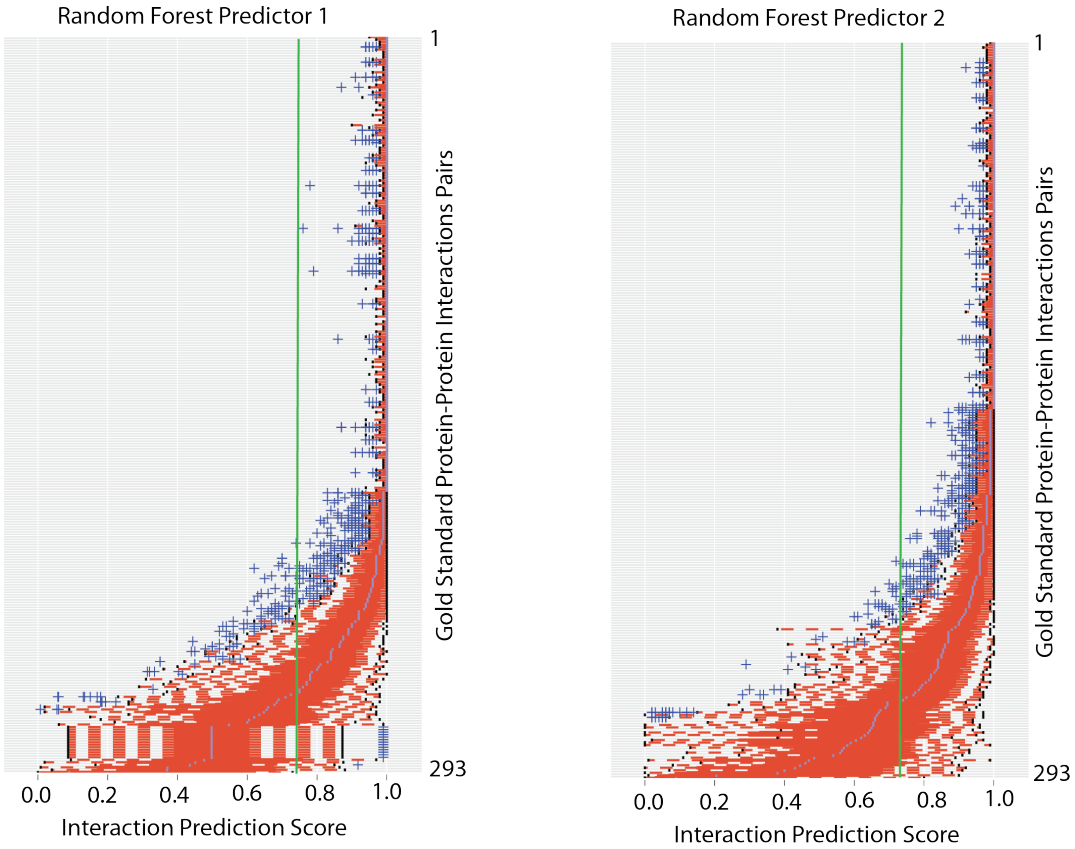


Figure 3.17: Machine learning interaction prediction score distributions. Interaction prediction score distributions of 293 gold standard protein-protein interaction pairs using random forest predictor 1 and 2. Vertical green line highlights the 0.75 interaction prediction score threshold.

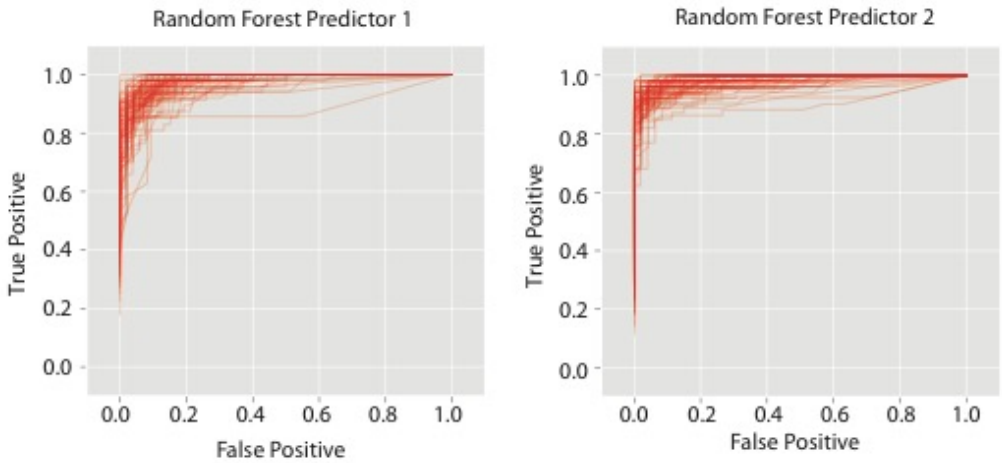


Figure 3.18: Receiver operator curves from both random forest predictors. Each random forest analysis was performed 100 times with a random selection of true negative interactors.

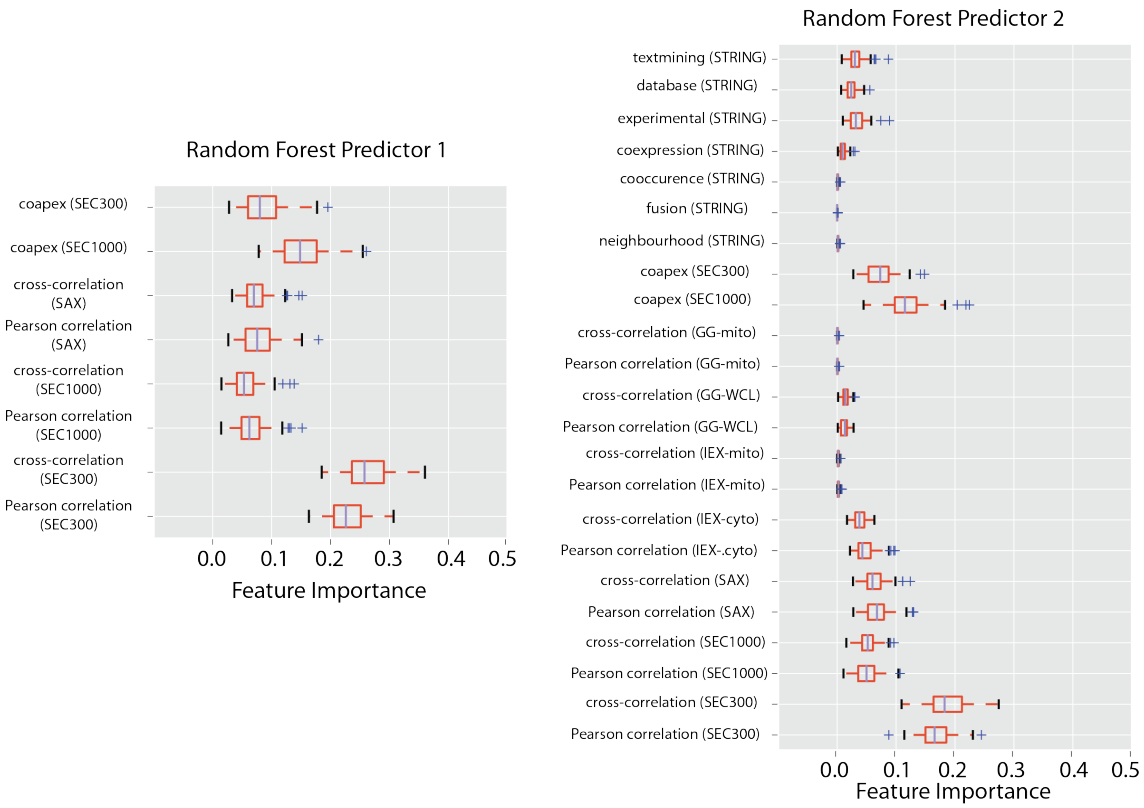


Figure 3.19: Feature importance output of random forest analyses. Feature importance reveals how often a scoring feature is used by random forest analysis to predict the gold standard protein complexes dataset. It therefore informs us on which are the most useful features for predicting protein interactions.

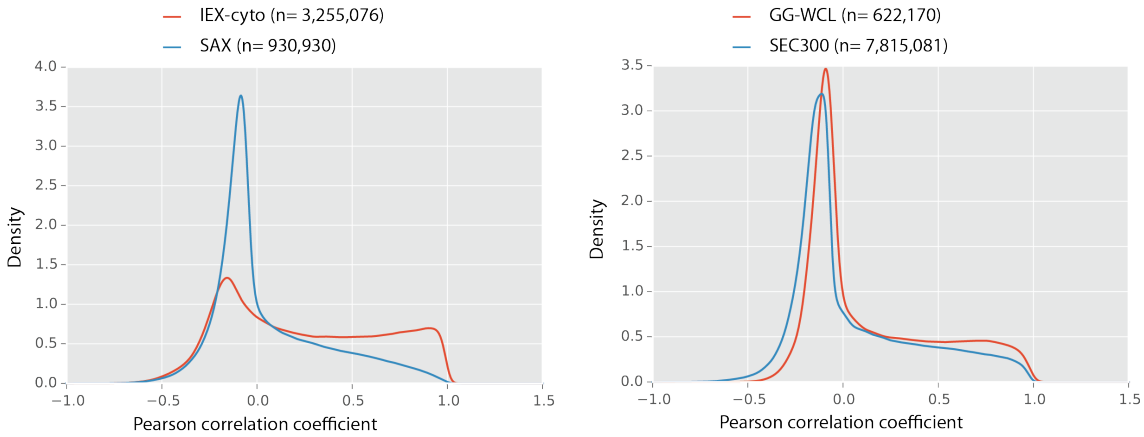


Figure 3.20: Comparative analysis of our data and published data. Distribution of Pearson correlation coefficients between all pairwise combinations of proteins in each dataset. Distributions of cytosolic IEX *(Gazestani et al., 2016)* and SAX fractionation (left) and glycerol gradient fractionation of whole cell lysate *(Gazestani et al., 2016)* and SEC300 (right). n = the number of pairwise protein correlations calculated within each dataset.

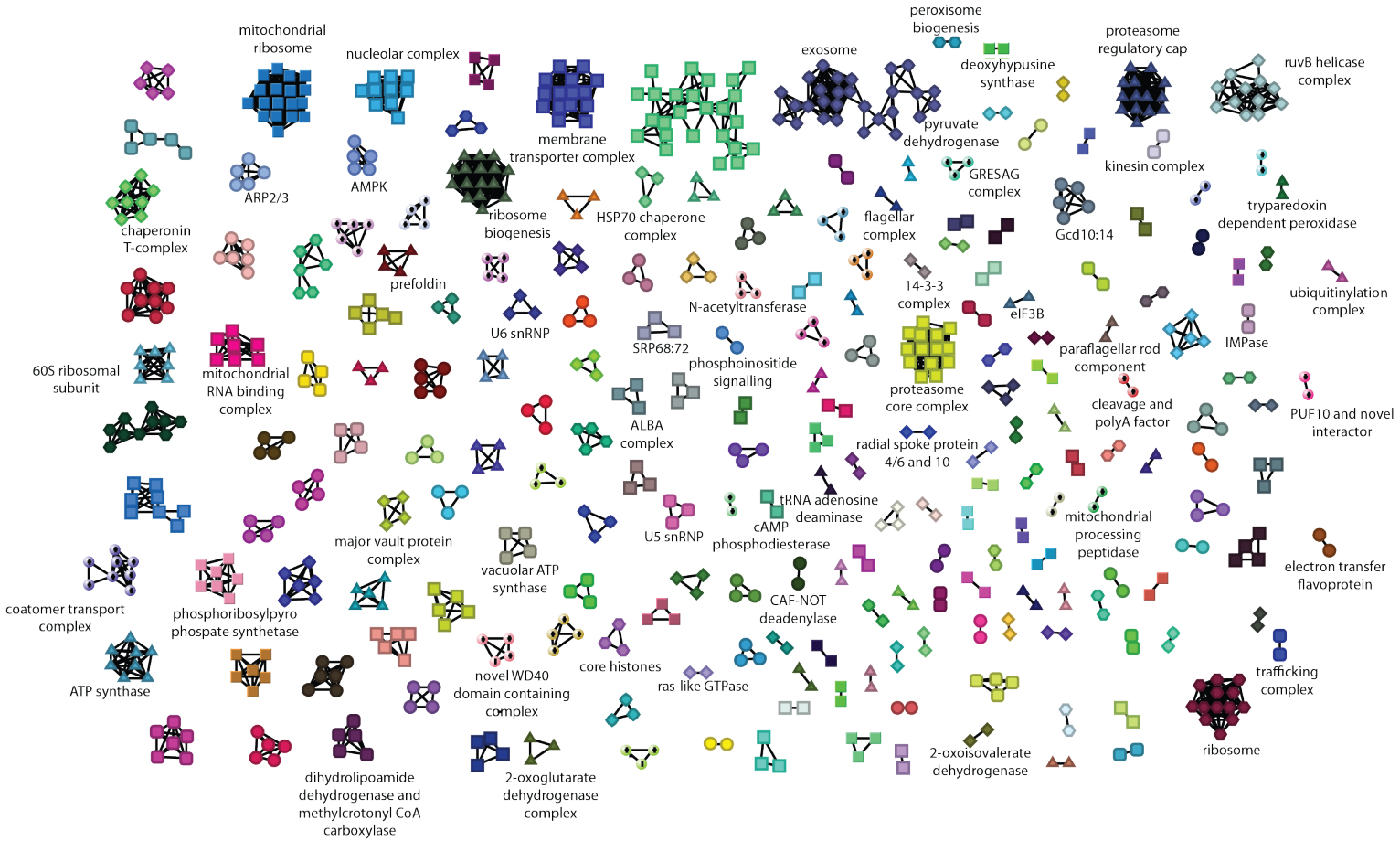


Figure 3.21: Machine learning prediction of protein-protein interactions and protein complexes. ClusterOne output of merged predictions from machine learning with both sets of predictors. Known protein complexes, or complexes with proteins of similar function have been manually annotated.

3.3.6 Prediction of novel protein complexes and interactions

Although many of the complexes we predict are highly abundant core protein complexes conserved across eukaryotic evolution, we also detect a number of novel complexes and protein-protein interactions not previously described in *T. brucei*. Some of these predicted interactions will shed light on the functions of some of the many 'hypothetical' proteins in the trypanosome genome. Some examples of previously uncharacterised associations are described in section *3.4.4*.

3.3.7 Formaldehyde and DSP cross-linking and protein fractionation

To increase the coverage of membrane protein complexes in our dataset I initiated work to optimise *in vivo* crosslinking of protein complexes with formaldehyde or dithiobis(succinimidyl propionate) (DSP), followed by lysis in denaturing SDS based buffers. A range of concentrations of either DSP (0-2 mM), or formaldehyde (0-6%) were tested and analysed before and after de-crosslinking by addition of reducing agent or heat, respectively.

SyPro stained SDS-PAGE gels were utilised to visualise the degree of protein cross-linking (Figure 3.22). The lowest concentration of DSP tested (0.125 mM) leads to the formation of large protein aggregates that could not enter the SDS-PAGE gel, indicating effective protein cross-linking. Formaldehyde only begins to show an obvious effect of crosslinking at concentrations >1%, with very few distinctive bands of monomeric proteins below 250 kDa observed at a concentration of 3%. De-crosslinking was highly effective for both methods as demonstrated by the loss of the high-molecular weight protein in the well of the gel, and the similarity in the banding pattern of proteins in each sample (Figure 3.22).

Western blotting of DSP crosslinked samples for enolase and tubulin (Figure 3.23), again shows the effective crosslinking of samples at the lowest concentrations of DSP tested. From 0.125-1 mM DSP a higher molecular weight band of ~100 kDa is detected for enolase, indicating formation of a dimer. At 2 mM this band disappeared, with a concomitant increase in the staining of material in the well of the lane. Tubulin is seen as a monomer at 49 kDa with no crosslinking, and a faint monomer and dimer can be seen upon addition of DSP, though most of the tubulin signal is lost upon cross-linking. This may be indicative of cross-linking large chunks of microtubules that may be lost from the sample upon centrifugation.

In formaldehyde cross-linked samples, enolase is detected as a monomer across the tested concentration range from 0-6%. A faint dimer band can be observed from 0.4%, and at 2% high-molecular weight material not entering the gel can be observed. A similar pattern is observed for tubulin, with the presence of a monomer and dimer band from 0.2-2% formaldehyde. Above this concentration, the monomeric band is lost and most tubulin stained material is found in the high molecular weight material (Figure 3.23).

Formaldehyde crosslinked and de-crosslinked samples were run on an SEC1000 column in an SDS based buffer. Comparison of the 215 nm chromatograms of these lysates clearly shows increasing material at high molecular weight at higher formaldehyde concentrations (Figure 3.24). Following de-crosslinking with heat, all samples looked identical.

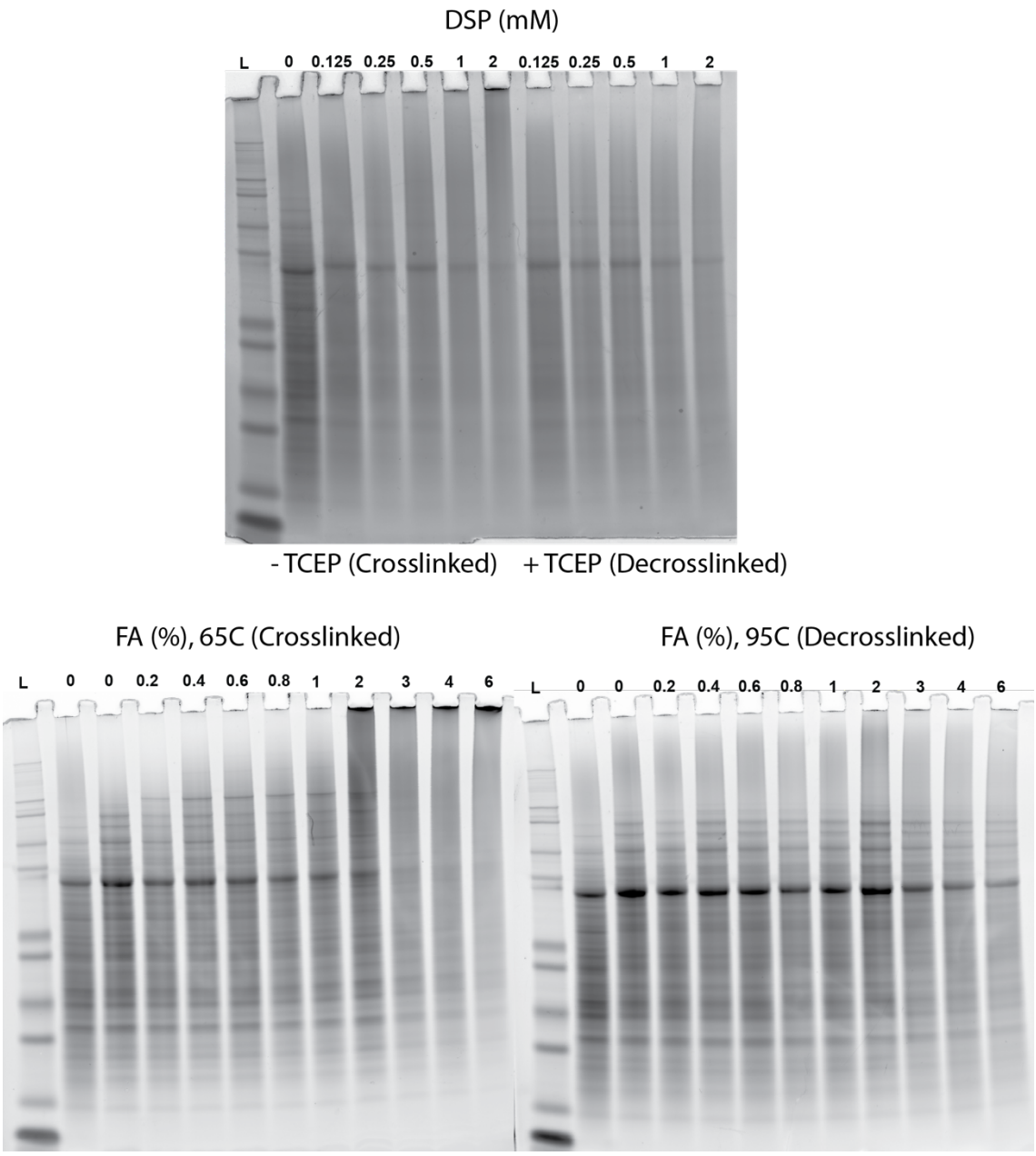


Figure 3.22: In vivo crosslinking and de-crosslinking analysis of *T. brucei*. SyPro staining of SDS-PAGE analysis of lysates from *T. brucei* cells treated with varying concentrations of DSP (top) or formaldehyde (bottom). Lysates from DSP treated cells were de-crosslinked through the addition of the reducing agent, TCEP (top, right five lanes). Lysates from formaldehyde treated cells were de-crosslinked by heating samples to 95˚C (bottom right gel).

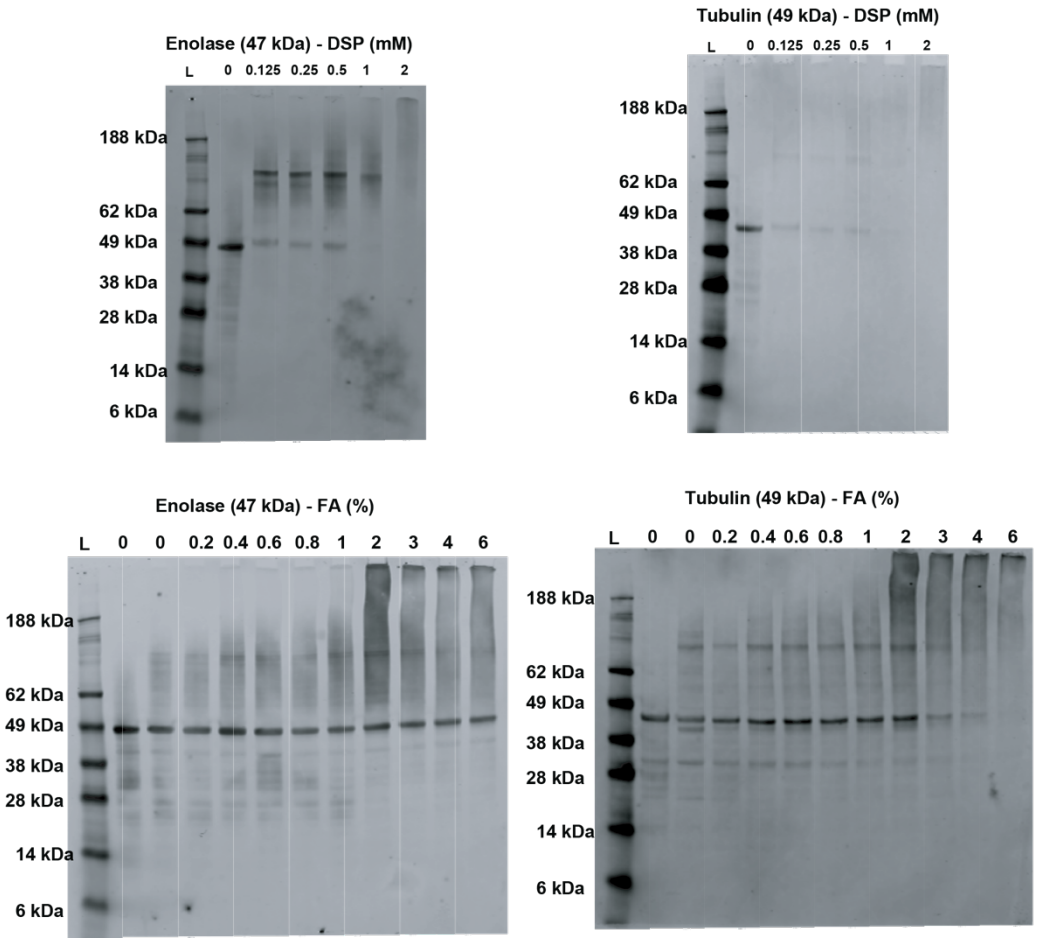


Figure 3.23: Western blotting for enolase and tubulin in lysates from DSP (top) or formaldehyde (bottom) crosslinked cells.

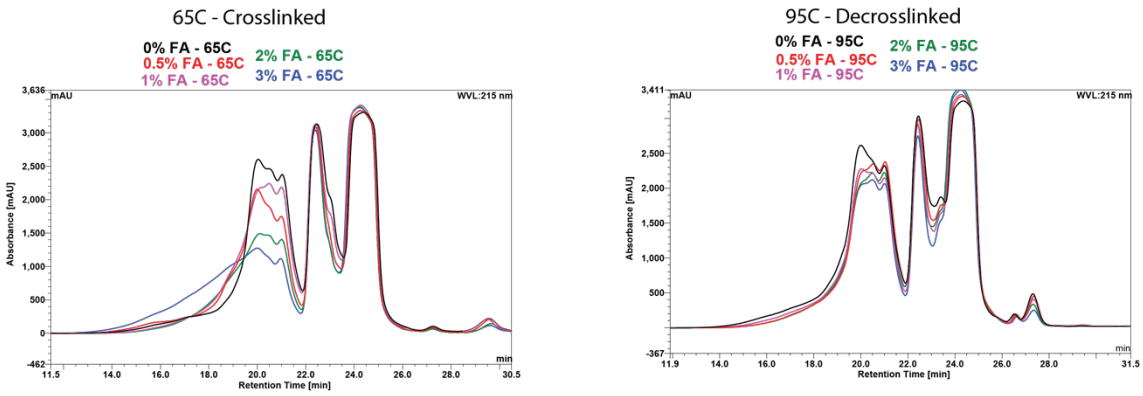


Figure 3.24: Denaturing SEC1000 on crosslinked lysates. Chromatograms of 215 nm absorbance of lysates from formaldehyde crosslinked *T. brucei*, fractionated using SEC1000 column in a denaturing buffer.

3.4 Discussion

In this study, I have produced information on the elution profiles across SEC and SAX chromatograms for 5,845 protein groups, providing a rich data source that predicts protein interactions. Computational analysis of these elution profiles allows us to predict 234 protein complexes, containing groups from two to eighteen proteins that we predicted to interact. This includes a number of well-characterised complexes in *T. brucei*, confirms some predictions from electronic annotations based on homology to known complexes from other organisms and provides evidence for the existence of previously uncharacterised complexes. We hope that these data will be a useful resource for trypanosome biologists working across a number of areas of trypanosome molecular cell biology.

Our data show that we can reliably characterise the interactions of thousands of untagged and natively expressed proteins directly from cellular lysates. Analysis of five individual biological replicates demonstrates the reproducibility of these methods to characterise protein elution profiles on a proteome wide scale. Through the pair-wise analysis of five replicates for each of the SEC columns tested, we could also detect outliers where deterioration of chromatography and/or some error in sample preparation occurred, and we could discount them from further analysis (Figures 3.11, 3.12). We would, therefore, advocate that researchers perform such pair-wise quality control analysis of datasets prior to counting data.

3.4.1 Utility of orthogonal modes of chromatography

I utilised two orthogonal modes of separation to fractionate *T. brucei* lysates, *i.e.*, by charge (SAX), or size/shape (SEC). For the latter I used two columns with different size/shape separation ranges (SEC300 and SEC1000). Through combining

evidence of co-elution of individual proteins from multiple forms of fractionation, we can have higher confidence in our protein-protein interaction predictions.

Each of the methods of chromatography used in this study have their own strengths and weaknesses with respect to their contributions to our protein complex predictions. One of the main benefits of SEC is that almost any buffer can be utilised for lysis and fractionation, and that the chemical properties of the buffer can be kept constant throughout chromatography. It is possible, therefore, lyse cells and fractionate lysates in a buffer such as PBS, which is similar in pH and ionic strength to the intracellular milieu, and thus limit the disruption of the protein-protein interactions we are trying to capture. Another benefit of SEC columns, is the ability to calibrate them with molecular weight markers (Figures 3.2, 3.3), which allows an estimation of the molecular weight of eluting proteins through regression analysis (Figure 3.4). The elution of a protein during SEC therefore informs us of the molecular weight of that proteoform (Figure 3.8). I have used this estimated molecular weight information in our SEC300 dataset to classify proteins into monomers and dimers, protein complexes and void material (i.e. too large to be separated by SEC300), providing a global overview of the protein interaction state in PBS based lysates of procyclic *T. brucei* (Figure 3.9). The weaknesses of SEC include its relatively low resolution, its relatively high elution volumes (requiring sample concentration post-chromatography) and its ultimate limitation in fractionation range. The latter can be ameliorated by using more than one fractionation range SEC column, as in this study with SEC300 and SEC1000 columns. Although I did not use density-gradient centrifugation in our study, this technique has many of the same strengths and weaknesses as SEC, with the added complication of the presence of the density-gradient forming solute in all fractions.

The advantages of ion-exchange chromatography (in our case SAX) include its relatively high resolution and relatively low elution volumes. The high-resolution of SAX chromatography can be seen in the elution profiles of the protein standards used to characterise the columns (Figures 3.2, 3.3, 3.5) and the heat-maps of elution profiles from all proteins detected by mass spectrometry (Figure 3.14). The proteasome core complex, for example, co-elutes consistently across SEC and SAX chromatography. However, in SEC the elution peak is very broad spanning six fractions whereas it spans only three in SAX (Figure 3.15). Its key weaknesses are that fractionation pH (in our case pH 9.0) and ionic strength (which varies from low to very high across the NaCl elution gradient) are non-physiological and can promote dissociation of part of or all protein complexes in the absence of chemical cross-linking. For example, although components of the proteasome regulatory subunit co-elute together in SEC, when separated by SAX a more complex elution pattern is observed at later stages of the gradient where there are high salt concentrations, indicating breaking apart of the complex (Figure 3.15).

The elution profiles of the proteasome core and regulatory subunits, demonstrates the advantage of using multiple forms of chromatography and lysis conditions, when as in this case, both forms of fractionation provide reinforcing evidence for protein interactions. However, individual protein complexes have a diverse range of required conditions necessary for their stability, and by using orthogonal lysis conditions and methods of chromatography, a wider range of cellular complexes that may only be stable under certain conditions are captured. Other forms of chromatography that could be considered for future studies include hydrophobic interaction chromatography, heparin affinity columns and a wide variety of anion and cation exchange columns.

3.4.2 Global analysis of protein interaction state

We estimate that 44% of the 3,326 proteoforms detected in our SEC300 dataset are present in monomeric or dimeric states, a large proportion of the proteome. This may be a true indication of how many proteins are present in the monomer/dimer pool within the cell, however, it cannot be discounted that some protein complexes may fall apart during cell lysis and sample chromatography, or that there may be some proteolysis when cells are lysed in PBS, even with the addition of protease inhibitors.

The optimal fractionation range (between the void and the monomer and dimer regions) of the SEC300 column contains 20% of detected proteoforms. GO enrichment analysis of the proteins eluting in these fractions highlights the detection of many annotated large macro-molecular complexes, validating that we can capture protein complexes using this method (Figure 3.10).

The remaining 36% of proteoforms are in material too large to be separated by SEC300. Proteins eluting in higher molecular weight fractions are enriched in GO terms associated with integral membrane and organellar proteins. This effect is likely due to presence of membranous vesicles, which have a large hydrodynamic radius, and hence elute in the void together with membrane proteins. We would therefore recommend that future PCP-MS studies, utilising SEC fractionation, incorporate ultracentrifugation of lysates prior to chromatography to remove vesicles and organelle fragments.

It would be interesting to compare the values reported here with future studies incorporating *in vivo* chemical crosslinking prior to lysis and fractionation. The physical crosslinking of proteins may prevent complexes dissociating into sub-complexes or monomeric subunits, providing a more realistic picture of how many proteins are truly monomeric in the cellular environment. The covalent association of proteins in

complexes would further allow lysis in detergent based buffers, solubilising vesicles, organelles and membrane protein complexes, increasing their coverage in the dataset and reducing the complexity of the void material (Larance et al., 2016).

3.4.3 Comprehensive characterisation of known complexes

Using machine-learning techniques, we have assessed the likelihood of all proteins within the dataset to interact in an exhaustive pairwise manner and predicted the formation of 234 protein complexes, comprising 805 proteins (Figure 3.21, Supplemenatry Tables 2 and 3). We have demonstrated the ability of this method to predict interaction between most, if not all, components of a number of highly abundant and well characterised protein complexes. For example, all fourteen proteins of the core 20S proteasome subunit are grouped together using this method, as all components co-elute in all three fractionation experiments. Other predicted complexes capture AMPK (complex 3), vacuolar ATP synthase (complex 20), the exosome (complex 28), prefoldin (complex 29), sub-components of the spliceosome (complexes 86 and 180), the nucleosome (complex 87), ARP2/3 (complex 112), T-complex (complex 129) and F_1F_0 ATP synthase (complex 130), among others detailed in Supplementary Tables 2 and 3. This validates the concept that by scoring how well proteins co-elute, we can predict genuine protein-protein interactions. This dataset of predicted interactions produced here is a powerful tool as a hypothesis generator; predicting interactions in a pairwise fashion across the dataset, which can then be directly tested or confirmed by orthogonal techniques such as immunoprecipitation or yeast-2-hybrid experiments.

This dataset can also be used to increase confidence in the annotation of the trypanosomatid proteome. For example, although most of the proteins detected in the proteasome complex (complex 31, Table 3.1) are annotated as proteasome alpha and beta subunits, one is annotated as ‘unspecified product’ (Tb927.9.11310). A bespoke

BLASTp search with this gene product reveals that it has homology to a proteasome beta subunit. Thus, the co-chromatograph of Tb927.9.11310 with the proteasome core complex has led to a re-evaluation of its likely identity.

Table 3.1: List of proteins identified in Complex 31 – proteasome core complex

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search
Tb11.v5.0196	Proteasome subunit A N-terminal signature/Proteasome subunit, putative	20S proteasome subunit alpha-6 (2.5e-141)	Proteasome subunit A N-terminal signature (IPR000426, 1.3e-13); Proteasome subunit (IPR001353, 1.4e-53)
Tb927.10.230	proteasome subunit alpha type-5, putative	20S proteasome subunit alpha-5 (3.6e-126)	Proteasome subunit A N-terminal signature (IPR000426, 2e-13); Proteasome subunit (IPR001353, 6.7e-57)
Tb927.10.290	proteasome alpha 2 subunit, putative	20S proteasome subunit alpha-2 (3.5e-121)	Proteasome subunit A N-terminal signature (IPR000426, 1.5e-9); Proteasome subunit (IPR001353, 2.7e-63)
Tb927.10.4710	20S proteasome subunit, proteasome subunit beta type-2, putative (PSB4)	20S proteasome subunit beta-4 (6.8e-109)	Proteasome subunit (IPR001353, 1e-27)
Tb927.10.6080	proteasome subunit beta type-5, putative, proteasome subunit beta type-5, putative	Proteasome subunit beta type-5 (2.8e-71)	Proteasome subunit (IPR001353, 2.2e-50)
Tb927.11.7020	proteasome alpha 7 subunit, putative (PSA4)	20S proteasome subunit alpha-4 (5.1e-127)	Proteasome subunit A N-terminal signature (IPR000426, 1.3e-14); Proteasome subunit (IPR001353, 4.8e-58)
Tb927.11.7270	proteasome beta 3 subunit, putative (PSB3)	20S proteasome subunit beta-3 (5.5e-107)	Proteasome subunit (IPR001353, 1e-39)
Tb927.3.780	proteasome alpha 7 subunit (TbPSA7)	20S proteasome subunit alpha-7 (1.6e-59)	Proteasome subunit A N-terminal signature (IPR000426, 5.7e-11); Proteasome subunit (IPR001353, 5.6e-44)
Tb927.4.430	proteasome beta 7 subunit	Proteasome subunit beta type-4 (1.2e-40)	Proteasome subunit (IPR001353, 7.2e-36)
Tb927.6.1260	proteasome beta-1 subunit, putative (PSB1)	Proteasome subunit beta type-6 (8.6e-54)	Proteasome subunit (IPR001353, 1e-42)
Tb927.7.4420	proteasome alpha 3 subunit, putative	20S proteasome subunit alpha-3 (4.5e-64)	Proteasome subunit A N-terminal signature (IPR000426, 1.3e-12); Proteasome subunit (IPR001353, 6.4e-48)
Tb927.7.4790	proteasome beta 6 subunit, 20S proteasome beta 6 subunit, putative (BETA6)	20S proteasome subunit beta-6 (2.6e-139)	Proteasome subunit (IPR001353, 2.2e-32)
Tb927.9.11310	unspecified product	Proteasome subunit beta type-2 (2.7e-73)	Proteasome subunit (IPR001353, 8.9e-47)
Tb927.9.9670	proteasome alpha 1 subunit, putative, 20S proteasome subunit alpha-6, (putative) (TbPSA6)	Proteasome subunit alpha type-1 (4e-120)	Proteasome subunit A N-terminal signature (IPR000426, 1.2e-11); Proteasome subunit (IPR001353, 1.7e-38)

Likewise, complex 130 (Table 3.2) contains all characterised subunits (α, β, γ, δ and ε) of the F_1 domain of the F_0F_1-ATP synthase complex (Zikova et al., 2009). Riboonucleoprotein p18 is also observed in this predicted complex, a protein also

demonstrated to be the b subunit of the F_o domain (Zikova et al., 2009). Complex 130 contains two other proteins (Tb927.11.13070 and Tb927.3.3410) with no prior evidence to indicate functional interaction with the ATP synthase complex. Additionally, complex 85 comprises three hypothetical proteins, two of which are also detected in the *T. brucei* ATP synthase (Table 3.3). Tb927.10.8030 is thought to be the homolog of the oligomycin sensitivity conferring protein of the F_o domain, and Tb927.11.6250 is a trypanosome specific ATP synthase associated protein (Zikova et al., 2009). The third and final protein, Tb927.5.1780, has no other functional information and may be a novel component of ATP synthase.

Table 3.2: List of proteins identified in Complex 130 – F_0F_1-ATP synthase complex

Gene ID	GeneDB Annotation	Homology search	BLAST search	Interpro domain search
Tb927.10.180	ATP synthase F1 subunit gamma protein, putative	ATP synthase gamma chain (2.1e-20)	ATP synthase (IPR000131, 4.7e-33); ATP synthase (F1-ATPase), gamma subunit (IPR023633, 4.5e-49)	Classed as gamma subunit of F_0F_1 ATP synthase in (Zikova et al., 2009)
Tb927.10.5050	Mitochondrial ATP synthase epsilon chain, putative	-	Epsilon subunit of mitochondrial F1F0-ATP synthase (IPR006721, 9.7e-13)	Classed as epsilon subunit of F_0F_1 ATP synthase in (Zikova et al., 2009)
Tb927.11.13070	O-phosphoseryl-tRNA(Sec) selenium transferase, putative	Selenocysteinyl-tRNA(Sec) synthase (9.4e-100)	Soluble liver antigen/liver pancreas antigen (IPR008829, 9.2e-59)	-
Tb927.3.1380	ATP synthase subunit beta, mitochondrial, ATP synthase F1, beta subunit (ATPB)	ATP synthase subunit beta (1.3e-171)	ATP synthase alpha/beta family, nucleotide-binding domain (IPR000194, 5.7e-64)	Classed as beta subunit of F_0F_1 ATP synthase in (Zikova et al., 2009)
Tb927.3.3410	aspartyl aminopeptidase, putative, metallo-peptidase, Clan MH, Family M20	Aspartyl aminopeptidase (3.3e-111)	-	-
Tb927.5.1710	ribonucleoprotein p18, mitochondrial precursor, putative	Protein P18 (1.4e-76)	-	Classed as b subunit of F_0F_1 ATP synthase in (Zikova et al., 2009)
Tb927.6.4990	ATP synthase, epsilon chain, putative	ATP synthase subunit delta (6.7e-15)	ATP synthase, Delta/Epsilon chain, beta-sandwich domain (IPR020546, 3.1e-17)	Classed as delta subunit of F_0F_1 ATP synthase in (Zikova et al., 2009)
Tb927.7.7430	ATP synthase alpha chain, mitochondrial precursor, ATP synthase F1, alpha subunit	ATP synthase subunit alpha (1.3e-107)	ATP synthase alpha/beta family, nucleotide-binding domain (IPR000194, 4.8e-69)	Classed as alpha subunit of F_0F_1 ATP synthase in (Zikova et al., 2009)

Table 3.3: List of proteins identified in Complex 85 – F_0F_1-ATP synthase complex

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search	Comments
Tb927.10.8030	hypothetical protein, conserved	-	ATP synthase delta (OSCP) subunit (IPR000711, 2.6e-5)	Classed as OSCP subunit of F_0F_1 ATP synthase in (Zikova et al., 2009)
Tb927.11.6250	hypothetical protein, conserved	-	-	Classed as trypanosome specific subunit of F_0F_1 ATP synthase in (Zikova et al., 2009)
Tb927.5.1780	hypothetical protein, conserved	-	-	-

Previous AP-MS analysis of the mitochondrial ribosome of *T. brucei* identified 133 proteins, 77 of which were classed as large-subunit and 56 as small-subunit associated proteins (Zikova et al., 2008). From our data analysis we predict three groups of complexes (complex 4, 92 and 134, Tables 3.4-3.6) which are composed solely of proteins identified from this previous publication, apart from one hypothetical protein (Tb927.7.3030) in complex 134. Within these three complexes 26 mitochondrial ribosomal proteins are identified, and one potentially novel component. Of these, 15 are identified as hypothetical proteins, which with prior evidence from AP-MS studies and this work may warrant re-classification as mitochondrial ribosome subunits.

Table 3.4: List of proteins identified in Complex 4 – mitochondrial ribosome

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search	Comments
Tb927.11.12930	DEAD-box helicase, putative	ATP-dependent rRNA helicase RRP3 (2e-30)	DEAD/DEAH box helicase (IPR011545, 3.8e-26)	-
Tb927.11.5990	hypothetical protein, conserved	Pseudouridylate synthase 7 homolog-like protein (6.9e-9)	Pseudouridine synthase (IPR020103, 8.8e-35)	mitochondrial LSU protein (Zikova et al., 2008)
Tb927.4.1070	50S ribosomal protein L13, putative	50S ribosomal protein L13 (6.5e-17)	Ribosomal protein L13 (IPR023564, 2e-30)	MRPL13 (Zikova et al., 2008)
Tb927.6.4080	hypothetical protein, conserved	-	-	mitochondrial SSU or LSU protein (Zikova et al., 2008)
Tb927.7.1640	ras-like small GTPase, putative (TbEAR)	GTPase Der (1.5e-56)	50S ribosome-binding GTPase (IPR006073, 1.8e-15)	mitochondrial LSU protein (Zikova et al., 2008)
Tb927.7.3460	hypothetical protein, conserved	-	-	mitochondrial LSU protein, KRIT2 (Zikova et al., 2008)

Table 3.5: List of proteins identified in Complex 92 – mitochondrial ribosome

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search	Comments
Tb927.10.3580	hypothetical protein, conserved	-	-	mitochondrial SSU protein (Zikova et al., 2008)
Tb927.11.2530	Mitochondrial SSU ribosomal protein, putative	-	-	mitochondrial SSU protein (Zikova et al., 2008)
Tb927.7.3050	hypothetical protein, conserved	-	-	mitochondrial SSU protein (Zikova et al., 2008)

Table 3.6: List of proteins identified in Complex 134 – mitochondrial ribosome

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search	Comments
Tb927.1.1200	SSU ribosomal protein, mitochondrial (MRPS15)	-	S15/NS1 RNA-binding domain (IPR009068, 1.9e-7)	MRPS15 (Zikova et al., 2008)
Tb927.10.7380	hypothetical protein, conserved	-	-	mitochondrial LSU protein (Zikova et al., 2008)
Tb927.11.11630	hypothetical protein, conserved	-	-	mitochondrial LSU protein (Zikova et al., 2008)
Tb927.11.1250	Mitochondrial SSU ribosomal protein, putative	-	-	mitochondrial SSU protein (Zikova et al., 2008)
Tb927.11.6000	ribosomal protein L4/L1 family, putative	50S ribosomal protein L4 (1.3e-6)	Ribosomal protein L4 (IPR023574, 4.6e-40)	MRPL4 (Zikova et al., 2008)
Tb927.11.870	hypothetical protein, conserved	-	Ribosomal proteins S24e, L23 and L15e (IPR012678, 7.6e-9)	MRPL23 (Zikova et al., 2008)
Tb927.2.4890	ribosomal protein L11, putative	54S ribosomal protein L19, mitochondrial (1.3e-13)	Ribosomal protein L11/L12 (IPR000911, 1.2e-9)	MRPL11 (Zikova et al., 2008)
Tb927.3.5610	ribosomal protein L3 mitochondrial, putative	39S ribosomal protein L3, mitochondrial (2e-27)	Ribosomal protein L3 (IPR000597, 1.2e-11)	MRPL3 (Zikova et al., 2008)
Tb927.5.3980	hypothetical protein, conserved	54S ribosomal protein L10, mitochondrial (1.9e-8)	Ribosomal proteins L15p and L18e (IPR021131, 3.1e-9)	MRPL15 (Zikova et al., 2008)
Tb927.5.4120	hypothetical protein, conserved	-	-	mitochondrial LSU protein (Zikova et al., 2008)
Tb927.6.2080	hypothetical protein, conserved	-	-	mitochondrial SSU protein (Zikova 2008)
Tb927.6.4560	hypothetical protein, conserved	-	-	mitochondrial SSU protein (Zikova et al., 2008)
Tb927.7.2760	ribosomal protein L22p/L17e, putative	50S ribosomal protein L22 (7.6e-9)	Ribosomal protein L22 (IPR001063, 6.4e-20)	MRPL22 (Zikova et al., 2008)
Tb927.7.3030	hypothetical protein, conserved	-	-	
Tb927.7.4140	ribosomal protein L21, putative	Probable 39S ribosomal protein L21, mitochondrial (9.8e-14)	Ribosomal prokaryotic L21 protein (IPR001787, 7e-20)	MRPL21 (Zikova et al., 2008)
Tb927.8.5200	hypothetical protein, conserved	-	-	mitochondrial SSU protein (Zikova et al., 2008)
Tb927.9.7170	Mitochondrial 39-S ribosomal protein L47 (MRP-L47), putative	39S ribosomal protein L47, mitochondrial (1.2e-13)	Mitochondrial 39-S ribosomal protein L47 (IPR010729, 2.3e-25)	MRPL47 (Zikova et al., 2008)
Tb927.9.8290	hypothetical protein, conserved	-	Ribosomal protein L30p/L7e (IPR016082, 9.9e-12)	MRPL30 (Zikova et al., 2008)

Complex 3 contains both experimentally verified members (Tb927.8.2450 and Tb927.10.3700, β and γ subunits respectively) of the AMPK complex (Table 3.7) (Clemmens et al., 2009). The three other proteins include Tb927.3.4560, annotated as the AMPKα subunit; Tb927.10.5310, an SNF1 related protein kinase, also with homology to AMPKα; and Tb927.9.9270, a hypothetical protein with little extra functional information. AMPK is thought to be a heterotrimeric complex, therefore it is possible that the two putative AMPKα subunits are isoforms, forming part of two independent protein complexes which co-elute. Whether the hypothetical protein is a novel component of the trypanosome AMPK complex, is a subunit isoform, or is simply a contaminating co-eluted protein warrants further investigation.

Table 3.7: List of proteins identified in Complex 3 - AMPK

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search	Comments
Tb927.10.3700	AMP-activated protein kinase, gamma regulatory subunit, SNF1-related protein kinase regulatory subunit gamma, AMPK subunit gamma (AMPKG)	AMPK subunit gamma-3 (2.2e-14)	-	identified in (Clemmens et al., 2009)
Tb927.10.5310	SNF1-related protein kinases, putative	AMPK subunit alpha-1 (3.1e-77)	-	-
Tb927.3.4560	5'-AMP-activated protein kinase catalytic subunit alpha, putative, AMPK subunit alpha, putative, SNF1-related protein kinase catalytic subunit alpha, putative	AMPK subunit alpha-2 (4.2e-90)	-	-
Tb927.8.2450	SNF1-related protein kinase regulatory subunit beta, 5'-AMP-activated protein kinase subunit beta, AMPK subunit beta (AMPKB)	AMPK subunit beta-1 (1.1e-10)	5'-AMP-activated protein kinase beta subunit, interation domain (IPR006828, 9.6e-41)	identified in (Clemmens et al., 2009)
Tb927.9.9270	hypothetical protein, conserved	Ankyrin repeat and zinc finger domain-containing protein 1 (1.6e-12)	-	-

3.4.4 Discovery of novel interactions and functional information

Our machine learning based interactome dataset predicts a number of complexes and interactions that have not previously been observed in *T. brucei* (Figures 3.25 and 3.26, Tables 3.8-3.15). These data have: (a) revealed putative functional information for 'hypothetical' proteins that are partnered with proteins of known function; (b) identified evolutionary conserved complexes characterised in other eukaryotes; (c) predicted novel

interactions between proteins of known function. Some examples of each are given below.

(a) We demonstrate the consistent co-elution of the RNA binding protein PUF10 with a hypothetical protein (Tb927.7.2170) across SEC300, SEC1000 and SAX fractionation in complex 72 (Table 3.8, Figure 3.25a). PUF proteins are known to bind to mRNA, and the hypothetical protein has also been predicted to bind mRNA through capture on oligo(dT) beads (Lueong et al., 2016). Taken together, these data indicate that Tb927.7.2170 interacts with PUF10 and may play a role in mRNA regulation. This outcome displays the power of PCP-MS, as there are now multiple lines of evidence for this interaction and it provides insights into the role of this 'hypothetical' protein as a mRNA binder, working in conjunction with PUF10.

Table 3.8: List of proteins identified in Complex 72 – PUF10 complex

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search	Comments
Tb927.11.6740	pumilio/PUF RNA binding protein 10, putative	-	ARM repeat (IPR016024, 2.4e-19)	mRNA binding (Lueong et al., 2016)
Tb927.7.2170	hypothetical protein, conserved	ANP32/acidic nuclear phosphoprotein-like protein 2 (3.3e-10)	-	mRNA binding (Lueong et al., 2016)

In complex 174 two proteins are detected, a hypothetical protein (Tb927.8.1960) together with subunit 10 of the CCR4-NOT complex (Tb927.10.8720)(Table 3.9, Figure 3.25b) (Schwede et al., 2008). The hypothetical protein has recently been co-purified with CAF1, a core component of the CAF1-NOT deadenylase complex (Färber et al., 2013). It is also the homologue of a human protein (C2ORF29) recently classified as subunit 11 of the CCR4-NOT complex that interacts with subunit 10 (Mauxion et al., 2013). The co-chromatography with CCR4-NOT subunit 10

provides further evidence for this functional classification of the hypothetical protein.

Table 3.9: List of proteins identified in Complex 174 – CNOT10 and 11 complex

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search	Comments
Tb927.10.8720	CCR4-NOT transcription complex subunit 10, putative	-	-	Identified in Caf1 pulldown (Schwede et al., 2008)
Tb927.8.1960	hypothetical protein, conserved	-	-	Copurified with Caf1 protein of CNOT complex, and Tb927.10.8720 and is human homolog of CNOT11 (Färber et al., 2013; Mauxion et al., 2013; Schwede et al., 2008)

Out of the 234 complexes predicted from machine-learning, 77 contain at least one protein annotated as 'hypothetical' and 19 are composed solely of 'hypothetical' proteins with no other orthogonal information on protein function.

(b) The trypanosome orthologue of Pwp2 (Tb927.10.13270) co-elutes with Tb927.11.10480, Tb927.11.460 and Tb927.7.4220, predicted to contain C-terminal Utp21, Utp13 and Utp12 domains respectively, in complex 108 (Table 3.10, Figure 3.25c). Work in yeast has shown that Pwp2 is known to associate with four other proteins (containing the same C-terminal Utp domains) in the U3 ribonucleoprotein assembly machinery, forming a complex necessary for pre-18S rRNA processing (Dosil and Bustelo, 2004). We therefore postulate that complex 108 performs a similar pre-18S rRNA processing function in *T. brucei*.

Table 3.10: List of proteins identified in Complex 108 – U3 ribonucleoprotein

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search
Tb927.10.13270	Periodic tryptophan protein 2 homolog, putative	Periodic tryptophan protein 2 homolog (2.2e-116)	WD40 repeat-like (IPR017986, 2.9e-52)
Tb927.11.10480	PQQ-like domain/WD domain, G-beta repeat/Utp21 specific WD40 associated putative domain containing protein, putative	U3 snoRNA-associated protein 21 (2.9e-21)	Utp21 specific WD40 associated putative domain (IPR007319, 4e-19)
Tb927.11.460	predicted WD40 repeat protein	U3 snoRNA-associated protein 13 (2.2e-51)	Utp13 specific WD40 associated domain (IPR013934, 2.3e-42)
Tb927.7.4220	WD domain, G-beta repeat/Dip2/Utp12 Family, putative	U3 snoRNA-associated protein 12 (1.6e-54)	Dip2/Utp12 Family (IPR007148, 2.6e-6)

Additionally, in complex 76 three proteins are detected: Tb927.10.170 (pseudouridine synthase, Cbf5p), Tb927.4.470 (snoRNP protein, GAR1 putative) and Tb927.4.750 (50S ribosomal protein L7Ae, putative) (Table 3.11, Figure 3.25d). Cbf5 is the enzymatic component of the H/ACA ribonucleoprotein complex which pseudouridylates target RNAs. There are over 100 H/ACA ribonucleoprotein complexes, formed through interaction of different RNAs with the same four core proteins (Cbf5, Gar1, Nhp2 and Nop10)(Meier, 2006). The co-elution of these three proteins provides evidence for the existence of this complex in trypanosomatids and confirms the putative identity of Tb927.4.470 as a Gar1 homolog. A BLAST-p search of the putative ribosomal subunit, Tb927.4.750 also indicates homology to Nhp2, further confirming the identity of this complex. A search of TriTrypDB indicates that there is one annotated Nop10 homolog (Tb927.10.4740) in *T. brucei*. Although this protein is not detected in our high-confidence protein complex prediction, using the "Profile Explorer" data visualisation tool (described in section *3.4.7*), we can see this protein is detected in our SAX fractionation experiment, and perfectly co-elutes with the three other components of complex 76.

Table 3.11: List of proteins identified in Complex 76 – H/ACA ribonucleoprotein

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search
Tb927.10.170	pseudouridine synthase, Cbf5p	H/ACA ribonucleoprotein complex subunit 4 (9.6e-142)	-
Tb927.4.470	snoRNP protein GAR1, putative	snoRNP protein GAR1 (8.6e-54)	Gar1/Naf1 RNA binding region (IPR007504, 1.4e-38)
Tb927.4.750	50S ribosomal protein L7Ae, putative	snoRNP protein NHP2 (4e-24)	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family (IPR004038, 2.1e-17)

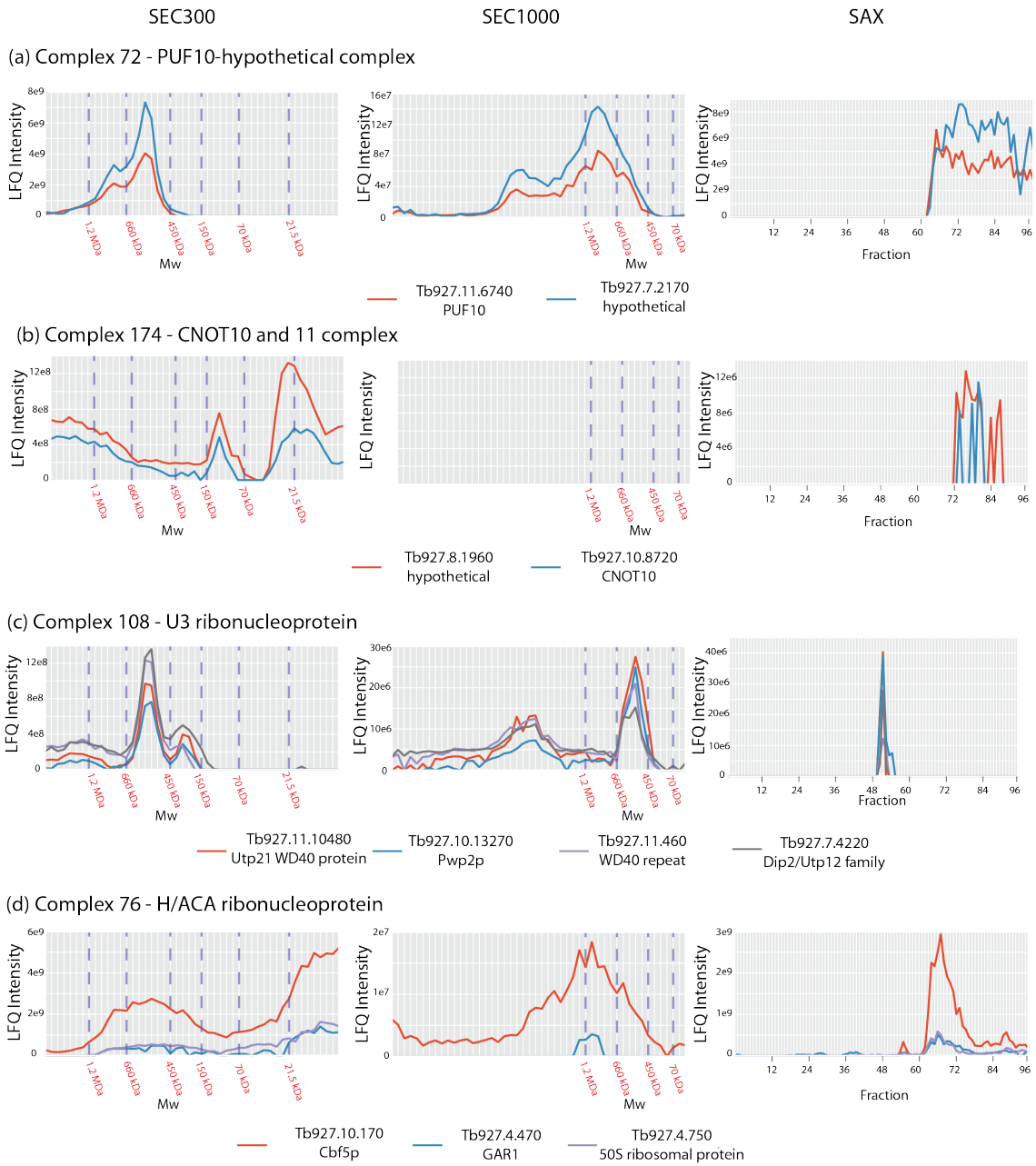


Figure 3.25: Elution profiles of components of novel predicted complexes across multiple forms of chromatography. LFQ intensity for each protein plotted across the fractionation range collected for mass spectrometry analysis. Molecular weight marker retention times on SEC300 and SEC1000 columns marked below the plot.

(c) Previous published studies have demonstrated the constitutive interaction of a heat-shock protein (HSP) 90 with protein phosphatase 5 (PP5) in *T. brucei* (Jones et al., 2008). Complex 12 contains five proteins, including an HSP90 (Tb927.3.3580), and the PP5 previously demonstrated to interact with HSP90 (Tb927.1013670) (Table 3.12, Figure 3.26a). The association of these proteins with a putative HSP70 protein (Tb927.9.9860) can also be observed. HSP90 chaperones function through the association with HSP70 proteins, which recruit and transfer substrate proteins to HSP90 (Folgueira and Requena, 2007). This novel association therefore matches our understanding of HSP90 function, and identifies a putative function for a previously uncharacterised HSP70.

Table 3.12: List of proteins identified in Complex 12 – HSP70/90 Complex

Gene ID	GeneDB Annotation	Homology search	BLAST search	Interpro domain search
Tb927.10.13670	serine/threonine protein phosphatase 5	Serine/threonine-protein phosphatase 5 (3.2e-132)	-	interacts with HSP90 (Jones et al., 2008)
Tb927.10.14030	hypothetical protein, conserved	-	-	-
Tb927.3.3580	heat shock protein 90, putative (LPG3)	HSP90.2 (7.8e-132)	ATPase domain of HSP90 chaperone/DNA topoisomerase II/histidine kinase (IPR003594, 8.6e-61)	-
Tb927.5.3260	WD domain, G-beta repeat, putative	WD repeat-containing protein 17 (4.7e-58)	WD40 repeat-like (IPR017986, 2.6e-46)	-
Tb927.9.9860	Hsp70 protein, putative	Hsp70-17 (8.6e-68)	Hsp70 protein (IPR013126, 1.6e-67)	-

In complex 99 the association of proteins from two distinct protein complexes are seen (Table 3.13, Figure 3.26b). Two proteins have been experimentally verified as members of the spliced leader RNA cap methyltransferase (Zamudio et al., 2009), including the enzymatic MTR1 subunit and a hypothetical protein (Tb927.10.7940 and Tb927.11.16490). The three other proteins in this predicted complex are identified as translation elongation factors (Tb927.10.5840, Tb927.11.13190 and

Tb927.4.3590) indicating an association between the methyltransferase capping enzymes of spliced leader RNA and the mRNA translation machinery.

Table 3.13: List of proteins identified in Complex 99 – MTR1 and eIFB Complex

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search	Comments
Tb927.10.5840	translation elongation factor 1-beta, putative	EF-1-beta (1.8e-62)	eEF-1beta-like (IPR014038, 7.7e-30)	-
Tb927.10.7940	methyltransferase, putative	MTr1 (4.8e-202)	FtsJ-like methyltransferase (IPR002877, 1.3e-30)	-
Tb927.11.13190	elongation factor 1 gamma, putative	eEF-1B gamma (8.3e-175)	eEF1-gamma domain (IPR001662, 8.4e-63)	-
Tb927.11.16490	hypothetical protein, conserved	Telomerase Cajal body protein 1 (2.4e-9)	WD40 repeat-like (IPR017986, 1.6e-14)	Identified interacting with Mtr1 in (Zamudio et al., 2009)
Tb927.4.3590	translation elongation factor 1-beta, putative	EF-1-beta (3.3e-84)	eEF-1beta-like (IPR014038, 1.3e-29)	-

Personal communication with Laurie Read also confirms predictions identified in complex 165(Table 3.14, Figure 3.26c), a complex dominated by nucleolar associated proteins. Within this complex two arginine-N-methyltransferases (Tb927.1.4690 and Tb927.10.3560) are seen, that are confirmed to form a protein complex, through directed studies in the Read lab.

Table 3.14: List of proteins identified in Complex 165 – nucleolar complex

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search
Tb927.10.11310	intraflagellar transport protein 57/55 (IFT57/55)	Intraflagellar transport protein 57 homolog (7.7e-53)	Intra-flagellar transport protein 57 (IPR019530, 8.3e-125)
Tb927.10.13860	GPI-anchor transamidase subunit 8 (GPI8)	GPI transamidase (1.8e-51)	Peptidase C13 family (IPR001096, 1.1e-31)
Tb927.10.4040	3-keto-dihydrosphingosine reductase	3-ketodihydrosphingosine reductase (1.9e-33)	short chain dehydrogenase (IPR002198, 6.7e-35)
Tb927.10.4610	dolicholphosphate-mannose synthase, putative (DPMS)	Dolichol-phosphate mannosyltransferase (2e-61)	Glycosyl transferase family 2 (IPR001173, 4.3e-33)
Tb927.11.13820	hypothetical protein, conserved	-	-
Tb927.11.15760	GPI transamidase subunit Tta1 (TTA1)	-	-
Tb927.2.1810	transcription silencer (ISWI)	ISW2-like (1.4e-168)	SNF2 family N-terminal domain (IPR000330, 1.2e-82)
Tb927.5.1930	signal peptidase subunit, putative	-	Signal peptidase subunit (IPR007653, 5.9e-13)
Tb927.8.5760	Ankyrin repeats (many copies)/Alpha/beta hydrolase family, putative	-	-

Complex 164 (Table 3.15, Figure 3.26d) further demonstrates how the methods described in this work can produce novel information on proteins with characterised functions, enhancing our knowledge of their biology. An association of the GPI transamidase with signal peptidase, suggests association of the GPI transamidase with the translocon complex in the endoplasmic reticulum (ER) (Johnson and van Waes, 1999). The suggested colocation of these components is novel, but consistent with the known co-translational addition of GPI anchors to nascent proteins in *T. brucei* (Ferguson et al., 1985).

Table 3.15: List of proteins identified in Complex 164 – GPI associated complex

Gene ID	GeneDB Annotation	BLAST search	Interpro domain search
Tb927.1.4690	arginine N-methyltransferase (PRMT1)	Histone-arginine N-methyltransferase PRMT1 (7.7e-85)	-
Tb927.10.12980	Multisite-specific tRNA:(cytosine-C(5))-methyltransferase, putative	tRNA (cytosine-5-)-methyltransferase NCL1 (1.5e-38)	NOL1/NOP2/sun family (IPR001678, 9.2e-21)
Tb927.10.14750	fibrillarin, putative	rRNA 2-X-methyltransferase fibrillarin (2e-102)	Fibrillarin (IPR000692, 6.7e-103)
Tb927.10.1960	hypothetical protein, conserved	-	ARM repeat (IPR016024, 1.5e-13)
Tb927.10.3560	arginine N-methyltransferase, putative	Probable protein arginine N-methyltransferase 1 (4.8e-37)	-
Tb927.10.7500	fibrillarin (NOP1)	rRNA 2-X-methyltransferase fibrillarin (4.2e-132)	Fibrillarin (IPR000692, 2.2e-106)
Tb927.8.3750	Nucleolar protein 56, putative (NOP56)	Nucleolar protein 56 (2.1e-107)	Putative snoRNA binding domain (IPR002687, 4.1e-56)
Tb927.8.900	splicing factor TSR1 (TSR1)	Splicing factor, arginine/serine-rich 2 (1.7e-18)	-
Tb927.9.5320	nucleolar RNA binding protein, putative	Nucleolar protein 58-2 (6.5e-100)	Putative snoRNA binding domain (IPR002687, 1.6e-52)
Tb927.9.6870	RNA-binding protein, putative (RBSR1)	Probable splicing factor, arginine/serine-rich 6 (7.2e-18)	-

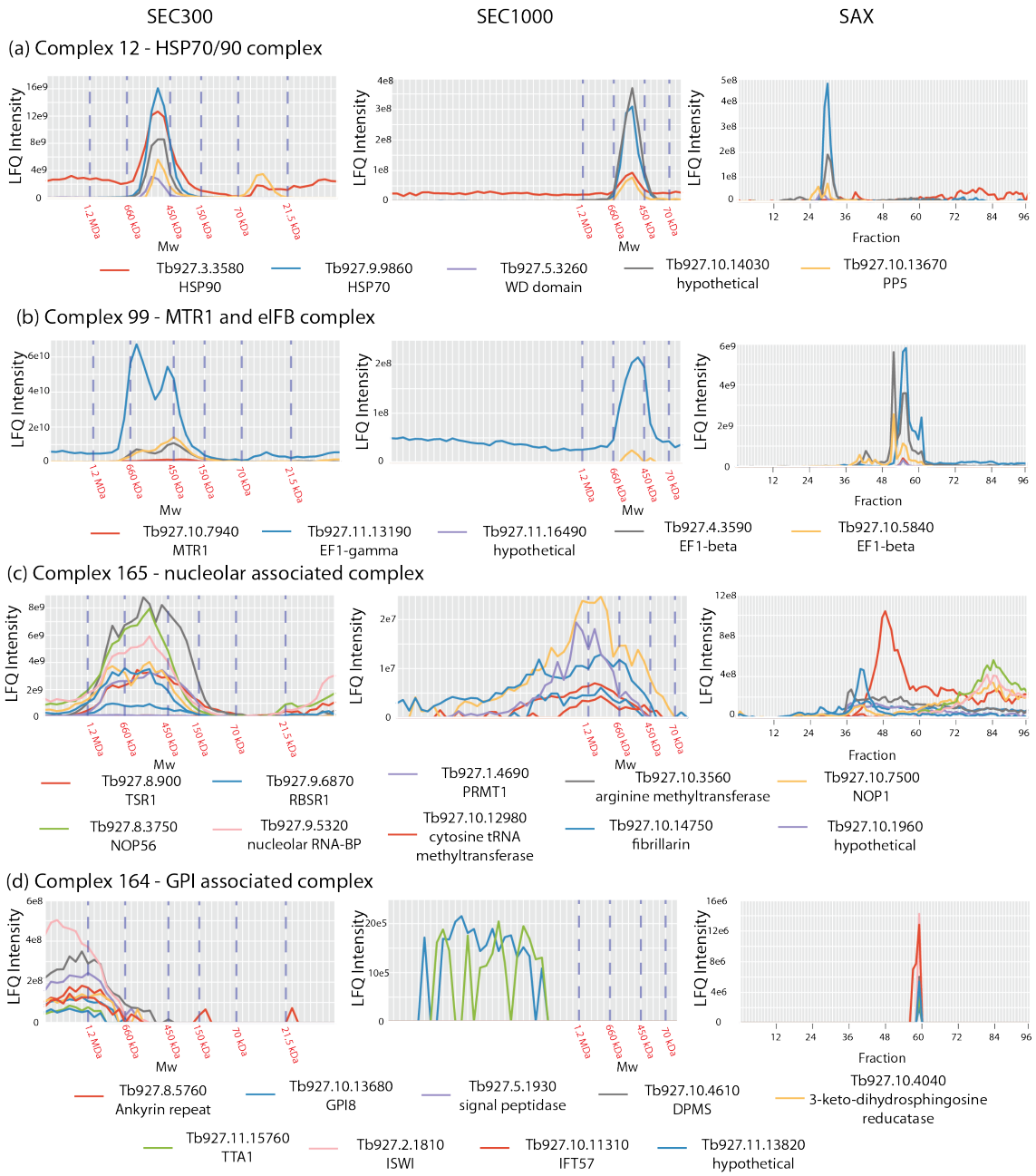


Figure 3.26: Elution profiles of components of predicted protein complexes with characterised functions across multiple forms of chromatography. LFQ intensity for each protein plotted across the fractionation range collected for mass spectrometry analysis. Molecular weight marker retention times on SEC300 and SEC1000 columns marked below the plot.

3.4.5 Co-chromatography complicates protein interaction prediction

The last of these examples also illustrates the limitations of our methodology, which is ultimately dependent on co-chromatography of the component parts of protein complexes. This means that ambiguities are bound to arise when two or more complexes co-chromatograph. It would be expected from our knowledge of the function of the genes described above that complex 164 is found in the ER. However, we also detect proteins with similar elution profiles that are not expected to be resident ER proteins; an intraflagellar transport protein, an ISWI protein and an Ankyrin repeat protein with alpha/beta hydrolase activity. As we are fractionating whole cell lysates, that contain a very complicated mixture of protein complexes that originate from various sub-cellular compartments, such ambiguities are not surprising. Furthermore, the mixing of different subcellular compartments during whole cell lysis might also lead to non-native interactions between proteins and complexes that do not usually see each other. To reduce these complexities, it would be interesting to utilise subcellular fractionation techniques prior to fractionation of proteins in future studies. This would physically separate components that are in different compartments during chromatography and allow for better predictions of protein-protein interactions. Another beneficial side effect would be to increase the coverage of lower abundance proteins in different sub-cellular compartments.

3.4.6 Comparison to other datasets

Figure 3.19 reveals that scoring features derived from data produced from SEC300 and SEC1000 elution profiles, are more highly used than features derived from STRING, or a recently published *T. brucei* interactome, by the random forest predictors trained in this study. A number of factors account for this result. For example, some of the components of the proteasome that co-elute in our datasets are not annotated as

interacting partners in the STRING database (Tb927.11.8310 and Tb927.10.9740). Hence, only in our dataset could information be produced to confirm the association of these component parts of the proteasome, which are backed up by BLASTp homology to subunits in other organisms. Furthermore, the glycerol gradient experiments performed in (Gazestani et al., 2016), have a low number of protein groups identified (1,300 and 593 for whole cell and mitochondrial extracts, respectively). A lower number of identified proteins means a lower number of potential gold standard positive interactors used to train the random forest algorithm, leading to a reduced use of these features to score interactions.

This effect is also seen in the ion-exchange chromatography (IEX) experiments from the same publication. In this case, only 19 (cytosol extract) or 20 (mitochondrial extract) fractions are collected across the chromatographic run. A smaller number of fractions limits the ability to resolve separate protein elution peaks, increasing the chance of detecting co-eluting profiles and predicting false positive interactions. The distribution of Pearson correlation coefficients between all detected proteins in each datasets confirms this result. Comparison of the glycerol gradient and SEC experiments, where 48 to 50 fractions were collected, shows most proteins within each experiment have no correlation between elution profiles, with a small proportion showing high correlation. This is what would be expected, as most proteins do not interact with each other. Focusing on the cytosolic IEX experiment, we see that a lower proportion of elution profiles have no correlation, with a distinct peak of proteins showing very high correlation >0.8. This highlights the need for extensive, high resolution fractionation to be able to distinguish closely co-eluting protein complexes in PCP-MS studies.

3.4.7 Data visualisation

We have made all of the processed data and predictions available to browse on a web server with a user friendly interface developed by Dr Michele Tinti (http://134.36.66.166:8083/complex_explorer), which will allow other researchers to explore and interrogate our predicted complexes and elution profiles. There are three distinct applications with which to browse the data:

The "Complex Explorer" (Figure 3.27) is a dynamic browser of the high-confidence predictions of protein complexes derived from machine learning of exhaustive pairwise comparisons of all protein elution profiles, matching the data presented in Figure 3.21 and Supplementary Table 2. Each cluster is displayed as a network of interconnected nodes (protein groups) with the lines between the nodes indicating evidence of pairwise associations. The stringency (cutting threshold) for the associations can be varied with a slider below the browser and the cluster browser can be queried to highlight individual protein groups and cluster numbers. Mousing over the nodes brings up their GO-terms within a word cloud of the GO-terms for the other nodes in the complex, which can give a general impression of possible cluster function. The dynamic browser can be adjusted through '*settings'* to highlight any or all of those: Nodes with human homologues. Nodes identified as essential in cell culture (Alsford et al., 2011). Nodes that were used as 'gold-standards' for the machine-learning. Clusters which agree with homologous associations in the STRING database (using a relatively high combined STRING score threshold value of >950). Clusters that are inter-related by the same STRING associations. Clusters predicted by those STRING associations alone. Nodes which appear in more than one cluster. Clicking on any node in a cluster brings up a table of the protein group components, alongside the SEC300, SEC1000 and SAX chromatograms for those protein groups. The chromatograms are dynamic and can

be expanded in the x- (time) axis and display either raw LFQ or normalised LFQ intensity data on the y-axis. Further, the colour-coded elution profiles of individual protein groups can be switched on or off by double-clicking on the gene IDs below the chromatograms. Below the dynamic cluster browser is a table of all the nodes in the browser, which can be searched in various ways and downloaded by the user for other applications.

The second application, "Profile Explorer" (Figure 3.28) allows exploration of protein-protein associations based on the machine learning predictions. The application allows the input of a list of up to twenty TriTrypDB geneIDs, and outputs the predicted interactions between them and the associated fractionation data in any of the SEC300, SEC1000 and SAX datasets, as well as in the density gradient and ion exchange fractionations recently published in (Gazestani et al., 2016).

The "Cluster Explorer" (Figure 3.29) application allows exploration of putative protein-protein associations based purely on hierarchical clustering of the protein elution profiles from SEC300, SEC1000 and SAX chromatography (Figure 3.14). These are lower confidence predictions of protein-protein associations than those based on machine learning but they allow the user to ask whether there is any evidence for the *possible* association of two or more proteins by co-chromatography. Thus, the application allows the input of a list of up to twenty TriTrypDB geneIDs, and outputs graphs showing which hierarchical clusters they belong to. From there, selecting the cluster number will display the relevant chromatogram.

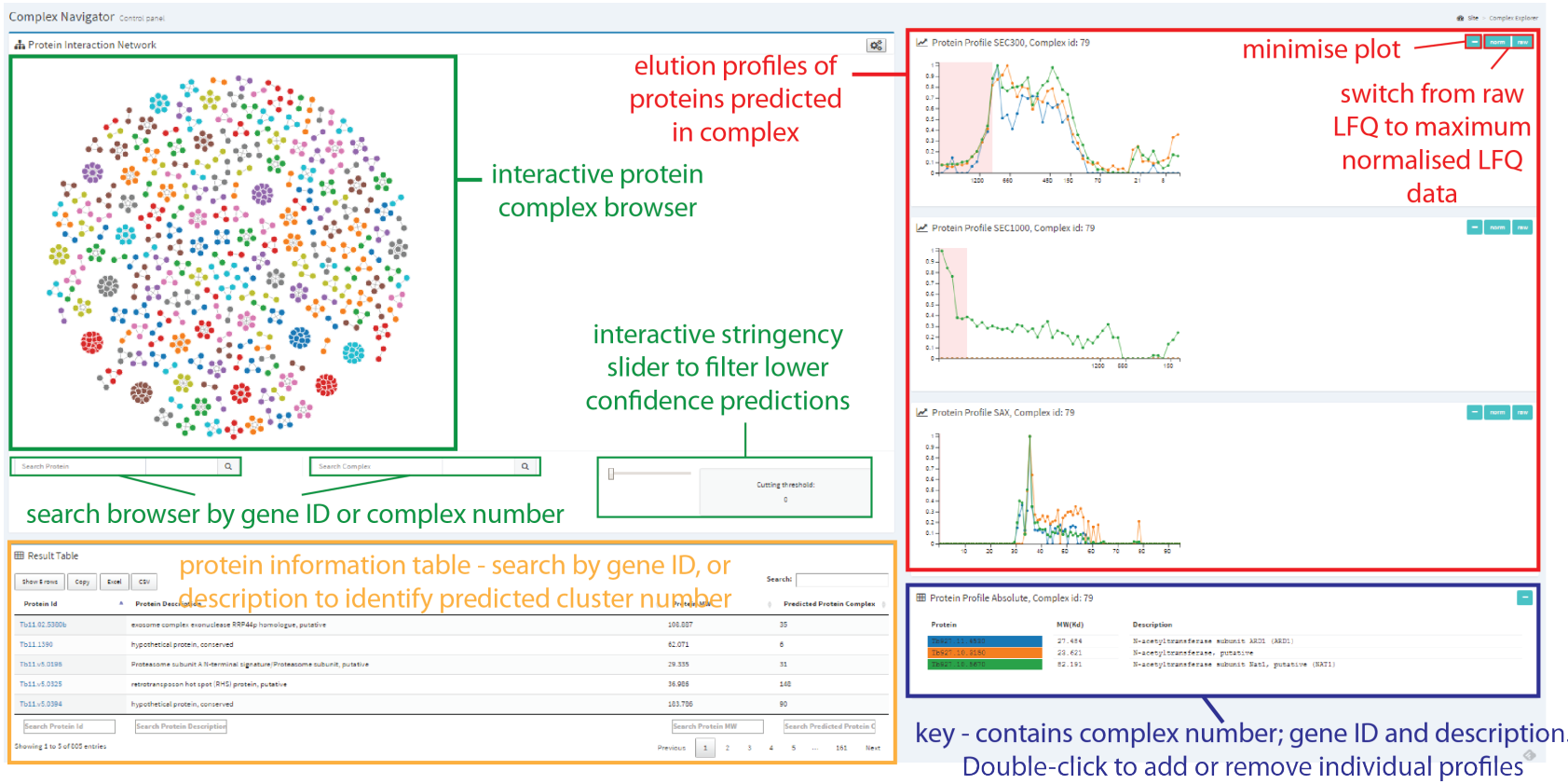


Figure 3.27: Data visualisation tools - Complex Explorer. As described in Section 3.4.7, allows browsing of our predictions of protein complexes from machine learning analysis detailed in Supplementary Table 2.

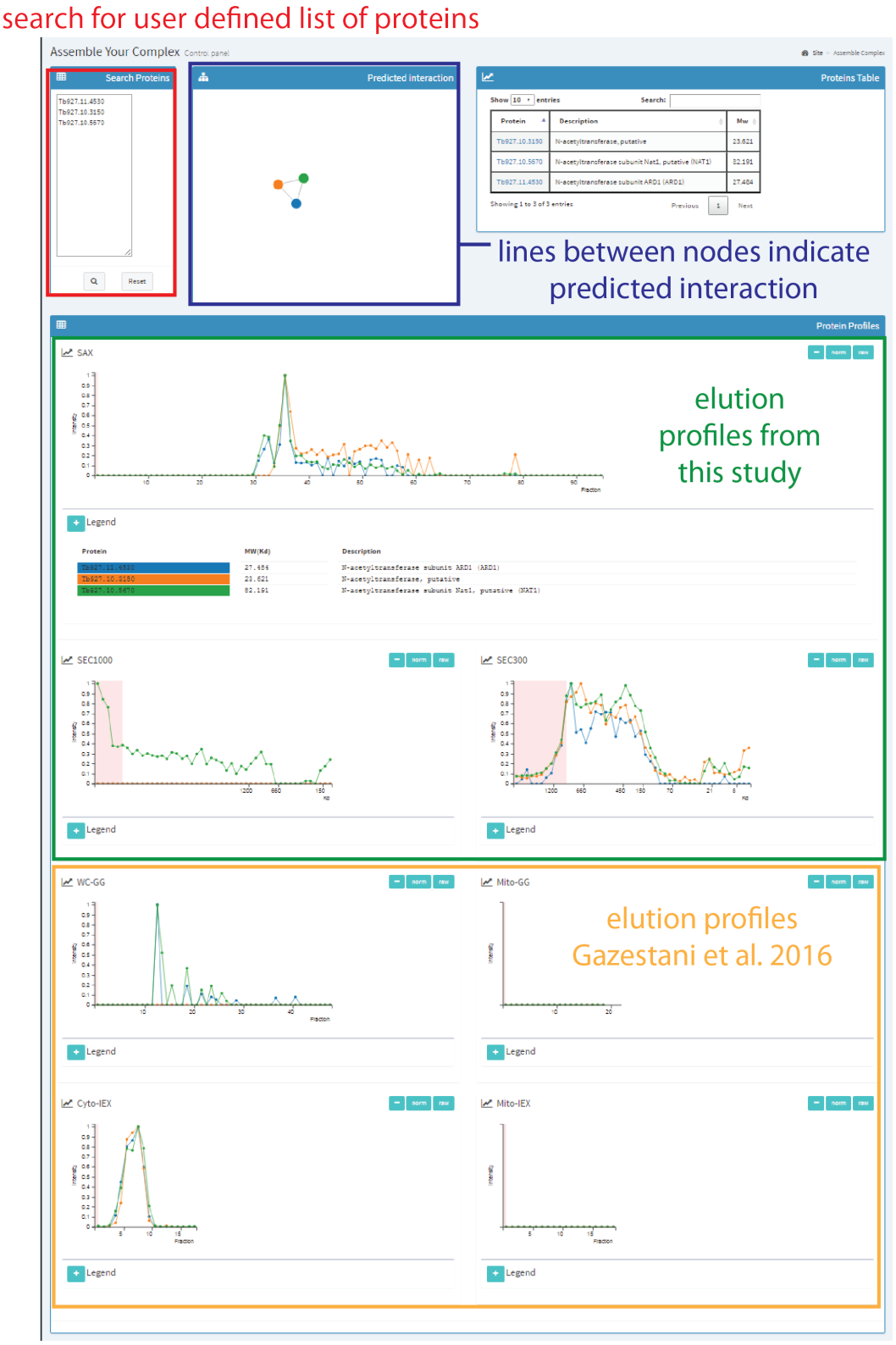


Figure 3.28: Data visualisation tools - Profile Explorer. As described in Section 3.4.7. allows users to search all data produced to compare elution profiles for any protein detected to assess their potential interaction.

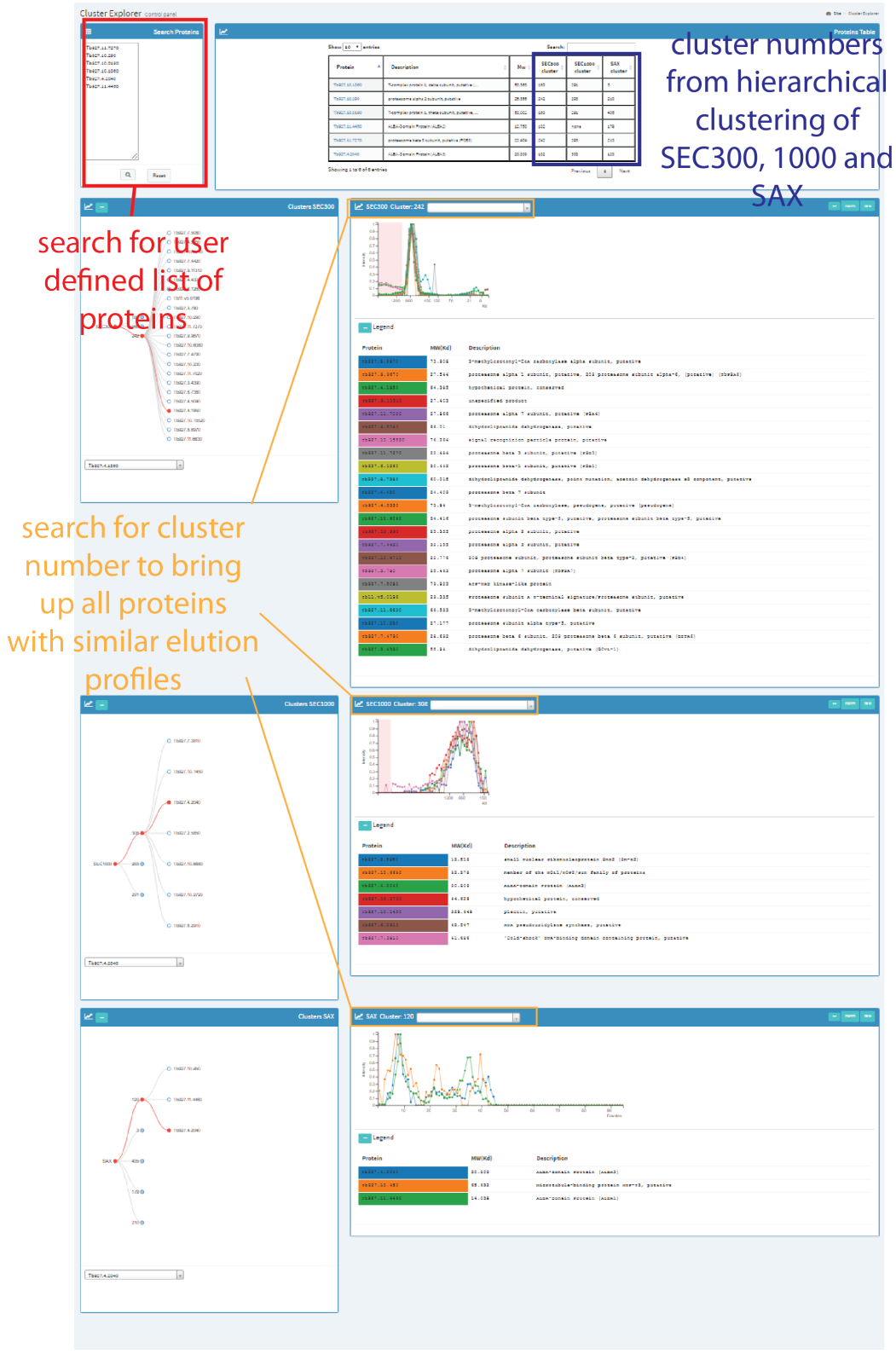


Figure 3.29: Data visualisation tools - Cluster Explorer. As described in Section 3.4.7, allows browsing for any protein detected in SEC300, SEC100 or SAX experiments and identify proteins with similar elution patterns, as identified by hierarchical clustering (Figure 3.14).

3.4.8 Initial development of cross-linking methodologies

Initial studies optimising *in vivo* crosslinking with formaldehyde and dithiobis(succinimidyl propionate) (DSP) indicate that different cross-linkers have differing effects in crosslinking distinct protein complexes. Even at the lowest concentration of DSP tested (0.125 mM), a lot of high-molecular weight protein at the top of the gel is observed, potentially indicating over-crosslinking. Formaldehyde shows a similar effect above concentrations of 2%, under the conditions utilised.

Western blotting for enolase shows that DSP stabilises a putative dimeric form of the protein up to concentrations of 1 mM, followed by an increase in high-molecular weight material. In contrast, under all formaldehyde concentrations utilised, the monomeric 47 kDa is the predominant form of enolase, with a faint dimeric band and high molecular weight forms above concentrations of 2%. In contrast, tubulin is almost undetectable when cross-linked with DSP, and a clear dimeric band is seen when utilising formaldehyde. Differing effects of cross-linking between individual proteins is likely due to the chemistries and structure of the cross-linkers. DSP is able to cross-link primary amines of different proteins across a distance of 12 Å, while the reactivity of formaldehyde is more complex. Formaldehyde can cross-link over a range of distances through the formation of polymeric para-formaldehyde, and reacts with a range of protein moieties. It is therefore likely that different complexes will require different cross-linkers at different concentrations to optimally stabilise physiologically relevant forms of the protein. Analysis of the 215 nm chromatograms from SEC of formaldehyde cross-linked lysates shows that high molecular weight proteins are stabilised at concentrations of formaldehyde above 3%.

3.4.9 Summary

To summarise, a major benefit of utilising a PCP-MS approach for proteomic analysis is that we produce information not just on protein abundance, but on protein interaction too, in a high-throughput and unbiased manner. This adds an extra dimension to the data we can extract on individual protein behaviour from a single proteomic experiment. It will be exciting to apply this approach to compare different biological samples, such as the comparison of different life-cycle stages or the effect of different drug-treatments. With the increasing ability for automation in sample preparation for chromatography (as demonstrated for SAX methodology reported here), and mass spectrometry, future studies should allow the quick production of more data looking at such effects on core protein complexes.

3.5 Distribution of work

The experimental design for the work in the chapter was devised with discussion and the advice of Dr Mark Larance. I performed all of the wet lab experiments, from cell culture to mass spectrometry data processing. The development of the machine learning pipeline for the prediction of protein complexes, all data analysis and development of data visualisation tools in this chapter was performed by Dr Michele Tinti, in discussion with myself and Professor Michael Ferguson.

Chapter 4: Proteomic analysis of the cell-division-cycle of *Trypanosoma brucei*

4.1 Introduction

In this chapter I will describe methods for producing populations of synchronous G1-phase procyclic form (PCF) *T. brucei* at a scale amenable for proteomic analysis, without the use of chemical agents to synchronise the cells, and the optimisation of mass spectrometry for protein quantitation employing 10-plex Tandem Mass Tag (TMT) technology. Our aim was to use these methods to characterise the cell-cycle regulated proteome of *T. brucei*, in a drug-free system, to produce data that might provide functional evidence for many uncharacterised proteins.

Previous transcriptomic analysis of the cell-cycle in *T. brucei* uncovered novel biological components of cell-division, unique to trypanosomatids, hence identifying highly attractive drug targets (Akiyoshi and Gull, 2014; Archer et al., 2011). The proteomic analysis carried out in this chapter should complement these transcriptomic data and further contribute to our understanding of cell-cycle control in trypanosomes. I have quantified the relative abundance of 5,034 proteins in PCF *T. brucei* across nine time-points of cell division, in all three biological replicates. From the results, I identified known cell-cycle regulators and describe cell-cycle regulated patterns of expression for 218 ‘hypothetical’ proteins, 57 of which are thought to be essential for parasite survival in culture, and may be interesting future candidates as drug targets.

4.2 Aims and hypothesis

Our hypotheses were that:

- We could identify key proteins involved in the regulation of cell division of *T. brucei* through the identification of proteins with high fold-changes across the cell-cycle.
- The identification of cell-cycle regulated patterns of protein expression could help to elucidate the biological function of 'hypothetical' proteins.
- We might identify novel drug targets through the identification of cell-cycle regulated 'hypothetical' proteins that have been classified as essential for growth in culture from RNA interference target sequencing (RITseq) studies.

Our aims were to:

- Adapt published methods for elutriation of PCF trypanosomes to a scale appropriate for proteomic analysis.
- Optimise methods for isobaric mass tag based proteomic quantitation, to carry out quantitative analysis across nine time-points of the *T. brucei* cell-cycle.
- Identify novel and essential proteins involved in the cell-cycle of trypanosomatids by cross-comparing our data to RITseq studies.

4.3 Results

4.3.1 Counterflow centrifugal elutriation

'Direct' counterflow centrifugal elutriation (Figure 4.1) was used to attempt to enrich for cells in G1, S, or G2&M phases of the cell-cycle. Fractions were collected by gradually increasing flow rates, and were analysed by flow cytometry to determine the cell-cycle distribution of collected populations (Figure 4.2). The maximum enrichment in any fraction collected for G1, S or G2&M-phase cells was 93%, 34% and 52% respectively (Table 4.1). Since the enrichment of S and G2&M-phase cells were relatively low, I chose instead to start from a G1-phase enriched (synchronised) population of cells and follow these in time to obtain S and G2&M-phase cells. Two methods were compared to produce a synchronised cell population. Single-cut elutriation splits an asynchronous culture into 'large' and 'small' cells. The small cells, enriched in G1-phase, are taken as the synchronous population (Figure 4.3). Double-cut elutriation (Archer et al., 2011) involves taking the large cell population, and culturing for 1-2 h before a second round of elutriation, where small, newly divided cells, are taken as the synchronous population (Figure 4.4). In both cases, aliquots were taken over an 11 h time-course for flow cytometry analysis. The maximum enrichment for G1, S and G2&M-phase cells was 88%, 53% and 61% using the single-cut method and 83%, 63% and 68% using the double-cut method (Table 4.1). Since the G1 cell enrichment was similar and since, single-cut elutriation returns 20% of the original culture while double-cut elutriation reduces this to about 5%, single-cut enrichment was utilised for all further studies.

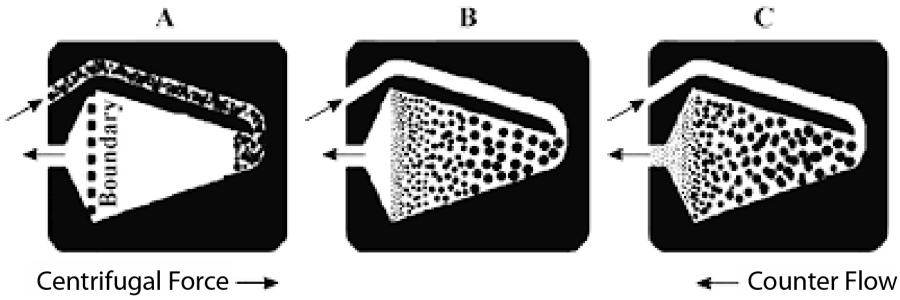


Figure 4.1: General principle of counterflow centrifugal elutriation. (A) An asynchronous culture of cells is loaded into an elutriation chamber. (B) At a constant centrifugal force and counter-flow of buffer a size gradient of cells forms, with small cells at the front of the chamber and large cells at the back of the chamber. (C) By gradually increasing the counter-flow rate of buffer smaller cells elute first followed by larger cells. Figure reproduced from *(Banfalvi, 2008)*.

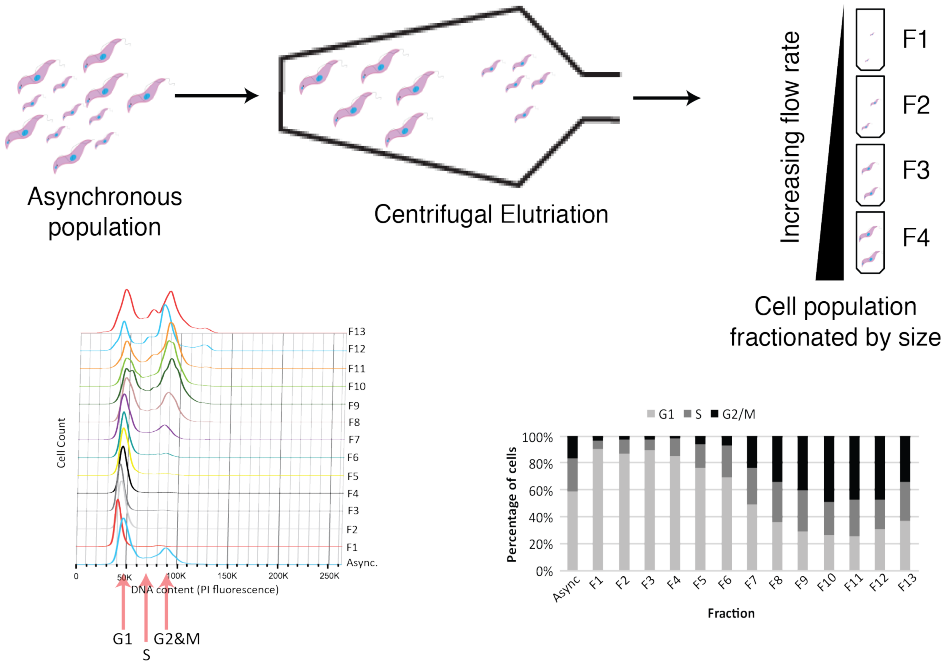


Figure 4.2: Direct counterflow centrifugal elutriation of trypanosomes. Fractions of PCF trypanosomes (small to large) were collected by gradually increasing the flow rate. Thirteen fractions were collected, and flow cytometry profiles of DNA content (propidium iodide fluorescence) were used to analyse the proportions of G1, S and G2&M-phase cells (bottom left panel). The distributions of cells in G1, S and G2&M-phase, in an asynchronous population of cells (Async) and the collected fractions are shown in the bottom right panel.

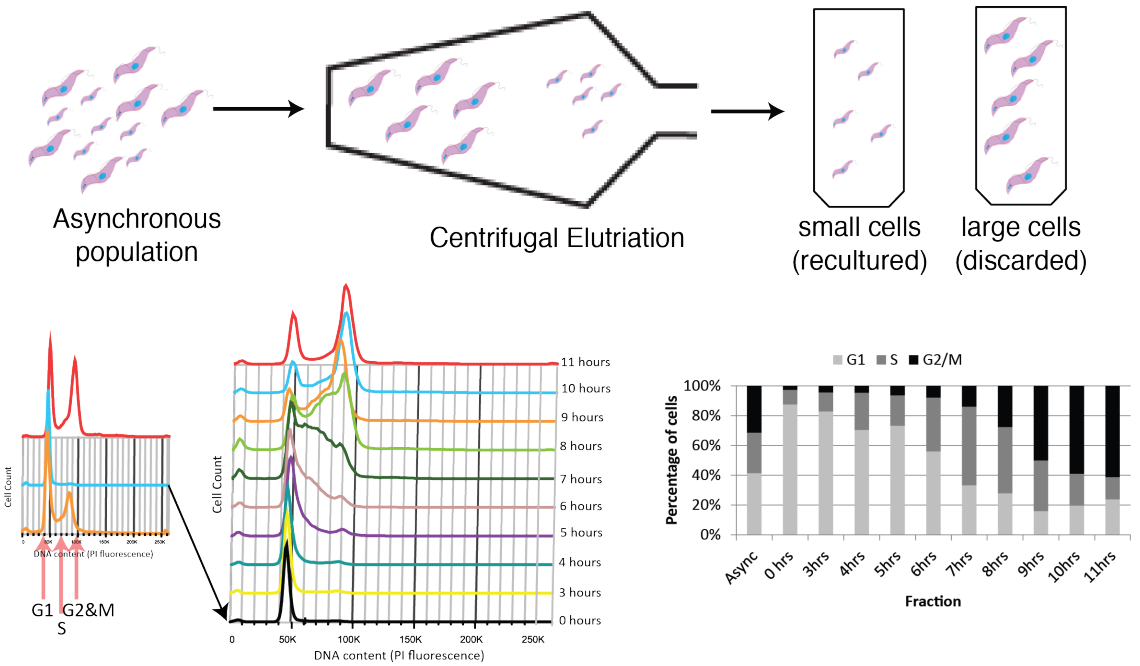


Figure 4.3: Single-cut elutriation method. Very small (G1-phase) cells were enriched from asynchronous cultures by elutriation (top panel) and these were used to seed synchronised cultures that were sampled at the times indicated for flow cytometry analysis (lower panels). Maximum enrichment for G1, S and G2&M-phase cells were obtained at t = 0, 7 and 11 h, respectively.

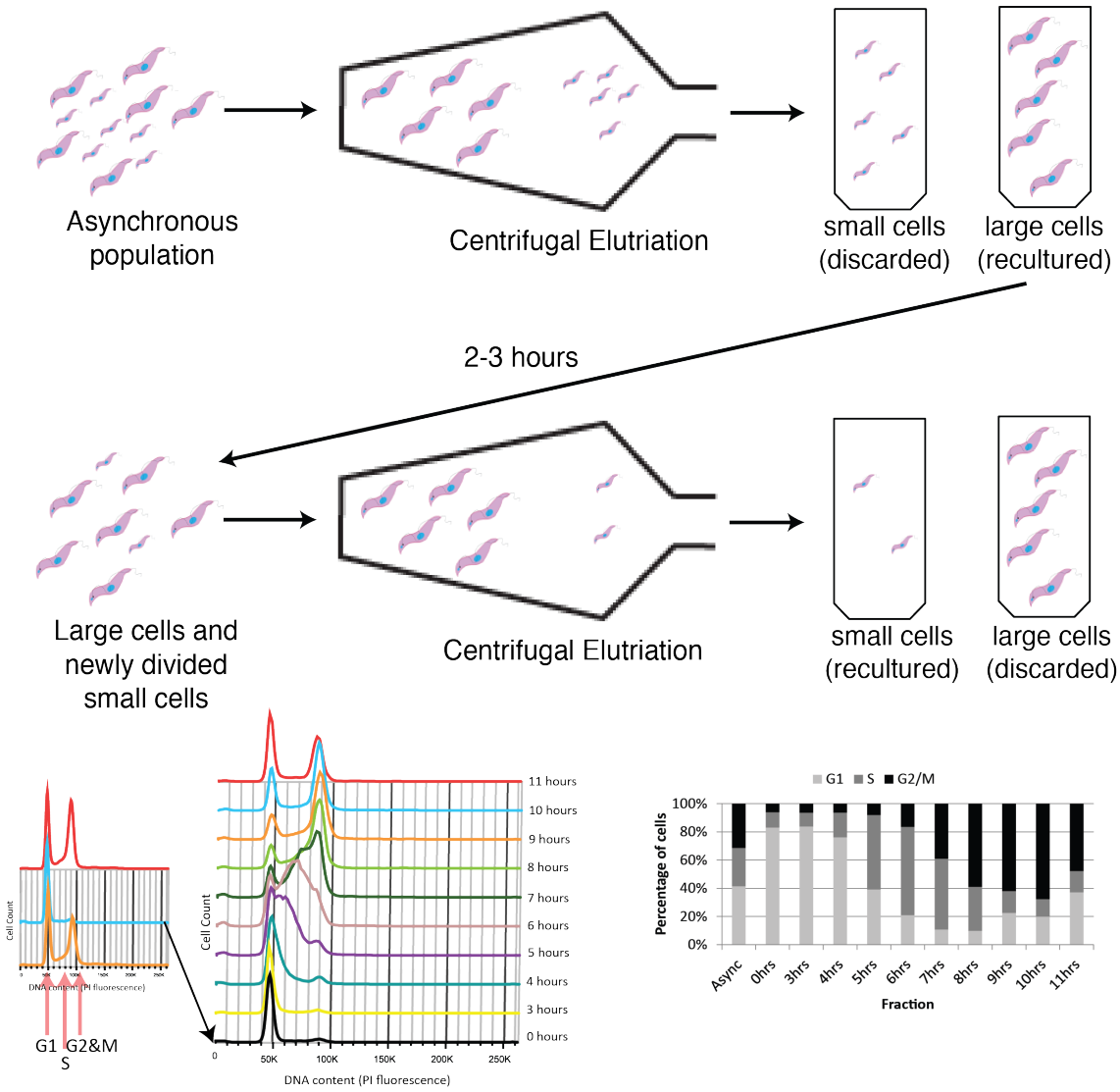


Figure 4.4: Double-cut elutriation method. Very small (G1-phase) cells were depleted from asynchronous cultures by elutriation (top panel) then cultured for 2-3 h and subjected to a second round of elutriation (middle panel) to obtain only newly divided small (G1-phase) cells to seed synchronized cultures. Maximum enrichment of G1, S and G2&M-phase cells were obtained at t = 0, 6 and 10 h respectively.

Cell-cycle phase	Direct	Single-cut	Double-cut
G1	93%	88%	83%
S	34%	53%	63%
G2&M	52%	61%	68%

Table 4.1: Maximum enrichments achieved for each cell-cycle population using either direct, single-cut or double-cut elutriation, as measured by flow cytometry.

4.3.2 Experimental design and chromatography of TMT labeled peptides

Three biological replicates of single-cut elutriation were performed, collecting nine time-points across 11 h in each case. TMT 10-plex technology was used to measure the relative abundance of proteins across the time-course (Figure 4.5). The chemical labelling of peptides with TMT increases the hydrophobicity of the peptides in question. I therefore adapted and optimised the gradients between aqueous and organic buffer across the off-line reverse-phase HPLC chromatographic run to produce an even elution of TMT labeled peptides (Figure 4.6). A shallow gradient of 35-60% acetonitrile containing buffer, followed by a sharp increase to 100% produced the most even elution pattern, based on the 215 nm absorbance profile.

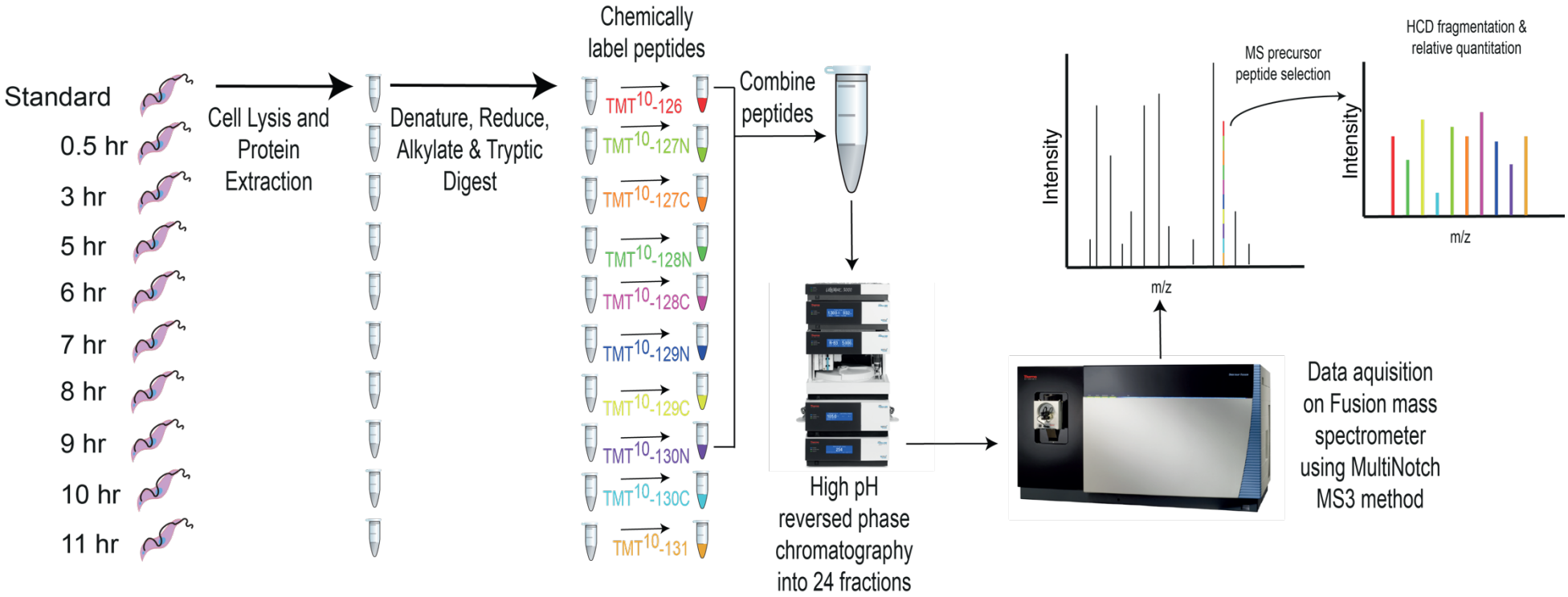


Figure 4.5: Workflow for Tandem Mass Tag (TMT) based quantitation of the cell-cycle regulated *T. brucei* proteome. Nine time-points were collected and prepared as indicated following the seeding of a synchronised culture with G1-phase enriched parasites from single-cut elutriation. The peptides from all time-points were mixed together after TMT labelling, fractionated into twenty-four fractions using high-pH reverse phase chromatography and analysed by LC-MS. A Fusion mass spectrometer using the MultiNotch SPS MS3 method *(McAlister et al., 2014)* was utilised for quantitation.

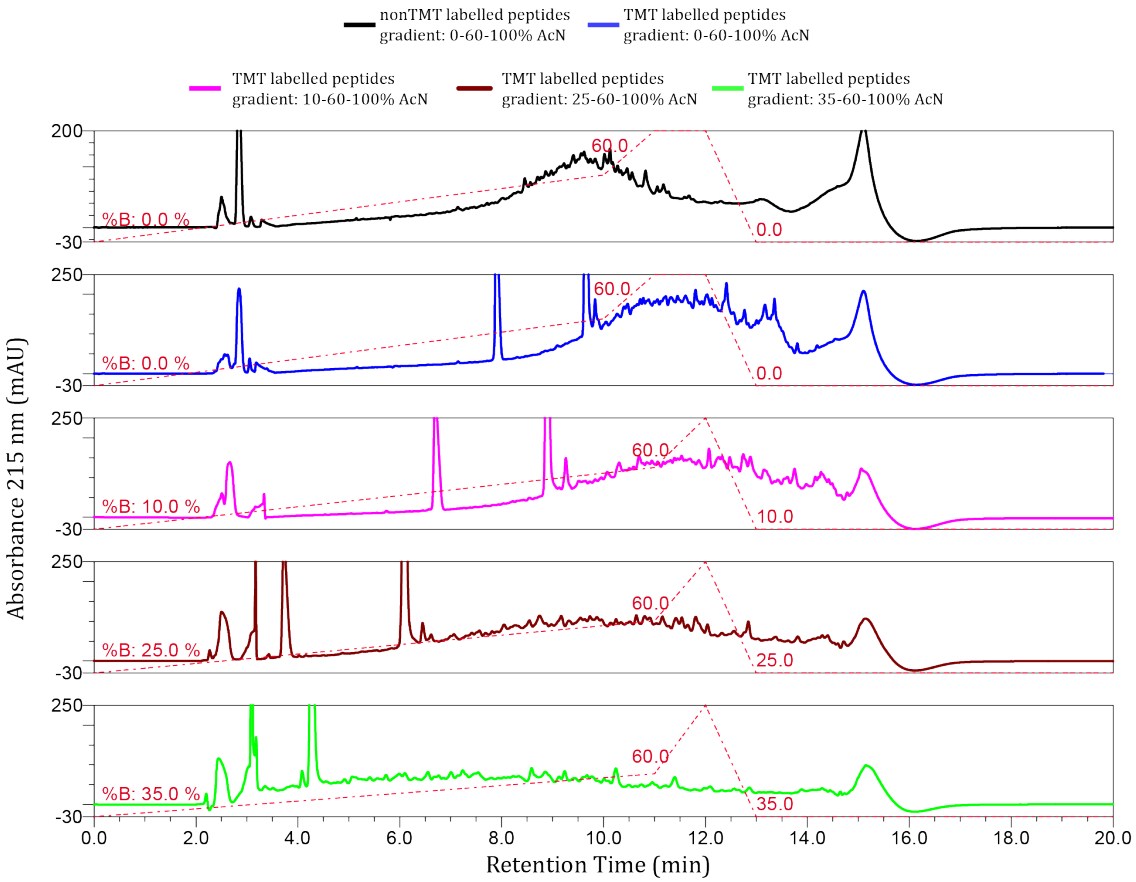


Figure 4.6: Optimisation of high-pH reverse phase fractionation of TMT labelled peptides. The 215 nm chromatogram of unlabeled and TMT-labelled peptides run on the same gradient are shown in black and blue, respectively. The chromatograms of the TMT labelled peptides with varying starting concentrations of acetonitrile are shown in pink, brown and green.

4.3.3 Optimising TMT quantitation

To develop a system to be able to analyse the precision and accuracy of TMT quantitation, and to determine the associated effects of ratio compression, I took advantage of a previously published experimental design (McAlister et al., 2014). Human peptides were labelled with six-plex TMT reagents that generate reporter ions at *m/z* 126, 127, 128, 129, 130 and 131, and mixed these in a ratio of 50:20:5:5:20:50 µg, respectively. Trypanosome peptides were labelled with TMT reagents 126, 127 and 128 and mixed at a ratio of 50:50:50 µg, and added to the human peptide mix. With this design, the quantification of human peptide ratios using the 129, 130 and 131 TMT

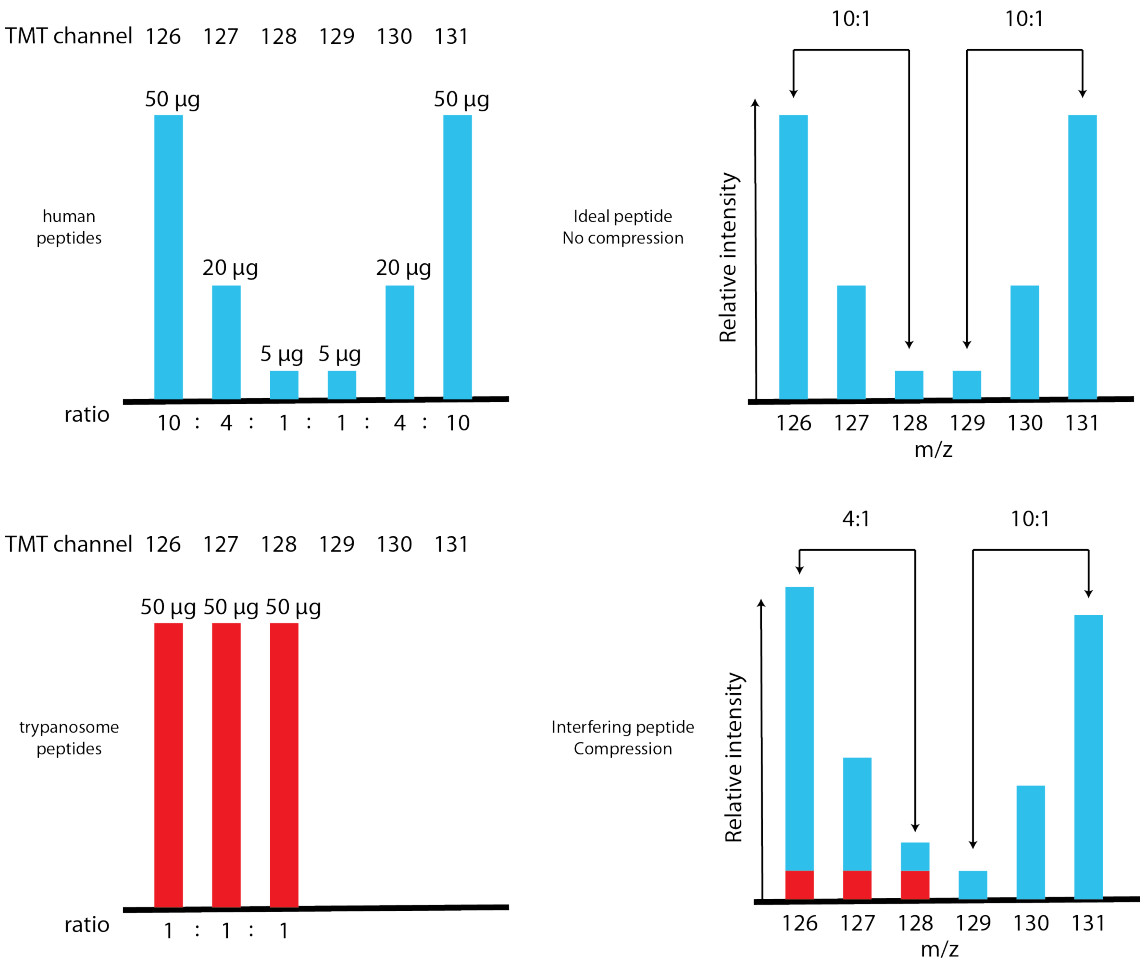


Figure 4.7: Experimental design for assessing TMT quantitation accuracy and precision. Human peptides and trypanosome peptides were labelled with the TMT reporters and mixed in the proportions indicated on the left. In the ideal situation, with no co-isolation and interference, the ratios between 126, 127 and 128 *m/z* should equal the 129, 130 and 131 *m/z* ratios (top right). However, if trypanosome peptides are co-isolated they will compress the ratios between 126, 127 and 128 *m/z*, but have no effect on the 129, 130 and 131 *m/z* ratios (bottom right). Figure adapted from *(Ting et al., 2011)*.

reporter ions will be uncompressed, whereas the 126, 127 and 128 TMT reporter ions will demonstrate effects of interference if trypanosome peptides are co-isolated with human peptides, leading to compression of the ratios measured (Figure 4.7). For the two sets of compressed and uncompressed reporter ions there are three expected ratio measurements: 126/128 or 131/129 as a ratio of 10-fold; 127/128 and 130/129 as a ratio of 4-fold; and 126/127 or 131/130 as ratio of 2.5-fold. This set-up was utilised to test

the effects on compression of a variety of parameters on QExactive+ and Fusion mass spectrometers.

Decreasing the isolation width used to select target precursor ions for fragmentation on a QExactive+ instrument from 4.0 to 1.5 and 0.4 *m/z* resulted in more accurate quantification of peptide spectral matches (psms) from compressed reporter channels, with an expected 10:1 ratio quantified as 2.8±0.9, 3.4±1.1 and 4.1±1.4, respectively. This equates to a 3.5-fold, 2.9-fold and a 2.4-fold compression, respectively. An increased number of psms is also detected, from 2,754 at 4.0 *m/z*, to 3,327 at 1.5 *m/z* and 3,422 at 0.4 *m/z* (Figure 4.8). It was not possible to detect a measurable difference between the compressed 4-fold and 2.5-fold ratios at any of the isolation widths tested. All uncompressed ratios were calculated within one unit of their expected values. An expanded set of isolation widths (0.4, 0.7, 1.0, 1.5, 2.0, 3.0 and 4.0 *m/z*) tested on a QExactive+ instrument is displayed in Supplementary Figure 1 in the Appendix.

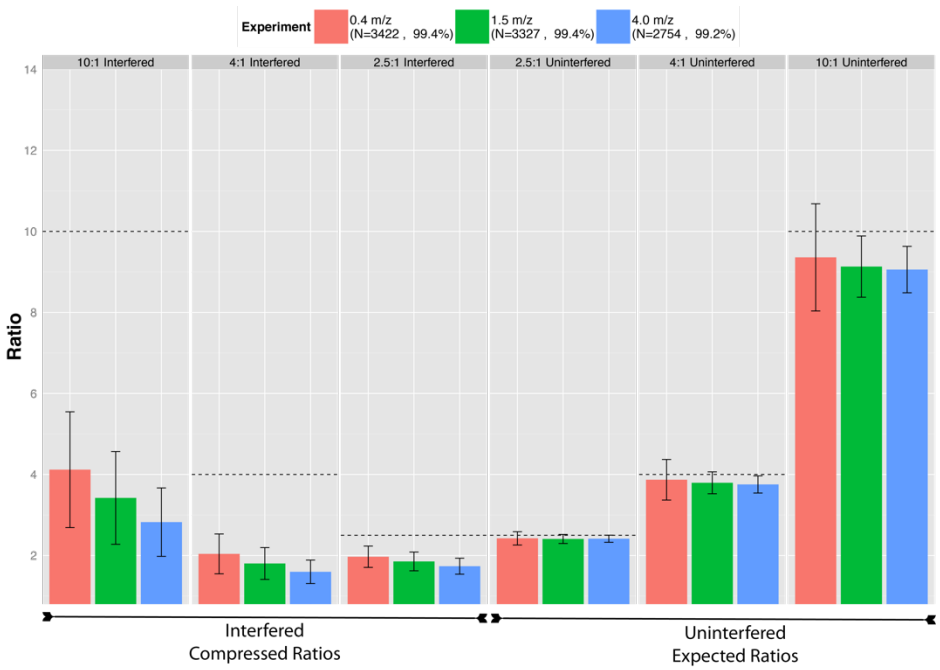


Figure 4.8: Effect of isolation width on TMT quantitation on a QExactive+. Dashed lines indicate the expected ratio in each facet. The key shows the isolation width used in each method together with N, the number of peptide spectral matches (psms) quantified in each method, and the percentage of psms identified that are modified with TMT, identified in all six TMT reporter channels and pass the intensity threshold of 100,000.

Data obtained on a Fusion mass spectrometer indicated high levels of inaccuracy in quantitation of ratios at lower total TMT reporter ion intensities (Figure 4.9). Many psms quantified with an expected 10-fold ratio, were detected as >20-fold ratios. To exclude these imprecise ratios an intensity threshold of 100,000 units was implemented for all data acquired in this set of samples. Following filtering of the psms by intensity, most methods on the Fusion produced similar distributions of compressed ratios, ranging from 7.5±2.2 to 6.8±2.1 for an expected ratio of 10:1, equating to 1.3 and 1.5-fold compression.

Using an initial default method provided from the mass spectrometer manufacturer ('Thermo') I tested the effects of: increasing the maximum injection time (maxIT) for MS3 ion accumulation from 105 ms to 300 ms ('300 ms scan'); modifying the isolation width for target ion selection from 0.8 to 1.6 *m/z* ('1.6 *m/z*'); and increased

the amount of peptide loaded from 1 to 2 µg ('2 µg') per run (Figure 4.10). To demonstrate the sequential effects of modifying the default 'Thermo' method in the parameters described above, a sub-set of experiments carried out are shown, highlighting the key effect from modifying each parameter. More extensive results are described in the Appendix and Supplementary Figures 2-5, showing a wider range of settings tested for isolation width and peptide load.

Only 30.7% of psms identified using the original default method passed the intensity threshold of 100,000. Increasing the maxIT for the MS3 scan to 300 ms led to a decrease in the number of identified psms, but an increase in identified and *quantified* psms, as 61.8% now passed the intensity filter. Further sequential modification of the 300 ms method by increasing the isolation window of the MS1 to MS2 scan from 0.8 to 1.6 *m/z* also increases the number of psms passing the intensity filter to >80%, as did increasing peptide load to 2 µg (Figure 4.11). I therefore utilised a 300 ms maxIT and 1.6 *m/z* mass spectrometry method with ~2 µg peptide load for analysis of the *T. brucei* cell-cycle regulated proteome.

In a second set of experiments with a mixture of mouse and trypanosome peptides a number of other parameters on quantitation were tested, including the number of SPS notches used to select fragment ions for quantitation; a titration of the Automatic Gain Control target; a titration of maxIT; and tested software allowing for separate isolation widths for MS2 and MS3 scans. The results of these variations are described in the Appendix and Supplementary Figures 6-11.

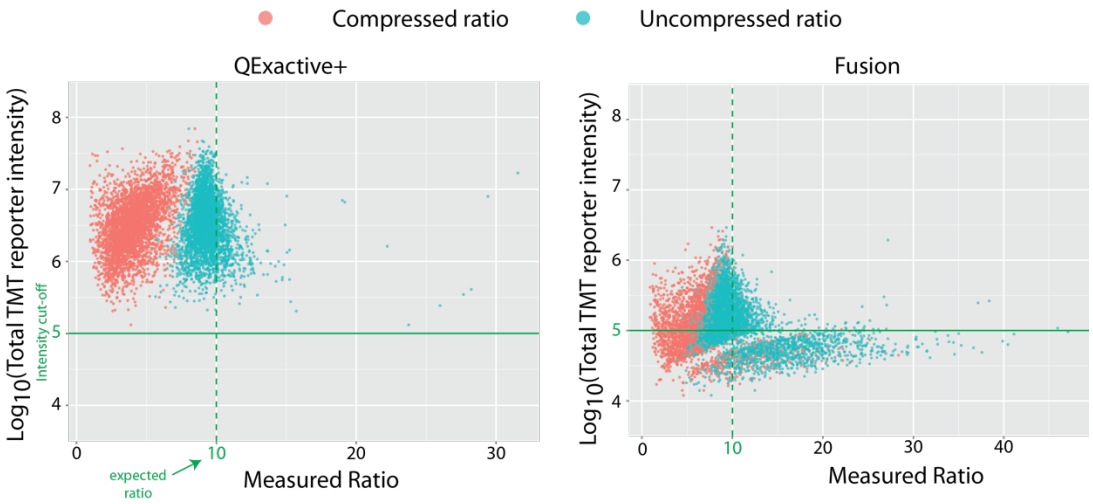


Figure 4.9: The use of an intensity filter to exclude inaccurate ratios. Scatter plots of the total TMT reporter intensity for a psm and the calculated compressed ratio (126/128 *m/z*) and uncompressed ratio (131:129 *m/z)* on a QExactive+ (left) and Fusion (right). Dashed green line shows the expected 10:1 ratio and the full green line shows the intensity filter of 100,000 implemented to remove inaccurate ratios calculated at low reporter intensities.

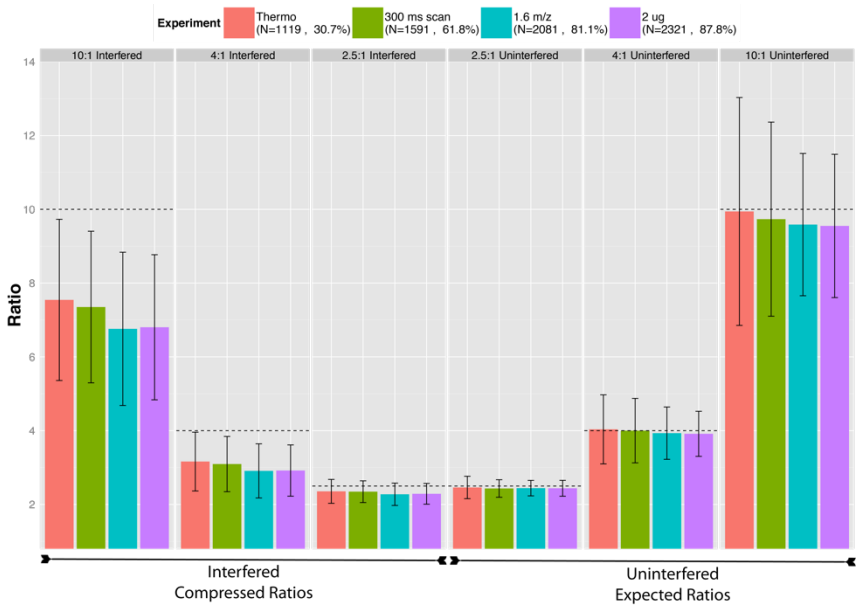


Figure 4.10: TMT quantitation on a Fusion mass spectrometer using different methods. The key shows method used together with N, the number of peptide spectral matches (psms) quantified in each method and the percentage of psms identified that are modified with TMT, identified in all six TMT reporter channels and pass the intensity threshold of 100,000.

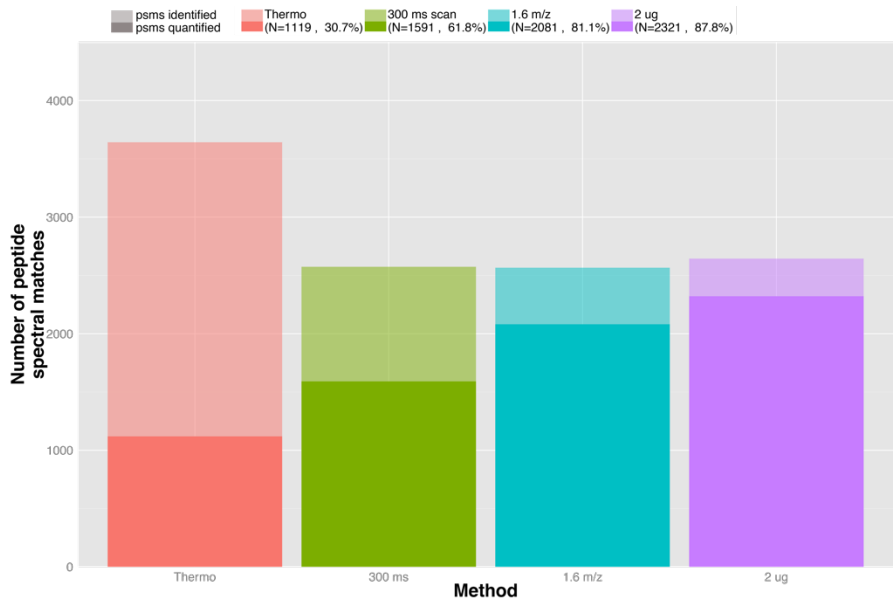


Figure 4.11: Number of psms identified and quantified across different Fusion methods. Transparent bar shows the number of psms identified in each method, and filled bar shows the number of identified psms that are TMT modified, quantified in all six TMT channels and pass the intensity filter of 100,000. The key shows the number of peptides quantified (N) and the percentage of psms identified that are quantified.

4.3.3 Cell-cycle regulated proteome

Following digestion of proteins to peptides, TMT labelling and mixing of time-points, each of the three biological replicates were separated by high-pH reverse phase fractionation into twenty-four fractions. Each fraction was run on a Fusion mass spectrometer using a 300 ms maxIT and 1.6 *m/z* isolation width in triplicate. A total of 45,777 peptide sequences were identified corresponding to 6,356 protein groups, with 5,034 detected and quantified across all nine time-points in all three biological replicates with ≥1 unique peptide. Protein levels were normalised by setting the maximum reporter intensity for a particular protein to 1.

To determine if there is more variation across a time-course than between biological replicates an Analysis of Variance (ANOVA) was performed on protein groups detected across all three biological replicates. Of the 5,034 protein groups quantified, 2,215 had p-values <0.05, indicating a statistically significant difference in

their measurement across the time-course, compared to within time-points across biological replicates; 166 of these proteins have a maximum fold-change >1.5, and 396 are >1.3 fold (Figure 4.12).

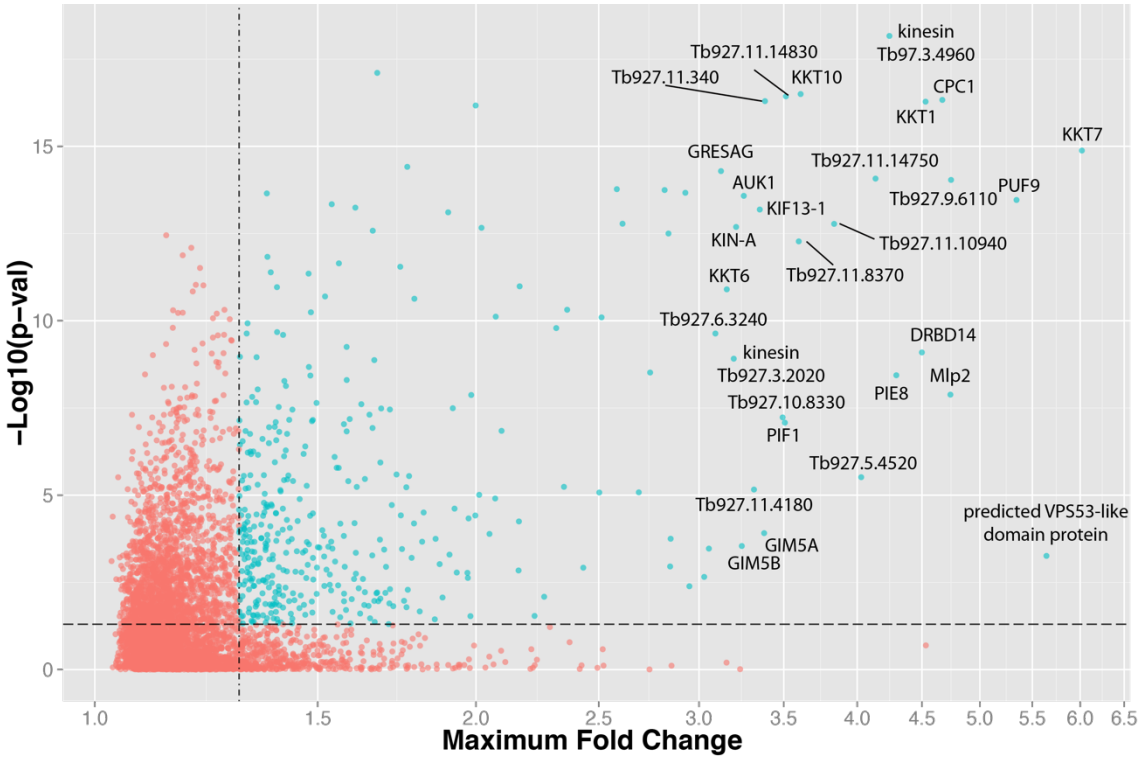


Figure 4.12: Scatter plot of maximum fold-change across the cell-cycle and ANOVA p-value. Dashed lines show a p-value of 0.05 and a maximum fold change of 1.3. Of 5,034 proteins detected, 396 (blue dots) can be classified as cell-cycle regulated with a fold change ≥1.3 and p-value ≤0.05.

The 2,215 protein groups were hierarchically clustered into ten groups classifying distinct patterns of cell-cycle regulation (Figure 4.13). Clusters were named based on the temporal expression patterns of proteins and cross-referencing to the flow cytometry profiles of collected samples (Figure 4.3). 1,787 proteins were classified as "Unchanging", 210 as "G2&M", 110 as "G1/early S", 52 as "early G1", 20 as "high G2&M", 16 as "high S", 13 as "high early G1", 2 as "high G1", 4 as "high early S" and 1 as "early G1/late G2&M", with "high" indicating clusters of proteins with larger fold-changes (Supplementary Table 4). Selected gene ontology (GO) terms enriched

within each cluster are displayed in the appropriate panels in Figure 4.13.

To display the data, I have produced radial visualisation plots where time-points are hours on the clock-face and proteins are plotted towards the time-point they are most abundantly detected in. A number of known regulators of the cell-cycle in *T. brucei* are upregulated at time-points that correlate well with their biological function (Figure 4.14). Thirty-eight proteins annotated with GO terms associated with the cell-cycle are detected; 24 in the "Unchanging" cluster and 14 in the remaining 9 clusters with patterns of cell-cycle regulation (Figure 4.15). By cross-comparison to RNA interfering target sequencing (RITseq) datasets it was determined that 287 of the 428 proteins in cell-cycle regulated clusters are essential for growth in one or more lifecycle stage of *T. brucei* in culture (Alsford et al., 2011). Of these, 57 are annotated as 'hypothetical' proteins (Figure 4.16). These data are also available to browse on an interactive html file (https://uod.box.com/v/cellcycle).

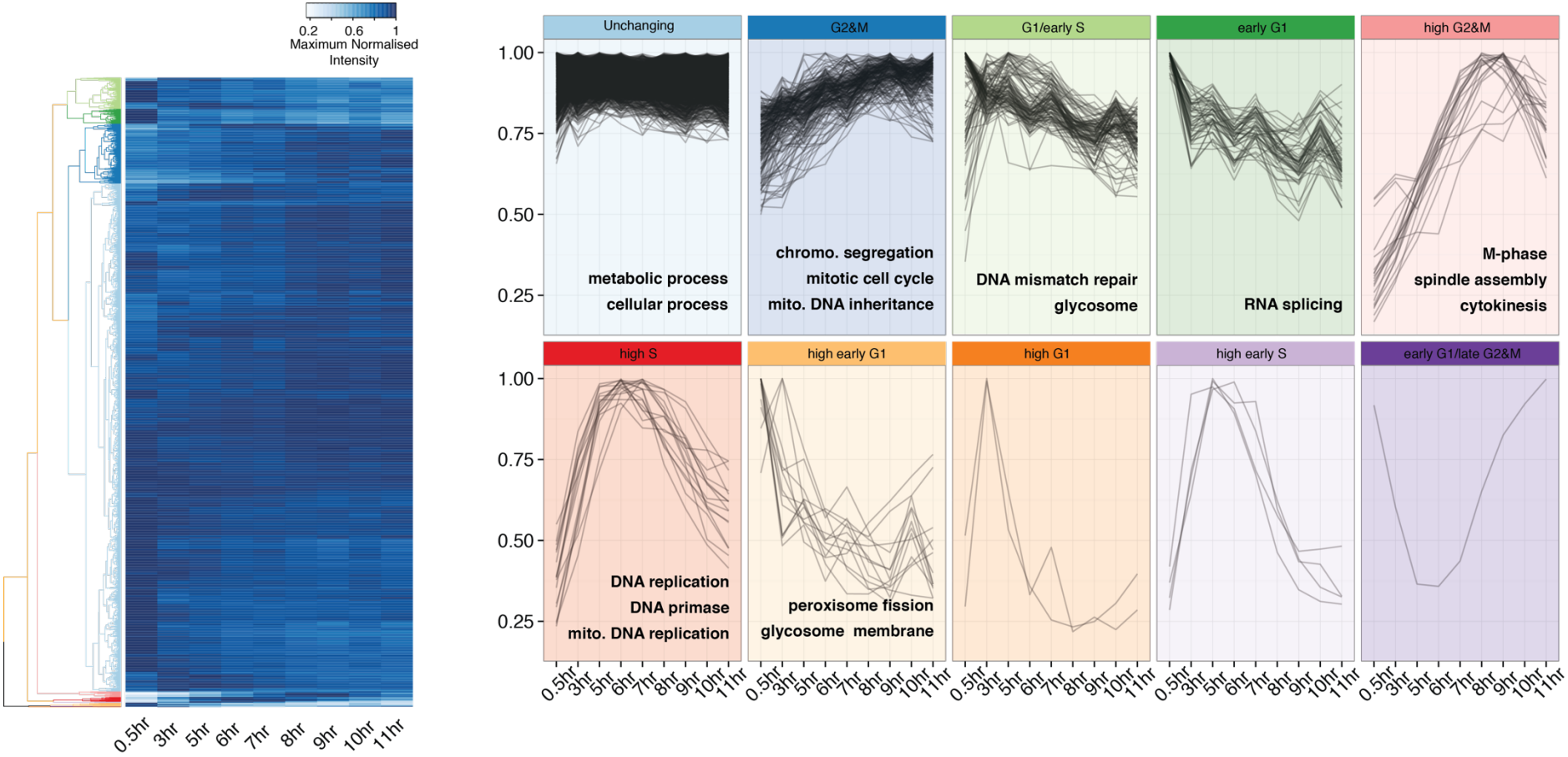


Figure 4.13: Hierarchical clustering. 2,215 proteins with a p-value ≤0.05 were hierarchically clustered into ten clusters. Heat-map of clustered normalised elution profiles on left panel. Right panel shows line-plots of proteins within each cluster. GO terms enriched in each cluster are annotated in line-plots. The colours of dendrogram branches match colours of the respective line-plots.

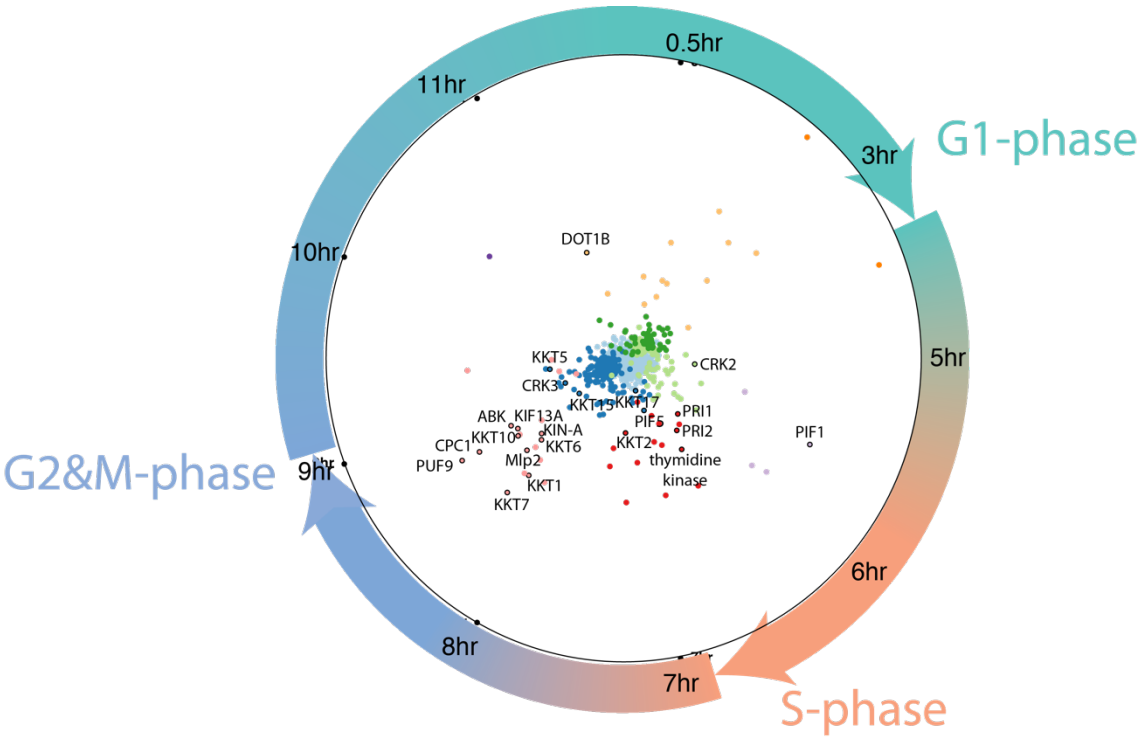


Figure 4.14: Radial visualisation plot of clusters and known cell-cycle regulated proteins. Each time-point is depicted as an hour on a clock-face. Coloured arrows are representative of the cell-cycle phase at each time-point. Proteins are pulled towards the time-point they are most abundantly expressed in. Dot colour refers to the cluster the protein is detected in (Figure 4.13). Proteins known to be involved in cell-cycle regulation are annotated.

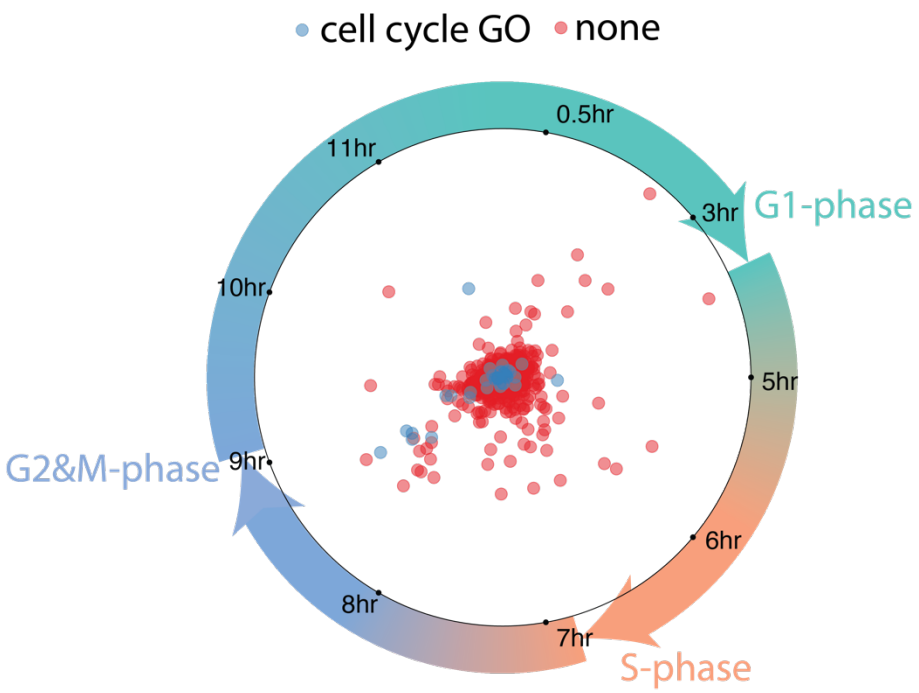


Figure 4.15: Many proteins detected as cell-cycle regulated lack cell-cycle GO annotation. Proteins in blue are annotated with a GO term associated with the cell-cycle, red proteins have no detectable cell-cycle GO term association.

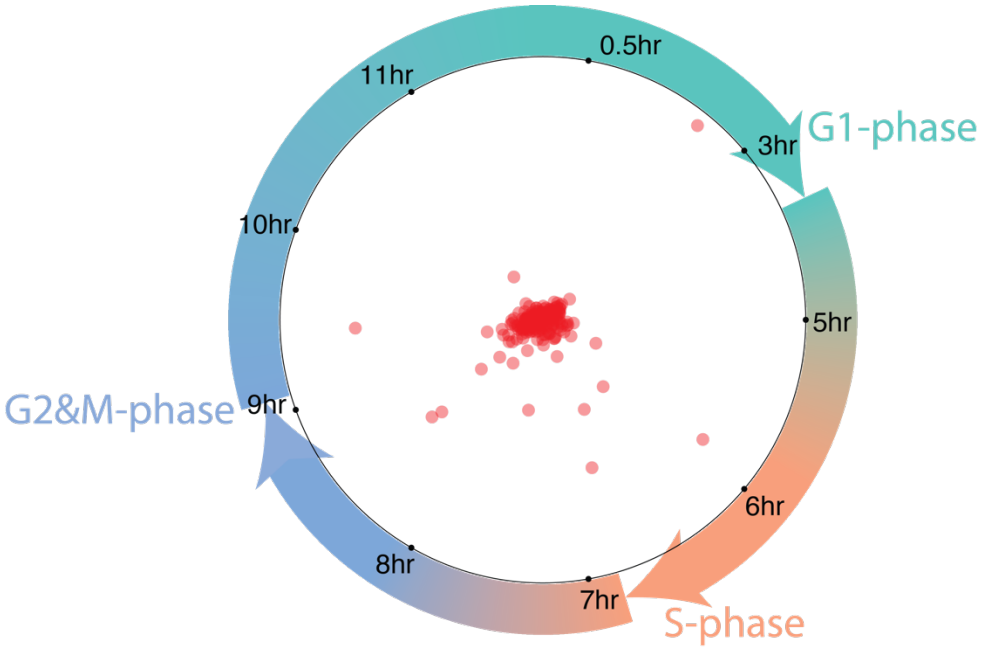


Figure 4.16: Radial visualization plot of 'hypothetical' proteins with an essential phenotype. Proteins displayed are annotated as 'hypothetical' and are essential for growth of one or more life-cycle stages in RITseq studies. Proteins at further edges of the plot are highly cell-cycle regulated, and their essentiality may be due to a cell-cycle function.

4.3.4 Comparison to transcriptomic dataset

We determined the overlap of proteins detected in our dataset to transcripts detected in a previously published transcriptome analysis of the cell-cycle of PCF trypanosomes (Figure 4.17)(Archer et al., 2011). Of the 5,034 proteins quantified in this work, 93% are detected in the transcriptomic dataset. Conversely 67% of the 6,948 transcripts identified are identified in the proteomic dataset. Proteomic and transcriptomic analyses classify 428 or 544 proteins/transcripts as regulated across the cell-cycle, respectively, a total of 887 unique transcripts/proteins. Of these, only 85 are classified as regulated in both datasets (Supplementary Table 5), while 302 (proteome, Supplementary Table 6) and 374 (transcriptome, Supplementary Table 7) are identified in both datasets, though only changing in one, and 41 (proteome, Supplementary Table 8) and 185 (transcriptome, Supplementary Table 9) are only identified in one dataset. GO enrichment analysis of each of these classes of transcripts/proteins was performed (Figure 4.18). From the 85 transcripts/proteins identified as changing in both transcriptomic and proteomic analyses, I produced a contingency table to assess how well the classification of cell-cycle phases associated with each transcript/protein matched between each dataset (Table 4.2). Early and late G1-phase classified transcripts from (Archer et al., 2011) were grouped to produce the G1-phase set of genes from transcriptomic data.

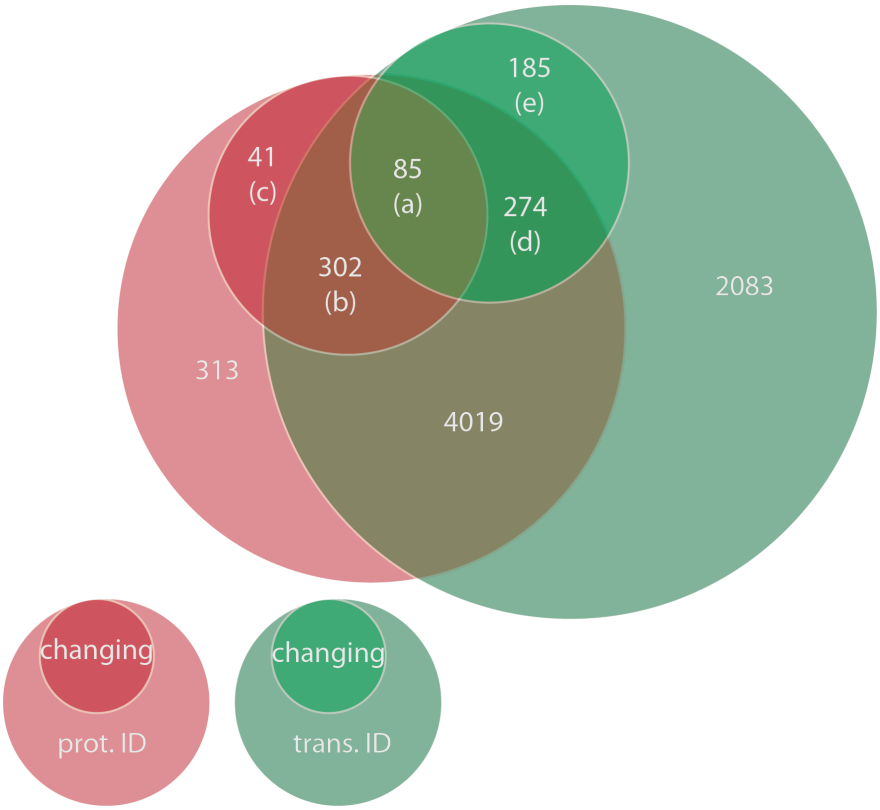


Figure 4.17: Overlap of proteomic and transcriptomic cell-cycle datasets. Large, lighter coloured circles depict transcript/protein IDs identified in either proteomic (red) or transcriptomic (green) datasets *(Archer et al., 2011)*. Small, darker coloured circles depict those transcripts/proteins classed as cell-cycle regulated. (a) Regulated transcripts/proteins in common. (b) Identified in both datasets but only regulated in proteomic dataset. (c) Changing proteins uniquely identified by proteomics. (d) Identified in both datasets but only regulated in transcriptomic dataset. (e) Changing transcripts uniquely identified by transcriptomics

	G1 - proteome	S - proteome	G2&M - proteome	total transcripts
G1 - transcriptome	29	12	11	52
S - transcriptome	3	2	18	23
G2&M - transcriptome	2	0	8	10
total proteins	34	14	37	85

Table 4.2: Contingency table of peak expression time-points for genes/proteins identified as cell-cycle regulated in both transcriptomic and proteomic datasets.

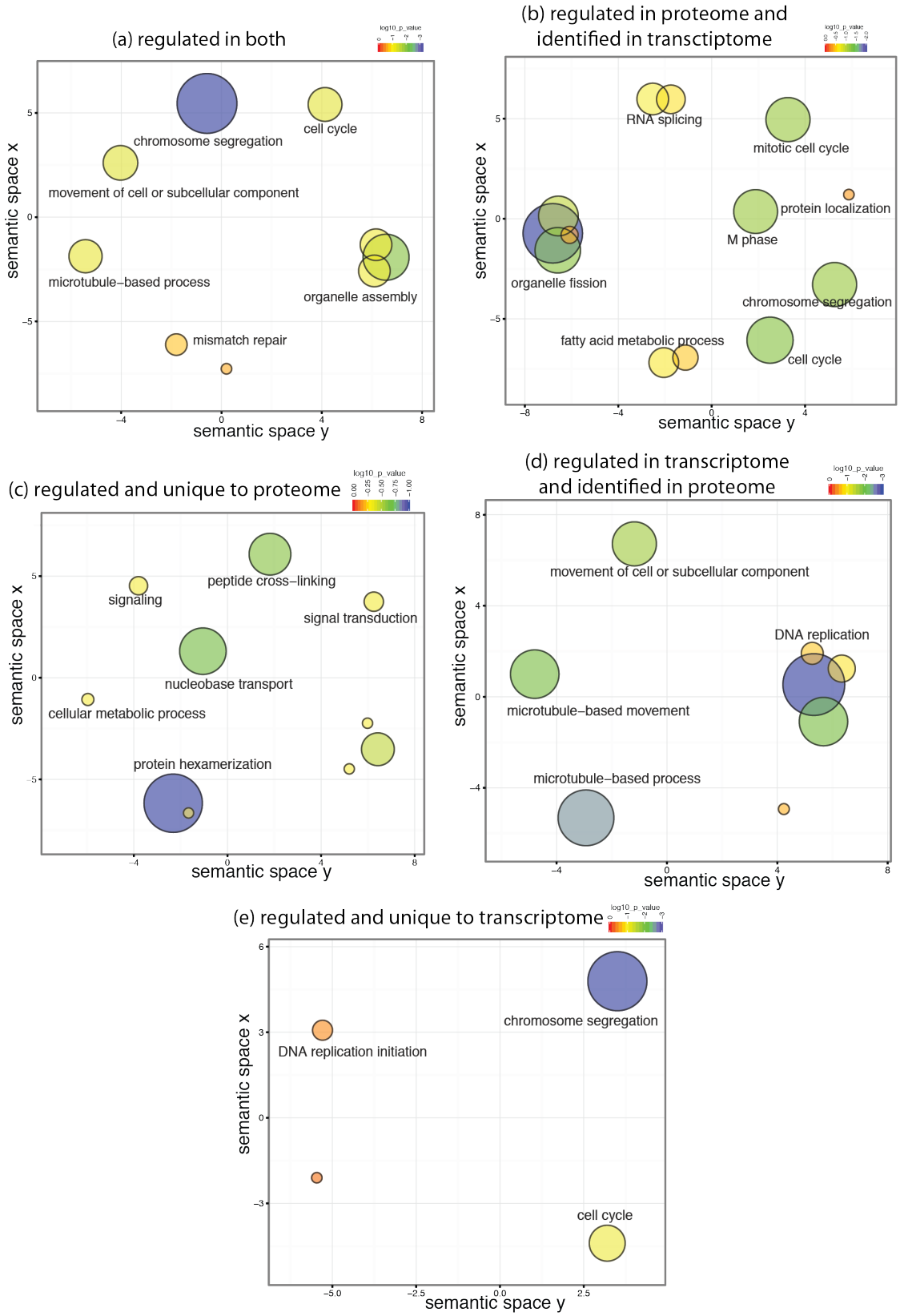


Figure 4.18: Gene ontology enrichment for the transcript/protein categories described in Figure 4.17. Dot sizes representative of the –log(p-value) associated with each GO term.

4.4 Discussion

I have optimised methods for centrifugal elutriation to produce synchronous G1-phase PCF trypanosome cell populations, without the aid of chemical agents, and methods for quantitation of TMT reporter ions on a Fusion mass spectrometer. Together, this has allowed us to quantify the relative abundance of 5,034 proteins in PCF trypanosomes across the cell-cycle, a useful resource that will contribute to our understanding of cell division in this parasite. Hierarchical clustering was used to classify patterns of temporal regulation, which allowed the identification of uncharacterised proteins upregulated in different phases of cell division, highlighting their potential functions. A number of known cell-cycle regulators in trypanosomes were detected, and a number of novel, essential, cell-cycle regulated proteins which are promising future drug targets were identified.

4.4.1 Comparison of elutriation methods

Using elutriation to directly enrich for different cell-cycle phases works well for G1-cells, (93% enrichment) however there does not appear to be enough of a size difference to fully separate S-phase and G2&M-phase cells (34 and 52% maximum enrichment, Table 4.1). Better results were obtained by seeding synchronised cultures with G1-phase cells. Of these, double-cut elutriation provides an extremely effective method to produce synchronous G1 cells as previously described (Archer et al., 2011). However, the single-cut method described here produced very similar results while providing a higher yield of G1-phase cells, which is beneficial for high proteomic coverage to capture low abundant proteins.

Single-cut elutriation compares well to other methods to produce populations enriched in different cell-cycle phases. It is possible to sort cells by flow cytometry based on DNA content either on live or fixed cells for proteomic analysis (Kabani et al.,

2010). However, to produce ~200-400 µg of protein per sample requires ~1 x 10^8 trypanosome cells, which would require very long sorting times using flow cytometry, especially for S-phase cells that constitute ~15% of asynchronous cultures. Other methods include drug treatments to synchronise cells, such as hydroxyurea treatment (Chowdhury et al., 2008; Forsythe et al., 2009), or starvation through removal of serum from culture (Gale et al., 1994). Although these are much simpler methods than elutriation, it has been shown to lead to the proteomic analysis of cell-cycle arrest rather than physiological cell-cycle regulation (Ly et al., 2015).

4.4.2 Comparison of methods for protein quantitation

To quantify cell-cycle regulated changes of the *T. brucei* proteome across nine time-points over 11 h I decided to use ten-plex TMT technology (Figure 4.5). For this study, the ability to quantify ten samples in a single mass spectrometry run dramatically reduced the amount of continuous instrument time required from ~5 months to just over two weeks in comparison to an equivalent label free experiment. This technique also eliminates the "*missing value problem*" that affects label free quantitation, as quantitative data from all nine time-points are detected in a single mass spectrometry scan as opposed to from multiple separate scan events from different samples in a label free set-up.

However, one of the major drawbacks of using isobaric tagging technology such as TMT for quantitative proteomics is the effect of interference. Interference occurs when peptides with a similar retention time and a similar *m/z* are co-isolated when selecting a target ion for quantitation. Quantitative information is therefore produced from btoh the target ion and the co-isolated contaminant ions, which distorts the quantitation of the target. As most peptides do not differ in abundance across most biological conditions, the global effect of this is to compress the measured ratios of

peptides which are changing. To lessen the effect of interference I tested the use of very narrow isolation widths on QExactive+ mass spectrometers and determined optimal parameters on a Fusion mass spectrometer, which can perform a function known as synchronous precursor selection (SPS). SPS is used to select up to ten fragment ions from the initial target ion for a third MS scan for quantitation, reducing the likelihood of co-isolating contaminant ions.

Narrowing the isolation width on a QExactive+ instrument from 4.0 to 1.5 and 0.4 *m/z* reduced the effect of ratio compression (Figure 4.8). It may be expected that narrowing isolation widths too much would lead to a reduced number of peptides being identified due to lower signal intensity, and hence poorer quality MS2 identification scans. However, an increasing number of psms is observed by decreasing the isolation width from 1.5 *m/z* (3,327) to 0.4 *m/z* (3,422). These results indicate that if samples are being quantified on a QExactive+ mass spectrometer then lower isolation widths reduce the effects of compression without a hit on numbers of psms identified. Unfortunately, even using a very narrow isolation width of 0.4 *m/z,* interference still causes dramatic effects of ratio compression, with a 2.4-fold compression for a 10:1 ratio, 2-fold for a 4:1 ratio and 1.25-fold for a 2.5:1 ratio.

The use of the SPS-MS3 method using a Fusion mass spectrometer reduces ratio compression significantly (Figure 4.10). A caveat for this method is that at low reporter ion intensities, inaccurate ratios are calculated (Figure 4.9). Implementing an intensity filter on identified psms improves the accuracy of quantification, though leads to a large drop in the number of psms utilised in quantification – 1,119 in the case of the original manufacturer method (30.7% of psms identified), 3-fold lower than detected on the QExactive+. Increasing the amount of material available for the quantitation scan improved the proportion of psms quantified. This was achieved by: increasing the

maxIT for the quantitative MS3 scan to 300 ms, allowing a longer period of time for lower abundance ions to accumulate; increasing the isolation width for target ion selection to 1.5 *m/z* from 0.8 *m/z*; and simply by injecting >2 µg of peptide per run (Figure 4.11). Together these measures boosted the number of quantified psms to 2,434, 90.6% of those identified and only 1.3 fold lower than the number detected on the QExactive+ system.

4.4.3 Identification of known cell-cycle regulated proteins

Of the 5,034 protein groups identified and quantified across the cell-cycle of PCF *T. brucei*, 2,155 have an ANOVA p-value <0.05 indicating a statistically significant change in abundance across the time-course. We classify 396 of these as being cell-cycle regulated, with a fold change >1.3. Proteins with a fold-change >3 across the cell-cycle include a mixture of hypothetical proteins of unknown function, proteins with a putative gene annotation and several proteins known to be cell-cycle regulated (Figure 4.12).

Of the proteins with prior evidence of cell-cycle regulation, increases in expression are detected at the expected time-points (Figures 4.12 and 4.14). Upregulated at the 5 h time-point, between G1 and S-phase, CRK2 is detected (a cdc2 related kinase), depletion of which results in a G1-phase block in *T. brucei* (Tu and Wang, 2004; 2005). PIF1, a DNA helicase necessary for kinetoplast DNA replication in early S-phase (Liu et al., 2009), is detected between the 5 and 6 h time-points, and thymidine kinase upregulated between 6 and 7 h, necessary for genomic DNA replication (Valente et al., 2016). Many of the proteins upregulated between 8 and 9 h are well annotated with ascribed G2&M-phase functions and include components of the chromosomal passenger complex (AUK1, CPC1 and KIN-A)(Li et al., 2008a; 2008b; Tu et al., 2006), another cdc2-related kinase (CRK3)(Tu and Wang, 2004), motor proteins involved in

spindle assembly (Mlp2 and KIF13)(Chan and Ersfeld, 2010; Holden et al., 2014; Morelle et al., 2015; Wickstead et al., 2010) and multiple kinetochore proteins (KKTs)(Akiyoshi and Gull, 2014). Finally, DOT1B is upregulated late in G2&M-phase and into G1-phase, a histone methyltransferase known to modify chromatin as cells exit mitosis, and necessary for cell division during differentiation from BSF to PCF (Dejung et al., 2016; Janzen et al., 2006).

4.4.4 Classification of temporal patterns of protein abundance

We have grouped the protein abundance profiles detected in this dataset into ten clusters with distinct patterns of regulation across the collected time-course (Figure 4.13). As expected, 80% of proteins (1,787) with an ANOVA p-value <0.05 do not have a strong pattern of regulation across the cell-cycle. Of the remaining nine clusters we see two clusters for proteins peaking in G2&M phase of the cell-cycle; three clusters each for proteins peaking in S-phase or G1; and a single cluster with one protein (DRBD17) downregulated in S-phase and upregulated in G1 and late G2&M. In total, this classifies 428 genes in clusters with patterns of cell-cycle regulated expression (Supplementary Table 4). We observe GO terms enriched in cell-cycle phases where we would expect them to be associated with. For example, both G2&M-phase clusters are associated with GO terms such as ‘M-phase’, ‘spindle assembly’, ‘chromosome segregation’ and ‘mitotic cell-cycle’. This result provides strength to the idea that we can begin to associate proteins of unknown function with roles in particular cell-cycle phases based on their clustering. To this end, 23 ‘hypothetical’ proteins are observed within G1-phase clusters, 71 in S-phase clusters and 124 in G2&M-phase clusters, indicating potential roles for these proteins in these distinct stages of cell division. Moreover, due to the ‘high-resolution’ temporal profile it is possible to distinguish between regulation of proteins involved in kinetoplast DNA replication (high early S) and nuclear genomic

DNA replication (high S), for example, identifying the mitochondrial helicase PIF1 upregulated in early S phase.

Surprisingly, the majority of proteins with a described cell-cycle GO term are found in our "Unchanging" cluster (Figure 4.15). If a protein performs a role during the cell-cycle we would expect a cell-cycle specific pattern of regulation, though this does not necessarily have to occur at the level of protein abundance. The unchanging, cell-cycle GO associated proteins, may therefore be regulated at the level of post-translational modification, or through modification of interaction partners or sub-cellular localisation, where protein abundance would remain relatively constant. Another explanation could be that the peptides used to quantify these proteins may be suffering effects of interference, leading to ratio compression, masking real changes in quantified protein abundance.

4.4.5 Comparative analysis of the cell-cycle regulated transcriptome and proteome

A comparison of a previously published transcriptomic cell-cycle study in PCF *T. brucei* (Archer et al., 2011) to our proteomic dataset highlights that transcriptomic analysis reached a much greater depth, identifying 93% of the proteins identified here, whereas we identified 67% of the transcripts identified. Both studies classify a similar number of genes as cell-cycle regulated, with 544 identified from transcriptomics and 428 identified in this study. However, there is surprisingly little overlap between these sets of transcripts/proteins, with only 85 in common (Figure 4.17a, Supplementary Table 5). Therefore, we now have evidence for the cell-cycle regulation of up to 887 individual transcripts/proteins from transcriptomic and/or proteomic studies.

As expected, the 85 proteins identified in common between both datasets contains many known cell-cycle regulated proteins, including CPC1, AUK1, CRK3 and

multiple kinetochore proteins (KKT1, 7 and 2). The classification of this group of proteins in two independent studies, increases our confidence that they are genuinely cell-cycle regulated. There are 343 proteins identified as regulated which are not corroborated by transcriptomic data; 302 (Figure 4.17b, Supplementary Table 6) of which are identified by transcriptomics and 41 (Figure 4.17c, Supplementary Table 8) which are not. In this class of proteins regulated at the protein level, and identified, but not regulated at the transcript level, an enrichment for GO terms associated with the cell-cycle is still observed (Figure 4.18b). Additionally, this protein set contains DOT1B, KIN-A, CYC4, CRK2 and multiple kinetochore proteins (KKT10, 17 and 5), demonstrating the identification of cell-cycle regulated proteins that do not appear to be regulated at the transcript level. Furthermore, an enrichment for the organelle fission GO term in this group is also observed, through the inclusion of glycosomal biogenesis proteins such as GIM5A and GIM5B, which are highly upregulated in early G1-phase time-points. Of the 41 proteins not identified by transcriptomics there is no enrichment for cell-cycle related GO terms (Figure 4.18c), and no known cell-cycle regulated proteins. Therefore, there is less confidence that these proteins are genuinely cell-cycle regulated.

From transcriptomic evidence, 459 mRNAs are classified as regulated which are not supported by the proteomic data produced here; 274 which are also identified at the protein level (Figure 4.17d, Supplementary Table 7), and 185 only identified by transcriptomics (Figure 4.17e, Supplementary Table 9). The former set of 274 transcripts is enriched for proteins associated with microtubule based movement (the flagellum) and DNA replication (Figure 4.18e). In this group many components of the trypanosome flagellum and various subunits of nuclear and kinetoplastid DNA polymerases are detected. The latter set of 185 transcripts, not detected in our dataset, is

also enriched for cell-cycle associated genes (Figure 4.18e), including CDC45, MCM5, CRK10, CYC8, KIN-B, DOT1A, PLK and multiple kinetochore components (KKT8, 11, 13 and 9).

These results demonstrate the complementarity of both datasets, as although there is little overlap in the transcripts/proteins classified as cell-cycle regulated, both are successful in identifying known regulated transcripts/proteins which the other dataset does not identify. There are a number of reasons why both experiments identify different sets of transcripts/proteins. For example, utilising proteomic techniques, it is a challenge to reliably identify and quantify low abundance proteins, as evidenced by our ability to identify only 67% of the transcripts identified. Many cell-cycle regulated proteins are low abundance, hence it is no surprise that, particularly in the class of transcripts not identified in our dataset (Figure 4.17e), there are many canonical cell-cycle proteins only identified by transcriptomics. Moreover, it is not surprising that some transcripts/proteins are only identified as regulated from proteomic evidence, as protein abundance can be regulated by factors independent of mRNA abundance, such as the rate of translation and protein degradation.

Using the remaining 85 transcripts/proteins found in common to be cell-cycle regulated between both datasets, we compared the classification of the cell-cycle phases that the transcript/protein peaks in. There appears to be broad agreement in the classification of G1-phase upregulated transcripts and proteins, with 29 out of 34 proteins or 52 transcripts both classed as G1. For S-phase and G2&M-phase classifications there appears to be more variation. From 23 S-phase classified transcripts, 18 are identified in G2&M-phase clusters in our proteomic data. Similarly, of 14 proteins classified as S-phase, 12 are classed as G1-phase in transcriptomic data. Of the 37 proteins classed as G2&M-phase, only 8 are similarly annotated in the

transcriptomic data. These results may indicate a lag between an increase in mRNA abundance translating into an increase in protein abundance, as we see the that S-phase and G2&M-phase proteome classified proteins are mainly identified as G1 and S-phase transcripts, respectively in transcriptomic data. Alternatively, the experimental design in our proteomic study may allow for more accurate classification of peak expression, due to a higher temporal resolution, using nine time-points, compared to four in the transcriptome study.

4.4.6 Cell-cycle regulatory role of PSP1 domain proteins

We would also like to note the enrichment of polymerase suppressor 1 (PSP1) domain containing proteins within the group of 887 transcripts/proteins with evidence for cell-cycle regulation. The PSP1 protein was first discovered in yeast, where it was found to suppress mutations in temperature sensitive DNA polymerases (Formosa and Nittis, 1998). The C-terminus of PSP1 contains a domain that is found in up to 13 proteins in *T. brucei* (Table 4.3). Two of these proteins have homologs in *Crithidia fasciculata* (RBP33 and RBP45) that are subunits of the cycling sequence binding protein (CSBP II), which bind directly to mRNAs that periodically accumulate across the cell-cycle. RBP33 and RBP45 are also known to be differentially phosphorylated across the cell-cycle, which may regulate their interaction with mRNA (Mittra and Ray, 2004). Of the remaining 11 PSP domain containing proteins in *T. brucei*, 4 are classified as cell-cycle regulated in both transcriptomic and proteomic datasets, and one more in the transcriptomic data alone. All four proteins detected are in the top-22 most significantly changing proteins in the proteomic data with maximum fold-changes across the cell-cycle >3.3 (Table 4.3). As there is now evidence for cell-cycle regulation of 7 out of 13 PSP1 domain containing proteins in *T. brucei*, either through changes in abundance or phosphorylation, we propose that this domain may be a conserved domain

intimately involved in cell-cycle associated processes in kinetoplastids.

4.4.7 Identification of novel cell-cycle regulated proteins

From the 428 proteins in the nine clusters with patterns of cell-cycle regulation, only 14 are associated with a cell-cycle GO term (Figure 4.15). We are therefore potentially describing novel cell-cycle associated functions for hundreds of proteins in *T. brucei*. However, within this group we find a few proteins such as PIF1, thymidine kinase and PUF9, all known to have key functions during cell division, but lacking GO annotation. This result highlights the need for better curation of trypanosomatid database resources, and studies such as this can contribute evidence through the data produced. It is also obvious from Figure 4.16 that proteins upregulated in G2&M-phase of the cell-cycle are more likely to be annotated, showing the bias of the field towards the study of mitosis.

To further identify novel proteins essential for the cell-cycle in trypanosomatids, our dataset was filtered to only display 'hypothetical' proteins that are essential for the growth of the parasites in culture (Figure 4.16)(Alsford et al., 2011). Of the 287 essential proteins in cell-cycle regulated clusters, 57 are classed as 'hypothetical' with fold-changes up to 4.13 across the time-course. That these proteins are changing in abundance across the cell-cycle, and are essential for growth in culture, points to the idea that their essentiality may be due to their role in cell-division of the organism. As these proteins are classed as 'hypothetical', and they lack obvious sequence homology to proteins characterised in other eukaryotes, they could be useful candidates for drug targets which would specifically interfere with trypanosomatid cell division.

Gene ID	Gene description	Evidence for cell-cycle regulation
Tb927.3.5080	PSP1 C-terminal conserved region, putative	ID in transcriptome – not changing
Tb927.8.3850	PSP1 C-terminal conserved region, putative	ID in transcriptome – late G1
Tb927.9.9370	PSP1 C-terminal conserved region, putative	ID in transcriptome – not changing
Tb927.10.8330	*S. cerevisiae* PSP1 homologue, putative	ID in transcriptome – late G1; ID in proteome – S-phase 3.5 MaxFC
Tb927.10.9910	PSP1 C-terminal conserved region, putative	ID in transcriptome and proteome – not changing
Tb927.10.11630	PSP1 C-terminal conserved region, putative	ID in transcriptome and proteome – not changing
Tb927.11.4180	PSP1 C-terminal conserved region, putative	ID in transcriptome – late G1; ID in proteome – S-phase 3.3 MaxFC
Tb927.11.14750	PSP1 C-terminal conserved region, putative	ID in transcriptome – late G1; ID in proteome – S-phase 4.1 MaxFC
Tb927.5.760	cell-cycle sequence binding phosphoprotein (RBP33), putative	ID in transcriptome – early G1 ID in proteome – not changing Cell-cycle phosphorylation (Mittra and Ray, 2004)
Tb927.6.2000	spliceosome associated protein, putative	ID in transcriptome and proteome – not changing
Tb927.6.2850	ESAG associated protein, putative (PIE8)	D in transcriptome – S; ID in proteome – S 4.3 MaxFC
Tb927.10.9330	hypothetical protein, conserved	ID in transcriptome and proteome – not changing
Tb927.11.7140	cell-cycle sequence binding phosphoprotein (RBP45), putative	ID in transcriptome and proteome – not changing Cell-cycle phosphorylation (Mittra and Ray, 2004)

Table 4.3: Evidence of cell-cycle regulation of PSP1 C-terminal domain containing proteins in *T. brucei*. Genes identified as containing PSP1 domains through Interpro annotation. MaxFC stands for maximum fold-change detected in the proteomic dataset. Transcript/protein evidence annotation is based on the classification into cell-cycle phase in *(Archer et al., 2011)*, or into cell-cycle regulated clusters (Figure 4.13) in proteomic analysis.

We highlight here a number of proteins we think are interesting for follow-up studies based on their pattern of expression and/or their predicted domain structure and function. For example, Tb927.8.710 (DRBD17) is predicted to contain two RNA binding domains and in our proteomic dataset is strongly downregulated between the 5 and 6 h time-points, in S-phase of the cell-cycle. Transcriptomic data from elutriated trypanosomes shows an anti-correlated pattern between the protein and mRNA data, with mRNA abundance peaking at a 5.5 h time-point in S-phase (Figure 4.19)(Archer et al., 2011). The functional relevance of this anti-correlation between RNA and protein is

unknown, though potentially indicates a crucial role in cell-cycle regulated expression of mRNA transcripts, a mechanism already described for other RNA binding proteins such as PUF9 (Archer et al., 2009).

A 'predicted WD40 repeat protein' (Tb927.10.10990) is seen peaking in expression at the 9 h time-point in the "high G2&M" cluster (Figure 4.20). BLAST-P searches indicate homology to cell-division-cycle protein 20 (cdc20), an activator of the Anaphase Promoting Complex/Cyclosome (APC/C) E3 ubiquitin ligase, recruiting substrates for ubiquitin mediated degradation once cells are ready to segregate replicated chromosomes during mitosis (Barford, 2011). However, as it lacks a KEN amino acid motif, and another gene in *T. brucei* is already annotated as a cdc20 orthologue (Tb927.8.6500), we propose that this gene is a cdh1 orthologue. Cdh1 is highly structurally homologous to cdc20 and also recruits substrates to the APC/C, including cdc20 itself, acting slightly later as cells exit mitosis, ensuring the mitotic apparatus is degraded prior to re-entry into G1 (Barford, 2011).

In a similar manner a putative 'serine-threonine protein kinase' (Tb927.6.5100) and a 'hypothetical' protein (Tb927.8.7540), are both upregulated in G1-phase, that have BLAST-P hits to polo-like kinase (PLK) and cyclin dependent kinase (CDK) domains (Figures 4.21 and 4.22). Proteins with these types of domain are known to play key regulatory roles during the cell-cycle of other organismsm and we propose these genes should be annotated as a likely PLK and CDK, respectively.

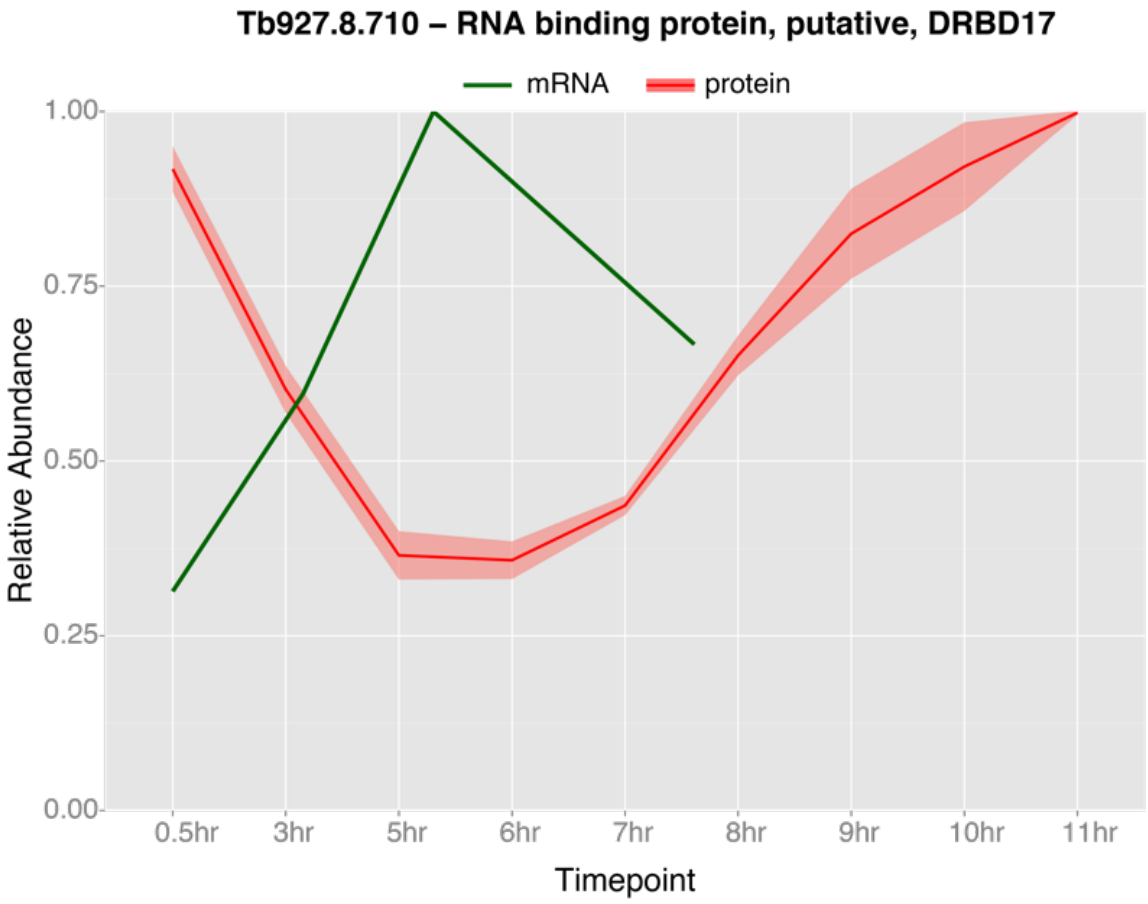


Figure 4.19: Relative protein and mRNA abundance of predicted DRBD17, Tb927.8.710 over the cell cycle. Red line shows mean of three biological replicates and red ribbon represents the standard deviation of measurements of TMT reporter ion intensity. Green line shows mRNA abundance (maximum normalized RPKM) as quantified in *(Archer et al., 2011)*.

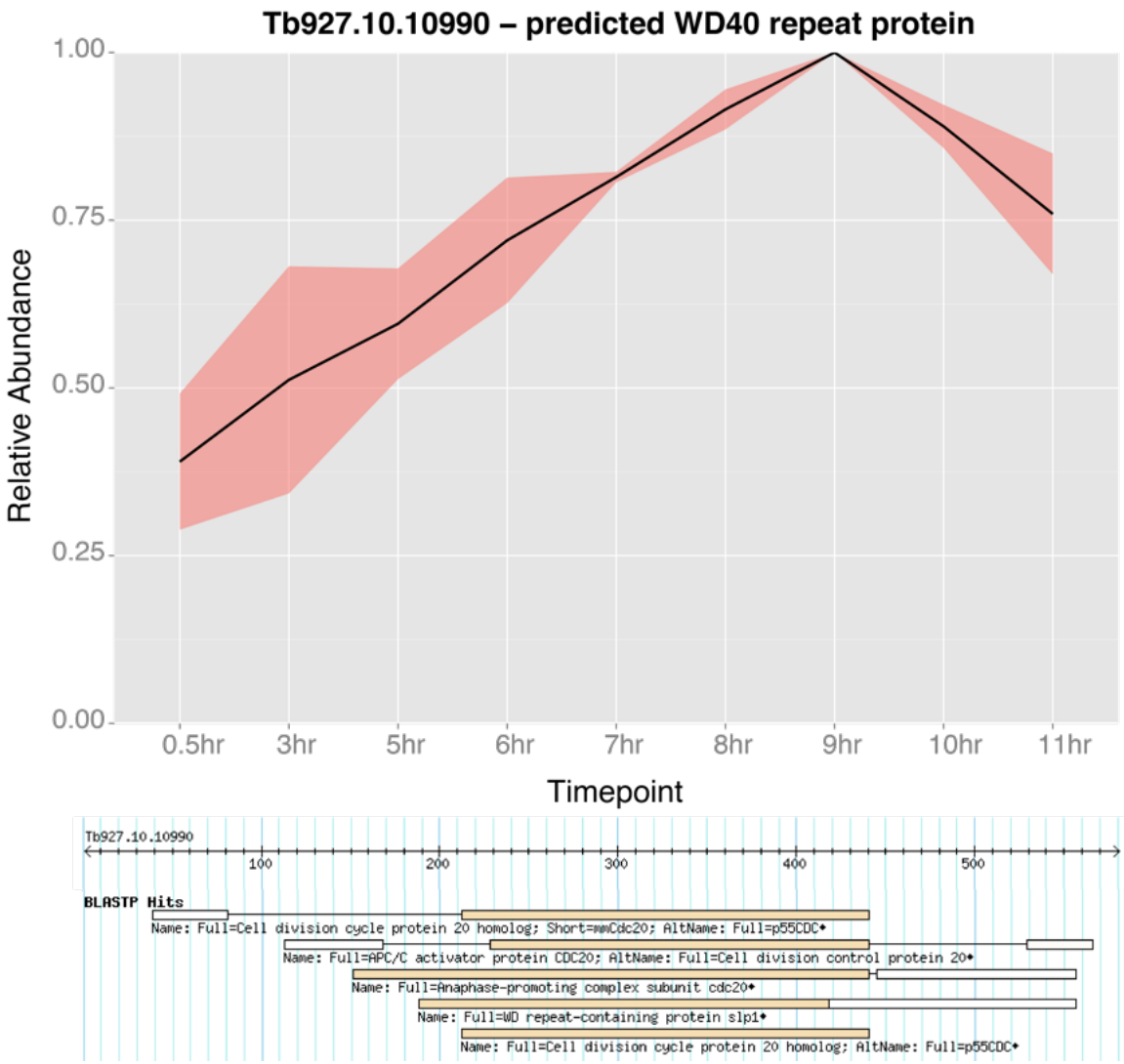


Figure 4.20: Relative protein abundance of predicted WD40 repeat protein, Tb927.10.10990 over the cell cycle. Black line shows mean of three biological replicates and red ribbon represents the standard deviation of measurements. Diagram below shows BLAST-P hits of Tb927.10.10990 displayed on TriTrypDB *(Aslett et al., 2010)*, with multiple hits to cdc20 related proteins.

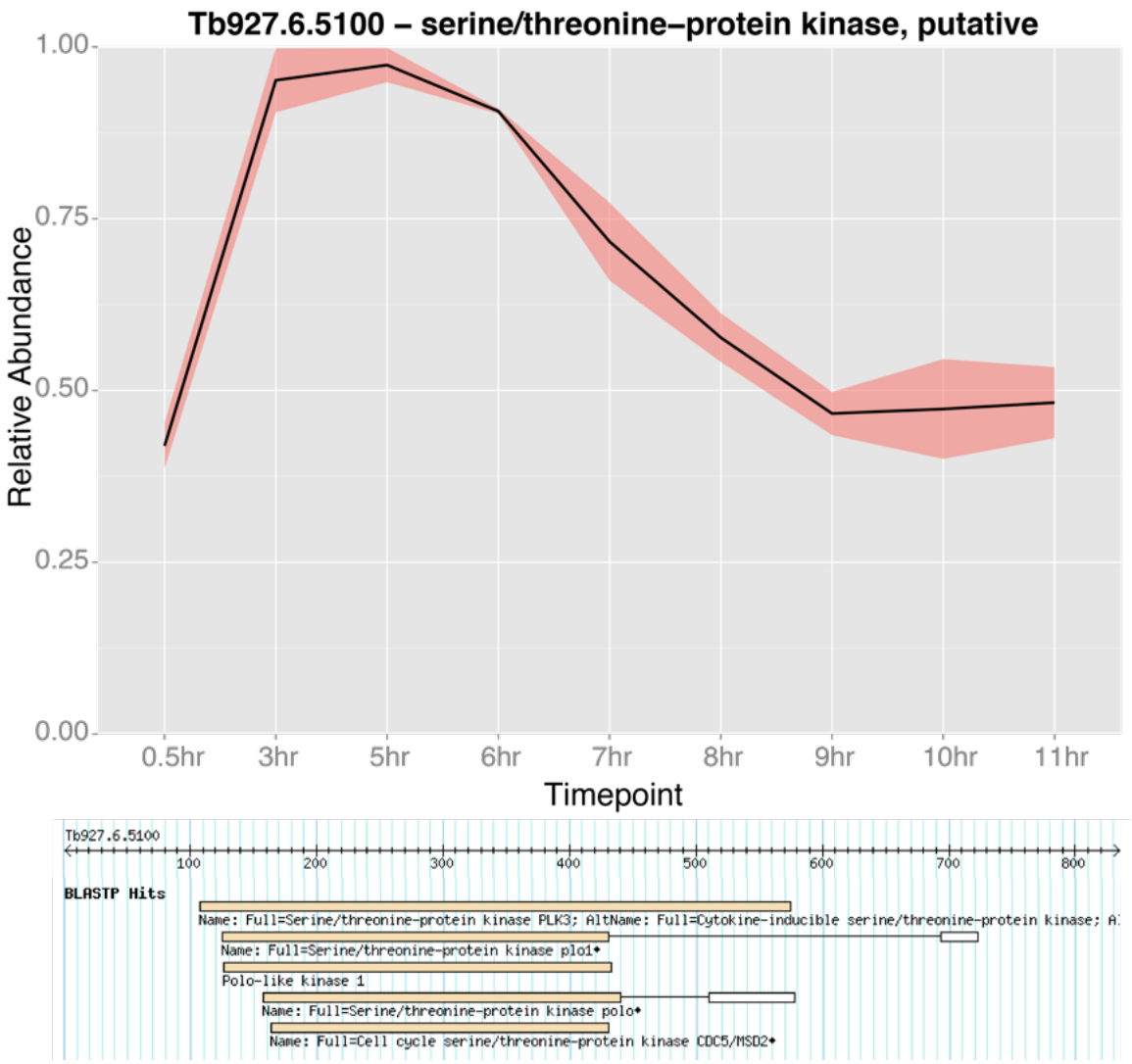


Figure 4.21: Relative protein abundance of putative S/T protein kinase, Tb927.6.5100 over the cell cycle. Black line shows mean of three biological replicates and red ribbon represents the standard deviation of measurements. Diagram below shows BLAST-P hits of Tb927.6.5100 displayed on TriTrypDB *(Aslett et al., 2010)*, with multiple hits to polo-like kinases.

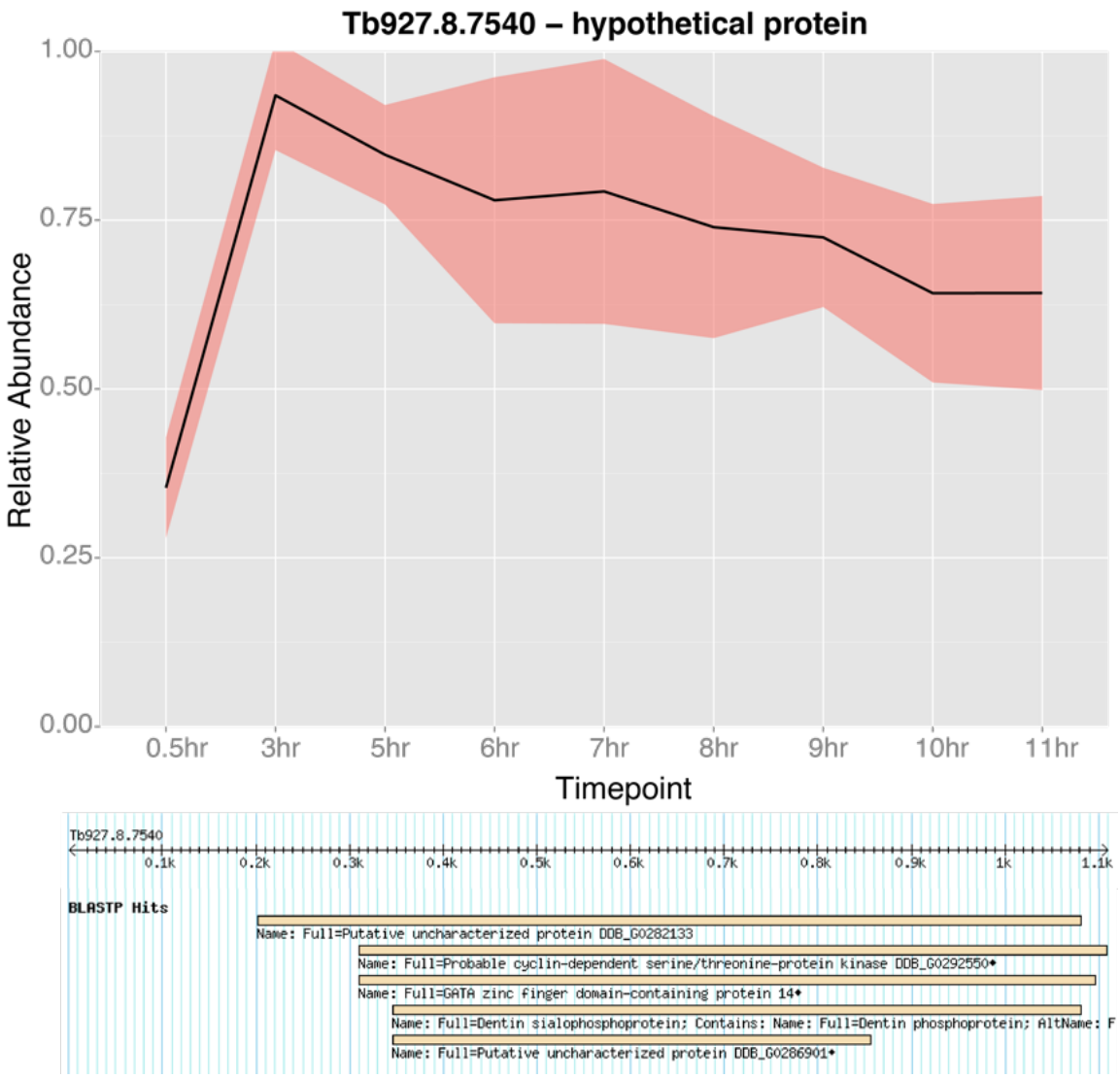


Figure 4.22: Relative protein abundance of hypothetical protein, Tb927.8.7540 over the cell cycle. Black line shows mean of three biological replicates and red ribbon represents the standard deviation of measurements. Diagram below shows BLAST-P hits of Tb927.8.7540 displayed on TriTrypDB *(Aslett et al., 2010)*, with a single hit to a cyclin-dependent kinase protein.

4.4.8 Summary

In summary, I present here the development of drug-free methodologies to study cell-cycle regulation of the *T. brucei* proteome through centrifugal elutriation. I have advanced mass spectrometry methods to enhance the accurate quantitation of isobaric tagging technology. Using 10-plex TMT and extensive fractionation of samples I have produced a comprehensive dataset, which quantifies 5,034 proteins in PCF *T. brucei* with a high temporal resolution across nine time-points of the cell-cycle. Through this analysis I have identified ~400 cell-cycle regulated proteins and classified these into different patterns of regulation, elucidating the biological role of many proteins with previously little or no evidence for function. Crucially, I identify 285 cell-cycle regulated proteins which are essential for growth of the parasite in culture, 57 of which are 'hypothetical' proteins, providing a comprehensive list of putative drug-targets with a role in cell-division of the parasite.

Chapter 5: Future directions and perspectives

5.1 Developments in understanding protein function in Trypanosoma brucei

One of the major challenges for the kinetoplastid research community is to understand the biological function of thousands of uncharacterised gene products identified from genome sequencing (Field et al., 2012). An understanding of protein function is important to further appreciate fundamental cell biological mechanisms of trypanosomes, and for the development of trypanocidal drugs with a clearer idea of putative mechanisms of action. In Chapter 1, I outlined a number of strategies which are being utilised to provide evidence for protein or gene function, ranging from computational, transcriptomic, proteomic and genetic techniques. In this thesis I have tried to contribute to this effort by applying unbiased proteomic methods, to understand protein-protein interactions, and cell cycle regulation of protein abundance in PCF *T. brucei*.

I have developed techniques to study soluble core protein complexes in *T. brucei*. Through the use of conventional chromatographic techniques, to separate complexes in whole cell lysates, in combination with quantitative mass spectrometry, I have produced information on elution profiles for 5,854 proteins. From this data we predicted 805 proteins in 234 protein complexes, including 135 proteins annotated as "hypothetical" and 488 proteins annotated with "putative" in their gene name. Protein-protein interactions are highly informative about the biological function of proteins, and I have highlighted key examples in Chapter 3 which demonstrate the power of our predictions to improve the annotation of the trypanosomatid proteome.

My analysis of the cell-cycle regulated proteome of PCF *T. brucei* identified and quantified the relative abundance of 5,034 proteins, across nine time-points over the course of 11 hours of cell-division. Of these, 428 were hierarchically clustered into groups with patterns of cell-cycle regulation; 219 were annotated as "hypothetical" and 139 had gene names annotated as "putative". The association of "hypothetical" or "putative" proteins with proteins of known function, and particular biological processes, such as DNA synthesis or mitosis, begins to provide evidence for function in a "guilt by co-regulation" approach, and leads to testable hypotheses of these functions. The identification of a cell-cycle regulated pattern of expression also throws new light onto putative predicted functions of proteins. For example, proteins with RNA binding domains may be playing key roles in translational control of proteins necessary for cell-cycle control, or proteins with kinase domains may be important players in signalling events necessary for cell-cycle progression.

I have highlighted, and can only work on, a handful of the proteins for which there may be novel evidence for protein function. A crucial component of large omics datasets is ensuring that the wider research community can access and work with the data produced in a fun and easy format. We have produced data visualisation applications for each of the datasets described in this work, which will allow the research community to explore this data and begin to identify other crucial proteins which they themselves may interested in characterising further.

5.2 Furthering analysis of protein complexes

I have taken a fairly straight-forward approach to my study of protein complexes, by simply lysing cells in PBS or ethanolamine, and loading the resulting lysates onto SEC or SAX columns. The use of orthogonal modes of chromatography increased the confidence in the predictions of complexes co-eluting across two conditions, and

allowed for the detection of complexes only stable under one condition. In future, the use of alternative chromatographic methods, such as hydrophobic interaction chromatography, may further increase the coverage of protein complexes and increase the confidence of predictions made so far.

Improvements in the global mapping of core protein complexes could also be made through subcellular fractionation of trypanosomes, prior to chromatographic fractionation. This would prevent complexes of similar size/shape or charge from co-eluting in SEC or SAX, allow for the reliable detection of lower abundance proteins that may have been excluded in this study and reduce the complexity of data analysis.

Moreover, under the 'native' lysis conditions used in these studies it is unlikely I would have solubilised membrane protein complexes. It would be interesting to test the effects of adding mild detergents, commonly used in immunoprecipitation studies, such as NP-40, digitonin, n-octyl glucoside or CHAPS, to lysis and fractionation buffers on the extraction and separation of membrane protein complexes. It is likely that buffers would have to be optimised for different subsets of complexes, and may result in conditions that disrupt other soluble cytoplasmic complexes.

Alternatively, it is possible to physically cross-link proteins *in vivo*, covalently capturing protein interactions in a physiological state. In the final stages of the work presented here I started the optimisation of formaldehyde and DSP cross-linking protocols for PCF *T. brucei*, followed by lysis in harsh SDS containing buffers. This methodology has previously been used to extract and fractionate membrane protein complexes in PCP-MS studies of human cancer cell lines (Larance et al., 2016).

Protein crosslinking mass spectrometry is commonly associated with the structural analysis of highly purified protein complexes. The heterogeneity and low

abundance of cross-linked peptides, in combination with complicated data analysis, has limited the use of these techniques in large-scale proteomic approaches. Many strategies have been implemented to counter these associated caveats, making global detection of cross-linked peptides more feasible. This has been made possible through: Affinity enrichable crosslinking reagents, to directly enrich for crosslinked peptides. Isotope labelled cross-linkers, allowing the detection of low abundance ions through the identification of specific mass shifts. MS-cleavable cross-linkers, allowing for simpler identification of cross-linked peptides (Tang and Bruce, 2010). The identification of directly cross-linked peptides would provide strong evidence to support the predictions of protein complexes made in this study, which are based on co-fractionation rather than any direct physical interaction.

Finally, we have demonstrated that elution profiles of individual proteins are highly reproducible between biological replicates. It is therefore possible to compare the elution profiles of the same protein between different biological conditions to determine if proteins have changed interaction state, possibly informing on biological function. Future studies could therefore compare BSF and PCF *T. brucei*, allowing for the detection of global changes in not just protein abundance, but protein interaction states between these distinct life-cycle stages. Furthermore, the formation of protein complexes at different stages of the cell-cycle is important for cell-cycle progression. The combination of elutriation with PCP-MS could provide key insights into the remodelling of the interactome over the cell-cycle in PCF trypanosomes.

5.3 Furthering analysis of the cell-division-cycle

I have comprehensively characterised the cell-cycle regulated proteome of PCF trypanosomes, using 10-plex TMT technology to produce a high resolution temporal map of relative protein abundance. However, there are a number of questions still to be

answered about cell-cycle control in these unusual eukaryotes. It is clear that different mechanisms control the cell-cycle in different life-stages of the parasite. Differences have been observed at the morphological level, with different positioning of nuclei and kinetoplasts prior to cytokinesis (Wheeler et al., 2013), and at the molecular level, as depletion of cell-cycle regulators by RNAi, or inhibition by small molecules, leads to different phenotypes. For example, the inhibition of DNA synthesis or the depletion of a mitotic cyclin, or the partner cyclin dependent kinase, results in a block in cytokinesis in BSF, yet PCF progress through cytokinesis (Hammarton, 2007; Li, 2012). Previous publications have indicated that BSF parasites are not amenable to elutriation as they are too small, however the parameters used for elutriation in this work were poorly reported (Archer et al., 2011). It is therefore worthwhile revisiting the elutriation of BSF trypanosomes in an attempt to carry out a comparative proteomic analysis to PCF cells, potentially revealing life-cycle specific factors controlling and regulating cell-division.

In this work, I utilised 10-plex TMT technology for a number of reasons. It reduced the amount of mass spectrometry acquisition time required by a factor of ten-fold. I could quantitate up to ten samples across a time-course in an individual mass spectrometry run, allowing for a high temporal resolution of our analysis of cell division. The nature of TMT quantitation, where all ten samples are quantified in a single mass spectrometry scan also reduced the number of missing values, a major caveat in label-free analysis when comparing many samples. Unfortunately, there is a well-documented effect of ratio compression when quantifying peptides using isobaric tags (Christoforou and Lilley, 2012), even when using SPS-MS3 methods on a Fusion MS. Although, this effect is unlikely to prevent the detection of proteins with high abundance changes between time-points, it may obscure proteins that have small changes in abundance, or proteins which are identified with low numbers of peptides.

Alternative methods of quantitation may reveal some of those proteins missed in this work, for example it would be possible to carry out a similar analysis using label-free quantitation either using Data Dependent Acquisition (DDA) or Data Independent Acquisition (DIA), as these methods have larger dynamic ranges and more robust quantitation. Conventional DDA analysis may require a reduction in the number of time-points collected, especially as extensive fractionation of peptides would be required for deep proteomic coverage using this method, leading to a large number of samples for analysis. Otherwise, publications utilising DIA-MS methods have demonstrated deep proteomic coverage for single-shot runs of complex lysates, without the difficulties of 'missing values' observed when using DDA methods (Bruderer et al., 2015). DIA-MS may therefore allow for a similar proteomic coverage seen in this dataset, without the need for a large number of samples that a DDA method would require.

Although protein abundance is a crucial factor for the regulation of biological processes such as the cell-cycle, it is not the only factor involved in such regulation. Post-translational modifications of proteins can change the function of a protein by affecting enzyme activity or modifying protein interaction partners, without a change in the abundance of a protein. It is also known that post-translational modifications such as phosphorylation play important regulatory roles across the cell-cycle (Fisher et al., 2012; Mocciaro and Rape, 2012). It would therefore be of interest to identify how phosphorylation varies across the cell-cycle on a proteome wide scale, potentially identifying regulated phospho-sites on proteins important for cell-division in trypanosomes. Identifying co-regulated phospho-sites may also allow for the identification of particular amino acid motifs that are the target of specific kinases. Furthermore, a number of kinases which are involved in cell-cycle regulation are

already known in *T. brucei*; for example, aurora-B kinase or cyclin-dependent kinases. However, the substrates of these kinases are unknown. It may be possible to identify the direct targets of these kinases through phospho-proteomic analysis of cells depleted of these enzymes, or through treatment of cells with specific inhibitors of these enzymes.

Finally, for the identification of putative drug candidates it is important to select proteins that are essential for the survival of the parasite. However, it is not necessarily the case that because a protein is regulated in terms of abundance or post-translational modification that it plays an essential role in cell-division. Therefore, in this work I performed a cross-comparison of our data to previously published RITseq screens that identified proteins essential for survival of *T. brucei* in cell culture (Alsford et al., 2011). If proteins are cell-cycle regulated and essential, it is likely that these proteins will be critical for cell-cycle progression. By understanding that these essential proteins are cell-cycle regulated we can also begin to further understand the mechanism of why these proteins are critical for parasite survival. Further evidence through RITseq studies with assays specifically designed to identify genes important for cell-division would provide another valuable data-set, with direct evidence for gene function. A theoretical example of such an assay is through the use of carboxyfluorescein succinimdyl ester (CFSE), which is used to stain cells in a pulse, followed by a chase where the number of cell-divisions can be counted through the cellular fluorescence of CFSE (Claes et al., 2009). It would therefore be possible to identify genes which are enriched in populations that arrest after one, two, three or four divisions following depletion by tetracycline induced RNAi in comparison to a non-induced sample.

5.4 Prediction of gene function

Together with other large-scale datasets produced from transcriptomic, proteomic or genetic techniques, the data produced here will contribute to our understanding of gene and protein function in this understudied eukaryote. These experiments by themselves can provide important information on individual biological questions, such as which proteins are cell-cycle regulated; which proteins form the flagella; or what proteins are depleted upon knock-out of another. However, they can also be highly useful when combined together to ask questions globally about protein function. Proteins involved in similar biological pathways, processes or complexes may demonstrate co-behaviours in these experiments, and the future computational analysis of such data may therefore identify groups of co-behaving proteins, and further our functional understanding of the *T. brucei* proteome. This kind of analysis has already been performed successfully on transcriptomic data from trypanosomes to annotate 'hypothetical' genes with unknown function, and therefore holds great promise for future analysis with the ever increasing output of large-scale data from many studies across the *T. brucei* field.

References

Acestor, N., Panigrahi, A.K., Ogata, Y., Anupama, A., and Stuart, K.D. (2009). Protein composition of Trypanosoma brucei mitochondrial membranes. Proteomics *9*, 5497–5508.

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., et al. (2012). The revised classification of eukaryotes. J. Eukaryot. Microbiol. *59*, 429–493.

Akiyoshi, B., and Gull, K. (2013). Evolutionary cell biology of chromosome segregation: insights from trypanosomes. Open Biol *3*, 130023.

Akiyoshi, B., and Gull, K. (2014). Discovery of unconventional kinetochores in kinetoplastids. Cell *156*, 1247–1258.

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell.

Alphey, M.S., Bond, C.S., Tetaud, E., Fairlamb, A.H., and Hunter, W.N. (2000). The structure of reduced tryparedoxin peroxidase reveals a decamer and insight into reactivity of 2Cys-peroxiredoxins. Journal of Molecular Biology *300*, 903–916.

Alsford, S., Eckert, S., Baker, N., Glover, L., Sanchez-Flores, A., Leung, K.F., Turner, D.J., Field, M.C., Berriman, M., and Horn, D. (2012). High-throughput decoding of antitrypanosomal drug efficacy and resistance. Nature *482*, 232–236.

Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-Fowler, C., and Horn, D. (2011). High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. Genome Research *21*, 915–924.

Andersen, J.S., Wilkinson, C.J., Mayor, T., Mortensen, P., Nigg, E.A., and Mann, M. (2003). Proteomic characterization of the human centrosome by protein correlation profiling. Nature *426*, 570–574.

Archer, S.K., Inchaustegui, D., Queiroz, R., and Clayton, C. (2011). The cell cycle regulated transcriptome of Trypanosoma brucei. PLoS ONE *6*, e18425.

Archer, S.K., Luu, V.-D., de Queiroz, R.A., Brems, S., and Clayton, C. (2009). Trypanosoma brucei PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. PLoS Pathog. *5*, e1000565.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., et al. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Research *38*, D457–D462.

Bacchi, C., Nathan, H., Hutner, S., McCann, P., and Sjoerdsma, A. (1980). Polyamine metabolism: a potential therapeutic target in trypanosomes. Science *210*, 332–334.

Baker, N., Hamilton, G., Wilkes, J.M., Hutchinson, S., Barrett, M.P., and Horn, D.

(2015). Vacuolar ATPase depletion affects mitochondrial ATPase function, kinetoplast dependency, and drug sensitivity in trypanosomes. Proc. Natl. Acad. Sci. U.S.a. *112*, 9112–9117.

Banfalvi, G. (2008). Cell cycle synchronization of animal cells and nuclei by centrifugal elutriation. Nat Protoc *3*, 663–673.

Banks, C.A.S., Boanca, G., Lee, Z.T., Florens, L., and Washburn, M.P. (2015). Proteins interacting with cloning scars: a source of false positive protein-protein interactions. Sci Rep *5*, 8530.

Barford, D. (2011). Structure, function and mechanism of the anaphase promoting complex (APC/C). Q. Rev. Biophys. *44*, 153–190.

Barrett, M.P., Boykin, D.W., Brun, R., and Tidwell, R.R. (2007). Human African trypanosomiasis: pharmacological re-engagement with a neglected disease. Br. J. Pharmacol. *152*, 1155–1171.

Barry, J.D., Graham, S.V., Fotheringham, M., Graham, V.S., Kobryn, K., and Wymer, B. (1998). VSG gene control and infectivity strategy of metacyclic stage Trypanosoma brucei. Molecular and Biochemical Parasitology *91*, 93–105.

Berberof, M., Vanhamme, L., Tebabi, P., Pays, A., Jefferies, D., Welburn, S., and Pays, E. (1995). The 3'-terminal region of the mRNAs for VSG and procyclin can confer stage specificity to gene expression in Trypanosoma brucei. Embo J. *14*, 2925–2934.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B., et al. (2005). The Genome of the African Trypanosome Trypanosoma brucei. Science *309*, 416–422.

Boulon, S., Ahmad, Y., Trinkle-Mulcahy, L., Verheggen, C., Cobley, A., Gregor, P., Bertrand, E., Whitehorn, M., and Lamond, A.I. (2010). Establishment of a protein frequency library and its application in the reliable identification of specific protein interaction partners. Molecular & Cellular Proteomics *9*, 861–879.

Brems, S., Guilbride, D.L., Gundlesdodjir-Planck, D., Busold, C., Luu, V.-D., Schanne, M., Hoheisel, J., and Clayton, C. (2005). The transcriptomes of Trypanosoma brucei Lister 427 and TREU927 bloodstream and procyclic trypomastigotes. Molecular and Biochemical Parasitology *139*, 163–172.

Brennand, A., Rico, E., Rigden, D.J., Van Der Smissen, P., Courtoy, P.J., and Michels, P.A.M. (2015). ATG24 Represses Autophagy and Differentiation and Is Essential for Homeostasy of the Flagellar Pocket in Trypanosoma brucei. PLoS ONE *10*, e0130365.

Bridges, D.J., Pitt, A.R., Hanrahan, O., Brennan, K., Voorheis, H.P., Herzyk, P., de Koning, H.P., and Burchmore, R.J.S. (2008). Characterisation of the plasma membrane subproteome of bloodstream form Trypanosoma brucei. Proteomics *8*, 83–99.

Broadhead, R., Dawe, H.R., Farr, H., Griffiths, S., Hart, S.R., Portman, N., Shaw, M.K., Ginger, M.L., Gaskell, S.J., McKean, P.G., et al. (2006). Flagellar motility is required for the viability of the bloodstream trypanosome. Nature *440*, 224–227.

Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinović, S.M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., et al. (2015). Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Molecular & Cellular Proteomics *14*, 1400–1410.

Brun, R., Blum, J., Chappuis, F., and Burri, C. (2010). Human African trypanosomiasis. The Lancet *375*, 148–159.

Butter, F., Bucerius, F., Michel, M., Cicova, Z., Mann, M., and Janzen, C.J. (2013). Comparative proteomics of two life cycle stages of stable isotope-labeled Trypanosoma brucei reveals novel components of the parasite's host adaptation machinery. Molecular & Cellular Proteomics *12*, 172–179.

Carmena, M., Wheelock, M., Funabiki, H., and Earnshaw, W.C. (2012). The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. Nat Rev Mol Cell Biol *13*, 789–803.

Chan, K.Y., and Ersfeld, K. (2010). The role of the Kinesin-13 family protein TbKif13-2 in flagellar length control of Trypanosoma brucei. Molecular and Biochemical Parasitology *174*, 137–140.

Chowdhury, A.R., Zhao, Z., and Englund, P.T. (2008). Effect of hydroxyurea on procyclic Trypanosoma brucei: an unconventional mechanism for achieving synchronous growth. Eukaryotic Cell *7*, 425–428.

Christoforou, A.L., and Lilley, K.S. (2012). Isobaric tagging approaches in quantitative proteomics: the ups and downs. Anal Bioanal Chem *404*, 1029–1037.

Claes, F., Vodnala, S.K., van Reet, N., Boucher, N., Lunden-Miguel, H., Baltz, T., Goddeeris, B.M., Büscher, P., and Rottenberg, M.E. (2009). Bioluminescent imaging of Trypanosoma brucei shows preferential testis dissemination which may hamper drug efficacy in sleeping sickness. PLoS Negl Trop Dis *3*, e486.

Clayton, C. (2013). The regulation of trypanosome gene expression by RNA-binding proteins. PLoS Pathog. *9*, e1003680.

Clayton, C.E. (2002). Life without transcriptional control? From fly to man and back again. Embo J. *21*, 1881–1888.

Clemmens, C.S., Morris, M.T., Lyda, T.A., Acosta-Serrano, A., and Morris, J.C. (2009). Trypanosoma brucei AMP-activated kinase subunit homologs influence surface molecule expression. Experimental Parasitology *123*, 250–257.

Cohen, P. (2000). The regulation of protein function by multisite phosphorylation–a 25 year update. Trends in Biochemical Sciences *25*, 596–601.

Colasante, C., Ellis, M., Ruppert, T., and Voncken, F. (2006). Comparative proteomics of glycosomes from bloodstream form and procyclic culture form Trypanosoma brucei brucei. Proteomics *6*, 3275–3293.

Cross, G.A.M. (1975). Identification, purification and properties of clone-specific

glycoprotein antigens constituting the surface coat of Trypanosoma brucei. Parasitology *71*, 393.

Cross, G. (1979). Crossreacting determinants in the C-terminal region of trypanosome variant surface antigens. Nature *277*, 310–312.

Cross, G.A.M., Kim, H.-S., and Wickstead, B. (2014). Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427. Molecular and Biochemical Parasitology *195*, 59–73.

Dacks, J.B., Walker, G., and Field, M.C. (2008). Implications of the new eukaryotic systematics for parasitologists. Parasitology International *57*, 97–104.

DeGrasse, J.A., Chait, B.T., Field, M.C., and Rout, M.P. (2008). High-yield isolation and subcellular proteomic characterization of nuclear and subnuclear structures from trypanosomes. Methods Mol. Biol. *463*, 77–92.

Dejung, M., Subota, I., Bucerius, F., and Dindar, G. (2016). Quantitative Proteomics Uncovers Novel Factors Involved in Developmental Differentiation of Trypanosoma brucei. PLoS Pathog. *12*, e1005439.

Desy, S., Schneider, A., and Mani, J. (2012). Trypanosoma brucei has a canonical mitochondrial processing peptidase. Molecular and Biochemical Parasitology *185*, 161–164.

Dosil, M., and Bustelo, X.R. (2004). Functional characterization of Pwp2, a WD family protein essential for the assembly of the 90 S pre-ribosomal particle. J. Biol. Chem. *279*, 37385–37397.

Doyle, J.J., Hirumi, H., Hirumi, K., Lupton, E.N., and Cross, G.A.M. (1980). Antigenic variation in clones of animal-infective Trypanosoma brucei derived and maintained in vitro. Parasitology *80*, 359.

Dunkley, T.P.J., Watson, R., Griffin, J.L., Dupree, P., and Lilley, K.S. (2004). Localization of organelle proteins by isotope tagging (LOPIT). Mol. Cell Proteomics *3*, 1128–1134.

Dyer, N.A., Rose, C., Ejeh, N.O., and Acosta-Serrano, A. (2013). Flying tryps: survival and maturation of trypanosomes in tsetse flies. Trends in Parasitology *29*, 188–196.

Elledge, S.J. (1996). Cell cycle checkpoints: preventing an identity crisis. Science *274*, 1664–1672.

Färber, V., Erben, E., Sharma, S., Stoecklin, G., and Clayton, C. (2013). Trypanosome CNOT10 is essential for the integrity of the NOT deadenylase complex and for degradation of many mRNAs. Nucleic Acids Research *41*, 1211–1222.

Ferguson, M.A., Low, M.G., and Cross, G.A. (1985). Glycosyl-sn-1,2-dimyristylphosphatidylinositol is covalently linked to Trypanosoma brucei variant surface glycoprotein. J. Biol. Chem. *260*, 14547–14555.

Fevre, E.M., Picozzi, K., Jannin, J., and Welburn, S.C. (2006). Human African

trypanosomiasis: epidemiology and control. Advances in Parasitology *61*, 167–221.

Field, M.C., Adung'a, V., Obado, S., Chait, B.T., and Rout, M.P. (2012). Proteomics on the rims: insights into the biology of the nuclear envelope and flagellar pocket of trypanosomes. Parasitology *139*, 1158–1167.

Fisher, D., Krasinska, L., Coudreuse, D., and Novák, B. (2012). Phosphorylation network dynamics in the control of cell cycle transitions. J. Cell. Sci. *125*, 4703–4711.

Folgueira, C., and Requena, J.M. (2007). A postgenomic view of the heat shock proteins in kinetoplastids. FEMS Microbiol. Rev. *31*, 359–377.

Formosa, T., and Nittis, T. (1998). Suppressors of the temperature sensitivity of DNA polymerase alpha mutations in Saccharomyces cerevisiae. Mol. Gen. Genet. *257*, 461–468.

Forsythe, G.R., McCulloch, R., and Hammarton, T.C. (2009). Hydroxyurea-induced synchronisation of bloodstream stage Trypanosoma brucei. Molecular and Biochemical Parasitology *164*, 131–136.

Foster, L.J., de Hoog, C.L., Zhang, Y., Zhang, Y., Xie, X., Mootha, V.K., and Mann, M. (2006). A mammalian organelle map by protein correlation profiling. Cell *125*, 187–199.

Franco, J.R., Simarro, P.P., Diarra, A., and Jannin, J.G. (2014). Epidemiology of human African trypanosomiasis. Clin Epidemiol *6*, 257–275.

Gale, M., Carter, V., and Parsons, M. (1994). Cell cycle-specific induction of an 89 kDa serine/threonine protein kinase activity in Trypanosoma brucei. J. Cell. Sci. *107 (Pt 7)*, 1825–1832.

Gavin, A.-C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631–636.

Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141–147.

Gazestani, V.H., Lu, Z., and Salavati, R. (2014). Deciphering RNA regulatory elements in trypanosomatids: one piece at a time or genome-wide? Trends in Parasitology *30*, 234–240.

Gazestani, V.H., Nikpour, N., Mehta, V., Najafabadi, H.S., Moshiri, H., Jardim, A., and Salavati, R. (2016). A Protein Complex Map of Trypanosoma brucei. PLoS Negl Trop Dis *10*, e0004533.

Gheiratmand, L., Brasseur, A., Zhou, Q., and He, C.Y. (2013). Biochemical characterization of the bi-lobe reveals a continuous structural network linking the bi-lobe to other single-copied organelles in Trypanosoma brucei. Journal of Biological Chemistry *288*, 3489–3499.

Gibson, W. (2015). Liaisons dangereuses: sexual recombination among pathogenic trypanosomes. Research in Microbiology *166*, 459–466.

Gingras, A.-C., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol *8*, 645–654.

Glover, L., Hutchinson, S., Alsford, S., McCulloch, R., Field, M.C., and Horn, D. (2013). Antigenic variation in African trypanosomes: the importance of chromosomal and nuclear context in VSG expression control. Cellular Microbiology *15*, 1984–1993.

Gunasekera, K., Wüthrich, D., and Braga-Lagache, S. (2012). Proteome remodelling during development from blood to insect-form Trypanosoma brucei quantified by SILAC and mass spectrometry. BMC Genomics *13*, 556.

Gunzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L.C., Chung, H.M., Lee, P.T., and Lee, M.G.S. (2003). RNA Polymerase I Transcribes Procyclin Genes and Variant Surface Glycoprotein Gene Expression Sites in Trypanosoma brucei. Eukaryotic Cell *2*, 542–551.

Günzl, A., Kirkham, J.K., Nguyen, T.N., Badjatia, N., and Park, S.H. (2015). Mono-allelic VSG expression by RNA polymerase I in Trypanosoma brucei: expression site control from both ends? Gene *556*, 68–73.

Güther, M., Urbaniak, M.D., and Tavendale, A. (2014). High-confidence glycosome proteome for procyclic form Trypanosoma brucei by epitope-tag organelle enrichment and SILAC proteomics. J. Proteome Res. *13*, 2796–2806.

Hammarton, T.C., Engstler, M., and Mottram, J.C. (2004). The Trypanosoma brucei cyclin, CYC2, is required for cell cycle progression through G1 phase and for maintenance of procyclic form cell morphology. Journal of Biological Chemistry *279*, 24757–24764.

Hammarton, T.C. (2007). Cell cycle regulation in Trypanosoma brucei. Molecular and Biochemical Parasitology *153*, 1–8.

Hart, S.R., Lau, K.W., Hao, Z., Broadhead, R., Portman, N., Hühmer, A., Gull, K., McKean, P.G., Hubbard, S.J., and Gaskell, S.J. (2009). Analysis of the trypanosome flagellar proteome using a combined electron transfer/collisionally activated dissociation strategy. J Am Soc Mass Spectrom *20*, 167–175.

Hashimi, H., Cicova, Z., Novotná, L., Wen, Y.-Z., and Lukeš, J. (2009). Kinetoplastid guide RNA biogenesis is dependent on subunits of the mitochondrial RNA binding complex 1 and mitochondrial RNA polymerase. Rna *15*, 588–599.

Havugimana, P., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A., Li, Z., Wang, P., Boutz, D., Fong, V., Phanse, S., et al. (2012). A Census of Human Soluble Protein Complexes. Cell *150*, 1068–1081.

Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., et al. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell *163*, 712–723.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature *415*, 180–183.

Holden, J.M., Koreny, L., Obado, S., Ratushny, A.V., Chen, W.-M., Chiang, J.-H., Kelly, S., Chait, B.T., Aitchison, J.D., Rout, M.P., et al. (2014). Nuclear pore complex evolution: a trypanosome Mlp analogue functions in chromosomal segregation but lacks transcriptional barrier activity. Molecular Biology of the Cell *25*, 1421–1436.

Horn, D., and Cross, G. (1997). Analysis of Trypanosoma brucei vsg expression site switching in vitro. Molecular and Biochemical Parasitology *84*, 189–201.

Horn, D. (2014). High-throughput decoding of drug targets and drug resistance mechanisms in African trypanosomes. Parasitology *141*, 77–82.

Hu, H., Hu, L., Yu, Z., Chasse, A.E., Chu, F., and Li, Z. (2012). An orphan kinesin in trypanosomes cooperates with a kinetoplastid-specific kinesin to maintain cell morphology by regulating subpellicular microtubules. J. Cell. Sci. *125*, 4126–4136.

Hu, P., Janga, S.C., Babu, M., Díaz-Mejía, J.J., Butland, G., Yang, W., Pogoutse, O., Guo, X., Phanse, S., Wong, P., et al. (2009). Global functional atlas of Escherichia coli encompassing previously uncharacterized proteins. PLoS Biol. *7*, e96.

Huang, G., Ulrich, P.N., Storey, M., Johnson, D., Tischer, J., Tovar, J.A., Moreno, S.N.J., Orlando, R., and Docampo, R. (2014). Proteomic analysis of the acidocalcisome, an organelle conserved from bacteria to human cells. PLoS Pathog. *10*, e1004555.

Huang, J., and Van der Ploeg, L.H. (1991). Maturation of polycistronic pre-mRNA in Trypanosoma brucei: analysis of trans splicing and poly (A) addition at nascent RNA transcripts from the hsp70 locus. Mol. Cell. Biol. *11*, 3180–3190.

Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., et al. (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. Cell *162*, 425–440.

Imboden, M.A., Laird, P.W., and Affolter, M. (1987). Transcription of the intergenic regions of the tubulin gene cluster of Trypanosoma brucei: evidence for a polyclstronic transcription unit in a eukaryote. Nucleic Acids Research *15*, 7357–7368.

Inoue, M., Nakamura, Y., Yasuda, K., Yasaka, N., Hara, T., Schnaufer, A., Stuart, K., and Fukuma, T. (2005). The 14-3-3 proteins of Trypanosoma brucei function in motility, cytokinesis, and cell cycle. J. Biol. Chem. *280*, 14085–14096.

Janzen, C.J., Hake, S.B., Lowell, J.E., and Cross, G.A.M. (2006). Selective di- or trimethylation of histone H3 lysine 76 by two DOT1 homologs is important for cell cycle regulation in Trypanosoma brucei. Molecular Cell *23*, 497–507.

Jensen, B.C., Ramasamy, G., Vasconcelos, E.J.R., Ingolia, N.T., Myler, P.J., and Parsons, M. (2014). Extensive stage-regulation of translation revealed by ribosome profiling of Trypanosoma brucei. BMC Genomics *15*, 911.

Jensen, B.C., Sivam, D., Kifer, C.T., Myler, P.J., and Parsons, M. (2009). Widespread

variation in transcript abundance within and across developmental stages of Trypanosoma brucei. BMC Genomics *10*, 482.

Johnson, A.E., and van Waes, M.A. (1999). The translocon: a dynamic gateway at the ER membrane. Annu. Rev. Cell Dev. Biol. *15*, 799–842.

Johnson, P.J., Kooter, J.M., and Borst, P. (1987). Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene. Cell *51*, 273–281.

Jones, C., Anderson, S., Singha, U.K., and Chaudhuri, M. (2008). Protein phosphatase 5 is required for Hsp90 function during proteotoxic stresses in Trypanosoma brucei. Parasitol. Res. *102*, 835–844.

Kabani, S., Fenn, K., Ross, A., Ivens, A., Smith, T.K., Ghazal, P., and Matthews, K. (2009). Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in Trypanosoma brucei. BMC Genomics *10*, 427.

Kabani, S., Waterfall, M., and Matthews, K.R. (2010). Cell-cycle synchronisation of bloodstream forms of Trypanosoma brucei using Vybrant DyeCycle Violet-based sorting. Molecular and Biochemical Parasitology *169*, 59–62.

Kennedy, P.G.E. (2008). The continuing problem of human African trypanosomiasis (sleeping sickness). Ann. Neurol. *64*, 116–126.

Kirkwood, K.J., Ahmad, Y., Larance, M., and Lamond, A.I. (2013). Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. Molecular & Cellular Proteomics *12*, 3851–3873.

Koumandou, V.L., Natesan, S.K.A., Sergeenko, T., and Field, M.C. (2008). The trypanosome transcriptome is remodelled during differentiation but displays limited responsiveness within life stages. BMC Genomics *9*, 298.

Köcher, T., and Superti-Furga, G. (2007). Mass spectrometry-based functional proteomics: from molecular machines to protein networks. Nature Methods *4*, 807–815.

Kristensen, A.R., Gsponer, J., and Foster, L.J. (2012). A high-throughput approach for measuring temporal changes in the interactome. Nature Methods *9*, 907–909.

Kristjanson, P.M., Swallow, B.M., Rowlands, G.J., Kruska, R.L., and de Leeuw, P.N. (1999). Measuring the costs of African animal trypanosomosis, the potential benefits of control and returns to research. Agricultural Systems *59*, 79–98.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637–643.

Lacomble, S., Vaughan, S., Gadelha, C., Morphew, M.K., Shaw, M.K., McIntosh, J.R., and Gull, K. (2009). Three-dimensional cellular architecture of the flagellar pocket and associated cytoskeleton in trypanosomes revealed by electron microscope tomography. J. Cell. Sci. *122*, 1081–1090.

Larance, M., Kirkwood, K.J., Tinti, M., Brenes Murillo, A., Ferguson, M.A.J., and Lamond, A.I. (2016). Global Membrane Protein Interactome Analysis using In vivo Crosslinking and Mass Spectrometry-based Protein Correlation Profiling. Molecular & Cellular Proteomics *15*, 2476–2490.

Li, S.S., and Giometti, C.S. (2007). A combinatorial approach to studying protein complex composition by employing size-exclusion chromatography and proteome analysis. J Sep Sci *30*, 1549–1555.

Li, Z. (2012). Regulation of the cell division cycle in Trypanosoma brucei. Eukaryotic Cell *11*, 1180–1190.

Li, Z., and Wang, C.C. (2003). A PHO80-like cyclin and a B-type cyclin control the cell cycle of the procyclic form of Trypanosoma brucei. J. Biol. Chem. *278*, 20652–20658.

Li, Z., Lee, J.H., Chu, F., Burlingame, A.L., Günzl, A., and Wang, C.C. (2008a). Identification of a novel chromosomal passenger complex and its unique localization during cytokinesis in Trypanosoma brucei. PLoS ONE *3*, e2354.

Li, Z., Umeyama, T., and Wang, C.C. (2008b). The chromosomal passenger complex and a mitotic kinesin interact with the Tousled-like kinase in trypanosomes to regulate mitosis and cytokinesis. PLoS ONE *3*, e3814.

Li, Z., Zou, C.-B., Yao, Y., Hoyt, M.A., McDonough, S., Mackey, Z.B., Coffino, P., and Wang, C.C. (2002). An easily dissociated 26 S proteasome catalyzes an essential ubiquitin-mediated protein degradation pathway in Trypanosoma brucei. J. Biol. Chem. *277*, 15486–15498.

Liu, B., Wang, J., Yaffe, N., Lindsay, M.E., Zhao, Z., Zick, A., Shlomai, J., and Englund, P.T. (2009). Trypanosomes Have Six Mitochondrial DNA Helicases with One Controlling Kinetoplast Maxicircle Replication. Molecular Cell *35*, 490–501.

Lueong, S., Merce, C., Fischer, B., Hoheisel, J.D., and Erben, E.D. (2016). Gene expression regulatory networks in Trypanosoma brucei: insights into the role of the mRNA-binding proteome. Molecular Microbiology *100*, 457–471.

Lukeš, J., Skalický, T., Týč, J., Votýpka, J., and Yurchenko, V. (2014). Evolution of parasitism in kinetoplastid flagellates. Molecular and Biochemical Parasitology *195*, 115–122.

Lustig, Y., Goldshmidt, H., Uliel, S., and Michaeli, S. (2005). The Trypanosoma brucei signal recognition particle lacks the Alu-domain-binding proteins: purification and functional analysis of its binding proteins by RNAi. J. Cell. Sci. *118*, 4551–4562.

Luz Ambrósio, D., Lee, J.H., Panigrahi, A.K., Nguyen, T.N., Cicarelli, R.M.B., and Günzl, A. (2009). Spliceosomal proteomics in Trypanosoma brucei reveal new RNA splicing factors. Eukaryotic Cell *8*, 990–1000.

Ly, T., Endo, A., and Lamond, A.I. (2015). Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells. eLife Sciences *4*, M111.013680.

MacGregor, P., Szöőr, B., Savill, N.J., and Matthews, K.R. (2012). Trypanosomal

immune evasion, chronicity and transmission: an elegant balancing act. Nature Reviews Microbiology *10*, 431–438.

Mao, Y., Najafabadi, H.S., and Salavati, R. (2009). Genome-wide computational identification of functional RNA elements in Trypanosoma brucei. BMC Genomics *10*, 355.

Marcello, L., and Barry, J.D. (2007). Analysis of the VSG gene silent archive in Trypanosoma brucei reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. Genome Research *17*, 1344–1352.

Martínez-Calvillo, S., Vizuet-de-Rueda, J.C., Florencio-Martínez, L.E., Manning-Cela, R.G., and Figueroa-Angulo, E.E. (2010). Gene expression in trypanosomatid parasites. J. Biomed. Biotechnol. *2010*, 525241.

Mauxion, F., Prève, B., and Seraphin, B. (2013). C2ORF29/CNOT11 and CNOT10 form a new module of the CCR4-NOT complex. RNA Biol *10*, 267–276.

McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wühr, M., Huttlin, E.L., Erickson, B.K., Rad, R., Haas, W., and Gygi, S.P. (2014). MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. Anal. Chem. *86*, 7150–7158.

McKean, P.G. (2003). Coordination of cell cycle and cytokinesis in Trypanosoma brucei. Current Opinion in Microbiology *6*, 600–607.

Mehta, V., and Trinkle-Mulcahy, L. (2016). Recent advances in large-scale protein interactome mapping. F1000Res *5*.

Meier, U.T. (2006). How a single protein complex accommodates many different H/ACA RNAs. Trends in Biochemical Sciences *31*, 311–315.

Milne, K.G., Prescott, A.R., and Ferguson, M.A. (1998). Transformation of monomorphic Trypanosoma brucei bloodstream form trypomastigotes into procyclic forms at 37 degrees C by removing glucose from the culture medium. Molecular and Biochemical Parasitology *94*, 99–112.

Mittra, B., and Ray, D.S. (2004). Presence of a poly(A) binding protein and two proteins with cell cycle-dependent phosphorylation in Crithidia fasciculata mRNA cycling sequence binding protein II. Eukaryotic Cell *3*, 1185–1197.

Mocciaro, A., and Rape, M. (2012). Emerging regulatory mechanisms in ubiquitin-dependent cell cycle control. J. Cell. Sci. *125*, 255–263.

Mony, B.M., MacGregor, P., Ivens, A., Rojas, F., Cowton, A., Young, J., Horn, D., and Matthews, K. (2014). Genome-wide dissection of the quorum sensing signalling pathway in Trypanosoma brucei. Nature *505*, 681–685.

Morelle, C., Sterkers, Y., Crobu, L., MBang-Benet, D.-E., Kuk, N., Portalès, P., Bastien, P., Pagès, M., and Lachaud, L. (2015). The nucleoporin Mlp2 is involved in chromosomal distribution during mitosis in trypanosomatids. Nucleic Acids Research *43*, 4013–4027.

Morriswood, B., Havlicek, K., Demmel, L., Yavuz, S., Sealey-Cardona, M., Vidilaseris, K., Anrather, D., Kostan, J., Djinovic-Carugo, K., Roux, K.J., et al. (2013). Novel bilobe components in Trypanosoma brucei identified using proximity-dependent biotinylation. Eukaryotic Cell *12*, 356–367.

Murray, A.W. (2004). Recycling the cell cycle: cyclins revisited. Cell *116*, 221–234.

Nett, I.R.E., Martin, D.M.A., Miranda-Saavedra, D., Lamont, D., Barber, J.D., Mehlert, A., and Ferguson, M.A.J. (2009). The phosphoproteome of bloodstream form Trypanosoma brucei, causative agent of African sleeping sickness. Molecular & Cellular Proteomics *8*, 1527–1538.

Nguyen, S., Jones, D.C., Wyllie, S., Fairlamb, A.H., and Phillips, M.A. (2013). Allosteric activation of trypanosomatid deoxyhypusine synthase by a catalytically dead paralog. Journal of Biological Chemistry *288*, 15256–15267.

Niemann, M., Wiese, S., Mani, J., Chanfon, A., Jackson, C., Meisinger, C., Warscheid, B., and Schneider, A. (2012). Mitochondrial Outer Membrane Proteome of Trypanosoma brucei Reveals Novel Factors Required to Maintain Mitochondrial Morphology. Molecular & Cellular Proteomics *12*, 515–528.

Oberholzer, M., Langousis, G., Nguyen, H.T., Saada, E.A., Shimogawa, M.M., Jonsson, Z.O., Nguyen, S.M., Wohlschlegel, J.A., and Hill, K.L. (2011). Independent analysis of the flagellum surface and matrix proteomes provides insight into flagellum signaling in mammalian-infectious Trypanosoma brucei. Molecular & Cellular Proteomics *10*, M111.010538.

Ohta, S., Bukowski-Wills, J.-C., Sanchez-Pulido, L., Alves, F. de L., Wood, L., Chen, Z.A., Platani, M., Fischer, L., Hudson, D.F., Ponting, C.P., et al. (2010). The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. Cell *142*, 810–821.

Olinares, P.D.B., Ponnala, L., and van Wijk, K.J. (2010). Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. Molecular & Cellular Proteomics *9*, 1594–1615.

Palfi, Z., Schimanski, B., Günzl, A., Lücke, S., and Bindereif, A. (2005). U1 small nuclear RNP from Trypanosoma brucei: a minimal U1 snRNA with unusual protein components. Nucleic Acids Research *33*, 2493–2503.

Panigrahi, A.K., Ogata, Y., Zikova, A., Anupama, A., Dalley, R.A., Acestor, N., Myler, P.J., and Stuart, K.D. (2009). A comprehensive analysis of Trypanosoma brucei mitochondrial proteome. Proteomics *9*, 434–450.

Parsons, M., and Nielsen, B. (1990). Trypanosoma brucei: two-dimensional gel analysis of the major glycosomal proteins during the life cycle. Exp. Parasitol. *70*, 276–285.

Pays, E., and Nolan, D.P. (1998). Expression and function of surface proteins in Trypanosoma brucei. Molecular and Biochemical Parasitology *91*, 3–36.

Piñeyro, M.D., Pizarro, J.C., Lema, F., Pritsch, O., Cayota, A., Bentley, G.A., and

Robello, C. (2005). Crystal structure of the tryparedoxin peroxidase from the human parasite Trypanosoma cruzi. J. Struct. Biol. *150*, 11–22.

Priotto, G., Kasparian, S., Mutombo, W., and Ngouama, D. (2009). Nifurtimox-eflornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis: a multicentre, randomised, phase III, non- …. The Lancet *374*, 56–64.

Pullen, T.J., Ginger, M.L., Gaskell, S.J., and Gull, K. (2004). Protein targeting of an unusual, evolutionarily conserved adenylate kinase to a eukaryotic flagellum. Molecular Biology of the Cell *15*, 3257–3265.

Reuner, B., Vassella, E., Yutzy, B., and Boshart, M. (1997). Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. Molecular and Biochemical Parasitology *90*, 269–280.

Rezende, A.M., Assis, L.A., Nunes, E.C., da Costa Lima, T.D., Marchini, F.K., Freire, E.R., Reis, C.R., and de Melo Neto, O.P. (2014). The translation initiation complex eIF3 in trypanosomatids and other pathogenic excavates – identification of conserved and divergent features based on orthologue analysis. BMC Genomics *15*, 1175.

Rhee, H.-W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science (New York, N.Y.) *339*, 1328–1331.

Rout, M.P., and Field, M.C. (2001). Isolation and characterization of subnuclear compartments from Trypanosoma brucei. Identification of a major repetitive nuclear lamina component. J. Biol. Chem. *276*, 38261–38271.

Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. The Journal of Cell Biology *196*, 801–810.

Ruepp, A., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Stransky, M., Waegele, B., Schmidt, T., Doudieu, O.N., Stumpflen, V., et al. (2007). CORUM: the comprehensive resource of mammalian protein complexes. Nucleic Acids Research *36*, D646–D650.

Salavati, R., and Najafabadi, H.S. (2010). Sequence-based functional annotation: what if most of the genes are unique to a genome? Trends in Parasitology *26*, 225–229.

Santaguida, S., and Musacchio, A. (2009). The life and miracles of kinetochores. Embo J. *28*, 2511–2531.

Schwede, A., Ellis, L., Luther, J., Carrington, M., Stoecklin, G., and Clayton, C. (2008). A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells. Nucleic Acids Research *36*, 3374–3388.

Shimogawa, M.M., Saada, E.A., Vashisht, A.A., Barshop, W.D., Wohlschlegel, J.A., and Hill, K.L. (2015). Cell Surface Proteomics Provides Insight into Stage-Specific Remodeling of the Host-Parasite Interface in Trypanosoma brucei. Molecular & Cellular Proteomics *14*, 1977–1988.

Siegel, T.N., Hekstra, D.R., Wang, X., Dewell, S., and Cross, G.A.M. (2010). Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. Nucleic Acids Research *38*, 4946–4957.

Silva, M.T.A.D., Ambrósio, D.L., Trevelin, C.C., Watanabe, T.F., Laure, H.J., Greene, L.J., Rosa, J.C., Valentini, S.R., and Cicarelli, R.M. (2011). New insights into trypanosomatid U5 small nuclear ribonucleoproteins. Memórias Do Instituto Oswaldo Cruz *106*, 130–138.

Simarro, P.P., Cecchi, G., Franco, J.R., Paone, M., Diarra, A., Priotto, G., Mattioli, R.C., and Jannin, J.G. (2015). Monitoring the Progress towards the Elimination of Gambiense Human African Trypanosomiasis. PLoS Negl Trop Dis *9*, e0003785.

Simpson, A.G.B., Stevens, J.R., and Lukeš, J. (2006). The evolution and diversity of kinetoplastid flagellates. Trends in Parasitology *22*, 168–174.

Singha, U.K., Hamilton, V., Duncan, M.R., Weems, E., Tripathi, M.K., and Chaudhuri, M. (2012). Protein translocase of mitochondrial inner membrane in Trypanosoma brucei. Journal of Biological Chemistry *287*, 14480–14493.

Smits, A.H., and Vermeulen, M. (2016). Characterizing Protein-Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. Trends in Biotechnology.

Steverding, D. (2010). The development of drugs for treatment of sleeping sickness: a historical review. Parasit Vectors *3*, 15.

Stuart, K., Brun, R., Croft, S., Fairlamb, A., Gürtler, R.E., McKerrow, J., Reed, S., and Tarleton, R. (2008). Kinetoplastids: related protozoan pathogens, different diseases. J. Clin. Invest. *118*, 1301–1310.

Subota, I., Julkowska, D., Vincensini, L., Reeg, N., Buisson, J., Blisnick, T., Huet, D., Perrot, S., Santi-Rocca, J., Duchateau, M., et al. (2014). Proteomic analysis of intact flagella of procyclic Trypanosoma brucei cells identifies novel flagellar proteins with unique sub-localization and dynamics. Molecular & Cellular Proteomics *13*, 1769–1786.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE *6*, e21800.

Swallow, B.M. (2000). Impacts of Trypanosomiasis on African Agriculture.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Research *43*, D447–D452.

Tang, X., and Bruce, J.E. (2010). A new cross-linking strategy: protein interaction reporter (PIR) technology for protein-protein interaction studies. Mol. BioSyst. *6*, 939–947.

Tetley, L., Turner, C.M., Barry, J.D., Crowe, J.S., and Vickerman, K. (1987). Onset of

expression of the variant surface glycoproteins of Trypanosoma brucei in the tsetse fly studied using immunoelectron microscopy. J. Cell. Sci. *87 (Pt 2)*, 363–372.

Ting, L., Rad, R., Gygi, S.P., and Haas, W. (2011). MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nature Methods *8*, 937–940.

Tu, X., and Wang, C.C. (2004). The involvement of two cdc2-related kinases (CRKs) in Trypanosoma brucei cell cycle regulation and the distinctive stage-specific phenotypes caused by CRK3 depletion. J. Biol. Chem. *279*, 20519–20528.

Tu, X., and Wang, C.C. (2005). Pairwise knockdowns of cdc2-related kinases (CRKs) in Trypanosoma brucei identified the CRKs for G1/S and G2/M transitions and demonstrated distinctive cytokinetic regulations between two developmental stages of the organism. Eukaryotic Cell *4*, 755–764.

Tu, X., Kumar, P., Li, Z., and Wang, C.C. (2006). An aurora kinase homologue is involved in regulating both mitosis and cytokinesis in Trypanosoma brucei. J. Biol. Chem. *281*, 9677–9687.

Ullu, E., Matthews, K.R., and Tschudi, C. (1993). Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. Mol. Cell. Biol. *13*, 720–725.

Urbaniak, M.D., Guther, M.L.S., and Ferguson, M.A.J. (2012a). Comparative SILAC proteomic analysis of Trypanosoma brucei bloodstream and procyclic lifecycle stages. PLoS ONE *7*, e36619.

Urbaniak, M.D., Martin, D.M.A., and Ferguson, M.A.J. (2013). Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of Trypanosoma brucei. J. Proteome Res. *12*, 2233–2244.

Urbaniak, M.D., Mathieson, T., Bantscheff, M., Eberhard, D., Grimaldi, R., Miranda-Saavedra, D., Wyatt, P., Ferguson, M.A.J., Frearson, J., and Drewes, G. (2012b). Chemical proteomic analysis reveals the drugability of the kinome of Trypanosoma brucei. ACS Chem. Biol. *7*, 1858–1865.

Valente, M., Timm, J., Castillo-Acosta, V.M., Ruiz-Pérez, L.M., Balzarini, T., Nettleship, J.E., Bird, L.E., Rada, H., Wilson, K.S., and González-Pacanowska, D. (2016). Cell cycle regulation and novel structural features of thymidine kinase, an essential enzyme in Trypanosoma brucei. Molecular Microbiology.

Van Den Abbeele, J., Claes, Y., van Bockstaele, D., Le Ray, D., and Coosemans, M. (1999). Trypanosoma brucei spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. Parasitology *118 (Pt 5)*, 469–478.

Vasquez, J.-J., Hon, C.-C., Vanselow, J.T., Schlosser, A., and Siegel, T.N. (2014). Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. Nucleic Acids Research *42*, 3623–3637.

Vassella, E., Reuner, B., Yutzy, B., and Boshart, M. (1997). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle

arrest via the cAMP pathway. J. Cell. Sci. *110 (Pt 21)*, 2661–2671.

Vertommen, D., Van Roy, J., Szikora, J.-P., Rider, M.H., Michels, P.A.M., and Opperdoes, F.R. (2008). Differential expression of glycosomal and mitochondrial proteins in the two major life-cycle stages of Trypanosoma brucei. Molecular and Biochemical Parasitology *158*, 189–201.

Vickerman, K. (1965). Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. Nature *208*, 762–766.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. Br. Med. Bull. *41*, 105–114.

Vickerman, K., Tetley, L., Hendry, K.A., and Turner, C.M. (1988). Biology of African trypanosomes in the tsetse fly. Biol. Cell *64*, 109–119.

Wang, C.C., Bozdech, Z., Liu, C.-L., Shipway, A., Backes, B.J., Harris, J.L., and Bogyo, M. (2003). Biochemical analysis of the 20 S proteasome of Trypanosoma brucei. J. Biol. Chem. *278*, 15800–15808.

Wheeler, R.J., Scheumann, N., Wickstead, B., Gull, K., and Vaughan, S. (2013). Cytokinesis in Trypanosoma brucei differs between bloodstream and tsetse trypomastigote forms: implications for microtubule-based morphogenesis and mutant analysis. Molecular Microbiology *90*, 1339–1355.

WHO (1998). Control and surveillance of African trypanosomiasis. Report of a WHO Expert Committee. World Health Organ Tech Rep Ser *881*, I–VI–1–114.

Wickstead, B., Carrington, J.T., Gluenz, E., and Gull, K. (2010). The expanded Kinesin-13 repertoire of trypanosomes contains only one mitotic Kinesin indicating multiple extra-nuclear roles. PLoS ONE *5*, e15020.

Wittmann, T., Hyman, A., and Desai, A. (2001). The spindle: a dynamic assembly of microtubules and motors. Nature Cell Biology *3*, E28–E34.

Zamudio, J.R., Mittra, B., Chattopadhyay, A., Wohlschlegel, J.A., Sturm, N.R., and Campbell, D.A. (2009). Trypanosoma brucei spliced leader RNA maturation by the cap 1 2'-O-ribose methyltransferase and SLA1 H/ACA snoRNA pseudouridine synthase complex. Mol. Cell. Biol. *29*, 1202–1211.

Zhou, Q., Gheiratmand, L., Chen, Y., Lim, T.K., Zhang, J., Li, S., Xia, N., Liu, B., Lin, Q., and He, C.Y. (2010). A comparative proteomic analysis reveals a new bi-lobe protein required for bi-lobe duplication and cell division in Trypanosoma brucei. PLoS ONE *5*, e9660.

Zikova, A., Panigrahi, A.K., Dalley, R.A., Acestor, N., Anupama, A., Ogata, Y., Myler, P.J., and Stuart, K. (2008). Trypanosoma brucei mitochondrial ribosomes: affinity purification and component identification by mass spectrometry. Molecular & Cellular Proteomics *7*, 1286–1296.

Zikova, A., Schnaufer, A., Dalley, R.A., Panigrahi, A.K., and Stuart, K.D. (2009). The F(0)F(1)-ATP synthase complex contains novel subunits and is essential for procyclic

Trypanosoma brucei. PLoS Pathog. *5*, e1000436.

Zoltner, M., Leung, K.F., Alsford, S., Horn, D., and Field, M.C. (2015). Modulation of the Surface Proteome through Multiple Ubiquitylation Pathways in African Trypanosomes. PLoS Pathog. *11*, e1005236.

Supplementary Tables

Supplementary Table 1: List of gold standard protein complexes used for random forest machine learning prediction of protein interactions. A representative protein complex name is shown along with the gene IDs of component proteins.

Protein Complex	Gene IDs
Acidic Ribosomal	Tb927.10.3380; Tb927.9.5690
ALBA	Tb927.4.2040; Tb927.11.4460; Tb927.4.2030; Tb927.11.4450
Arginine N-methyltransferase	Tb927.1.4690; Tb927.10.3560
ARP2/3	Tb927.10.4540; Tb927.9.5350; Tb927.10.15800; Tb927.2.2900; Tb927.8.4410
ATP synthase	Tb927.3.1380; Tb927.10.180; Tb927.5.1710; Tb927.6.4990; Tb927.10.5050; Tb927.7.7430
Dihydrolipoamide dehydrogenase	Tb927.8.7380; Tb927.3.4390; Tb927.4.5040; Tb927.5.3800
DNA polymerase	Tb927.11.8890; Tb11.v5.0480
Exosome	Tb927.10.7450; Tb927.11.11030; Tb927.11.16600; Tb927.4.1630; Tb927.5.1200; Tb927.6.670; Tb927.9.7070
Fibrillarin	Tb927.10.14750; Tb927.10.7500
Nucleosome	Tb927.10.10590; Tb927.5.4260; Tb927.7.2940
Mitochondrial peptidase	Tb927.11.3980; Tb927.5.1060
Mitochondrial RNABP	Tb927.11.13280; Tb927.11.1710
Mitochondrial SSU	Tb927.11.2530; Tb927.10.3580; Tb927.7.3050
MRB1 complex	Tb927.2.3800; Tb927.7.2570
MVP	Tb927.5.4460; Tb927.10.1990; Tb927.10.6310
N-acetyltransferase	Tb927.10.5670; Tb927.11.4530; Tb927.10.3150
Nucleosome assembly	Tb927.10.15180; Tb927.9.5730
Oxoglutarate dehydrogenase	Tb927.11.1450; Tb927.11.9980; Tb927.11.11680
Oxoisovalerate dehydrogenase	Tb927.10.4330; Tb927.10.660
Phosphoribosylpyrophosphate synthetase	Tb927.5.2960; Tb927.11.3030; Tb927.10.9430; Tb927.5.3170
Prefoldin	Tb927.7.570; Tb927.7.2590; Tb927.11.12680; Tb927.5.580; Tb927.6.2280; Tb927.11.800; Tb927.11.16040; Tb927.11.12910

Proteasome core	Tb927.10.290; Tb927.7.4790; Tb927.9.11310; Tb927.4.430; Tb927.6.1260; Tb927.7.4420; Tb927.9.9670; Tb927.11.7020; Tb927.3.780; Tb927.10.4710; Tb927.10.230; Tb11.v5.0196; Tb927.10.6080; Tb927.11.7270
Proteasome regulatory	Tb927.11.9220; Tb927.10.3520; Tb927.10.15720; Tb927.10.9740; Tb927.11.14430; Tb927.10.1550; Tb927.11.16030; Tb927.7.2550; Tb927.6.1090
Pyruvate dehydrogenase	Tb927.10.12700; Tb927.3.1790; Tb927.10.7570
Ribosome 60S	Tb927.9.3990; Tb927.8.1340; Tb927.3.3320; Tb927.7.1730
ruvB-like DNA helicase	Tb927.4.1270; Tb927.4.200
SMD	Tb927.2.4540; Tb927.2.5850; Tb927.4.890
Spliceosome	Tb927.5.2290; Tb927.11.15430; Tb927.9.11110
Succinyl-CoA	Tb927.3.2230; Tb927.10.2560
T-Complex	Tb927.10.1060; Tb927.10.8190; Tb927.11.16760; Tb927.11.3240; Tb927.11.14250; Tb927.11.1900; Tb927.9.11270; Tb927.8.3150
Translation elongation	Tb927.10.5840; Tb927.4.3590; Tb927.11.13190
tRNA processing complex	Tb927.9.5210; Tb927.8.5330; Tb927.10.1250; Tb927.10.10030

Supplementary Table 2: List of predicted protein complexes from machine learning analysis. Protein complex numbers match those displayed on 'Complex Explorer' web application described in Chapter 3.

Predicted Protein Complex	Protein Id	Protein Description
1	Tb927.10.2290	chaperone protein DNAj, putative
	Tb927.11.7380	glycerol-3-phosphate dehydrogenase (FAD-dependent), mitochondrial
	Tb927.8.6080	Glycerophosphoryl diester phosphodiesterase family, putative (POMP42)
	Tb927.9.15000	proteasome complex subunit Rpn13 ubiquitin receptor, putative
2	Tb927.4.1330	DNA topoisomerase IB, large subunit
	Tb927.9.10530	hypothetical protein, conserved
3	Tb927.10.3700	AMP-activated protein kinase, gamma regulatory subunit, SNF1-related protein kinase regulatory subunit gamma, AMPK subunit gamma (AMPKG)
	Tb927.10.5310	SNF1-related protein kinases, putative
	Tb927.3.4560	5'-AMP-activated protein kinase catalytic subunit alpha, putative, AMPK subunit alpha, putative, SNF1-related protein kinase catalytic subunit alpha, putative
	Tb927.8.2450	SNF1-related protein kinase regulatory subunit beta, 5'-AMP-activated protein kinase subunit beta, AMPK subunit beta (AMPKB)
	Tb927.9.9270	hypothetical protein, conserved

4	Tb927.11.12930	DEAD-box helicase, putative
	Tb927.11.5990	hypothetical protein, conserved
	Tb927.4.1070	50S ribosomal protein L13, putative
	Tb927.6.4080	hypothetical protein, conserved
	Tb927.7.1640	ras-like small GTPase, putative (TbEAR)
	Tb927.7.3460	hypothetical protein, conserved
5	Tb927.11.11290	heat shock protein 70, putative
	Tb927.11.7150	NGG1 interacting factor 3-like
	Tb927.5.3520	queuine tRNA-ribosyltransferase, putative
	Tb927.6.4920	S-adenosylmethionine synthetase, putative (METK1)
	Tb927.9.5190	proliferative cell nuclear antigen (PCNA), putative
	Tb927.9.9820	glyceraldehyde-3-phosphate dehydrogenase, putative
6	Tb11.1390	hypothetical protein, conserved
	Tb927.1.4050	protein phosphatase with EF-Hand domains (PPEF), ser/thr protein phosphatase, putative
	Tb927.10.12500	P-type H+-ATPase, putative
	Tb927.10.12510	P-type H+-ATPase, putative
	Tb927.10.7700	ABC transporter, putative
	Tb927.10.8530	glucose transporter 2A (THT2A)
	Tb927.10.9080	pteridine transporter, putative
	Tb927.11.540	ABC transporter, mitochondrial, putative, multidrug resistance protein, mitochondrial, putative (ABCT)
	Tb927.11.5970	phosphoinositide-specific phospholipase C, putative
	Tb927.11.6040	Nodulin-like, putative
	Tb927.4.4360	monoglyceride lipase, putative
	Tb927.4.4490	multidrug resistance protein E, p-glycoprotein (MRPE)
	Tb927.8.2160	multidrug resistance protein A, p-glycoprotein (PGPA)
	Tb927.9.15460	calcium motive p-type ATPase, putative
	Tb927.9.6310	ABC transporter, putative
7	Tb11.v5.0807	ribonuclease, putative
	Tb927.10.5620	fructose-bisphosphate aldolase, glycosomal (ALD)
	Tb927.11.6830	Domain of unknown function(DUF2779), putative
	Tb927.11.7310	RNA binding protein, putative
	Tb927.2.820	retrotransposon hot spot protein (RHS, pseudogene), putative, retrotransposon hot spot protein 1 (RHS1), interrupted
	Tb927.9.9580	tubulin tyrosine ligase protein, putative
8	Tb927.10.11530	hypothetical protein, conserved
	Tb927.11.10300	hypothetical protein, conserved
	Tb927.11.5510	dynein light chain p28, axonemal, putative
	Tb927.3.5580	tryptophanyl-tRNA synthetase
	Tb927.7.4570	inosine-guanine nucleoside hydrolase (IG-NH)
	Tb927.9.12290	Peroxin 19
9	Tb927.1.870	deoxyhypusine synthase, putative
	Tb927.10.2750	deoxyhypusine synthase, putative

10	Tb927.11.10240	hslVU complex proteolytic subunit, threonine peptidase, Clan T(1), family T1B, ATP-dependent protease subunit HslV (HsIV)
	Tb927.11.2230	carnitine O-acetyltransferase, putative (CAT)
	Tb927.4.1270	ruvB-like DNA helicase, putative
	Tb927.4.2000	ruvB-like DNA helicase, putative (RUVBL)
	Tb927.6.3610	SET domain containing protein, putative
	Tb927.6.950	cysteinyl-tRNA synthetase, putative
	Tb927.7.5890	hypothetical protein, conserved
	Tb927.9.5900	glutamate dehydrogenase (GDH)
11	Tb927.1.1670	ARM-like helical domain-containing protein
	Tb927.4.1890	hypothetical protein, conserved
	Tb927.5.1090	threonyl-tRNA synthetase, putative
	Tb927.7.760	hypothetical protein, conserved
12	Tb927.10.13670	serine/threonine protein phosphatase 5
	Tb927.10.14030	hypothetical protein, conserved
	Tb927.3.3580	heat shock protein 90, putative (LPG3)
	Tb927.5.3260	WD domain, G-beta repeat, putative
	Tb927.9.9860	Hsp70 protein, putative
13	Tb927.5.2250	META domain/Domain of unknown function (DUF1935), putative
	Tb927.6.4950	mago nashi-like protein, putative
	Tb927.7.1170	RNA-binding protein, putative (Y14)
14	Tb927.11.15370	hypothetical protein, conserved (TbKap123)
	Tb927.3.1670	hypothetical protein, conserved
	Tb927.3.4250	hypothetical protein, conserved
15	Tb927.10.10030	hypothetical protein, conserved
	Tb927.10.1250	Aminoacyl-tRNA editing domain containing protein, putative
	Tb927.5.4450	hypothetical protein, conserved
	Tb927.8.5330	tyrosyl/methionyl-tRNA synthetase, putative
	Tb927.9.5210	glutaminyl-tRNA synthetase, putative
16	Tb927.10.12700	pyruvate dehydrogenase E1 alpha subunit, putative
	Tb927.10.7570	dihydrolipoamide acetyltransferase E2 subunit, putative
	Tb927.3.1790	pyruvate dehydrogenase E1 beta subunit, putative
	Tb927.5.1090	threonyl-tRNA synthetase, putative
17	Tb927.1.1670	ARM-like helical domain-containing protein
	Tb927.10.15180	nucleosome assembly protein, putative
	Tb927.11.6370	leucine-rich repeat protein (LRRP), putative
	Tb927.9.5730	nucleosome assembly protein-like protein
18	Tb927.10.2620	CS domain containing protein, putative
	Tb927.2.1890	E2-like ubiquitin-conjugation enzyme (ATG3)
19	Tb927.10.15170	hypothetical protein, conserved
	Tb927.10.1990	major vault protein, putative (MVP)
	Tb927.10.6310	major vault protein, putative (MVP)
	Tb927.5.4460	major vault protein, putative (MVP)

20	Tb927.8.5900	Sedlin, N-terminal conserved region, putative
	Tb927.9.12150	transport protein particle (TRAPP) subunit, putative
21	Tb927.10.3710	proteasome activator protein PA26 (pa26)
	Tb927.3.4040	Ankyrin repeats (3 copies), putative
22	Tb927.10.11390	60S ribosomal protein L6, putative
	Tb927.3.3320	60S ribosomal protein L13, putative
	Tb927.4.2180	60S ribosomal protein L35a, putative
	Tb927.6.720	40S ribosomal protein L14, putative
	Tb927.7.1730	60S ribosomal protein L7, putative
	Tb927.8.1340	60S ribosomal protein L7a, putative
	Tb927.9.3990	ribosomal protein S7, putative
23	Tb927.11.9710	60S ribosomal protein L10a, putative (RPL10A)
	Tb927.5.2770	SET domain containing protein, putative
24	Tb927.10.10960	heat shock protein, putative
	Tb927.8.3100	coronin, putative (CRN12)
25	Tb927.10.1100	60S ribosomal protein L9, putative
	Tb927.11.14130	ribosomal protein L18, putative
	Tb927.11.3600	40S ribosomal protein S4, putative
	Tb927.11.9730	60S ribosomal protein L34, putative
	Tb927.4.1100	ribosomal protein L21E (60S), putative
	Tb927.4.1860	ribosomal protein S19, putative
	Tb927.5.1610	60S ribosomal protein L13a, putative
	Tb927.7.1050	40S ribosomal protein S16, putative
	Tb927.7.5020	60S ribosomal protein L19, putative
	Tb927.8.6160	40S ribosomal protein S8, putative
	Tb927.9.11410	60S ribosomal protein L23, putative
	Tb927.9.11490	60S ribosomal protein L27a, 60S ribosomal protein L28, 60S ribosomal protein L29 (RPL27A)
	Tb927.9.8420	60S ribosomal protein L10, putative, QM-like protein
26	Tb927.10.3760	vacuolar ATP synthase subunit d, putative
	Tb927.10.730	ATP synthase, putative
	Tb927.11.11690	Vacuolar proton pump subunit B, putative, V-type proton ATPase subunit B, putative
	Tb927.11.9420	ATP synthase, putative
27	Tb927.10.9020	Gcd10p family, putative
	Tb927.11.11660	conserved protein
	Tb927.11.9210	NOL1/NOP2/sun family, putative
	Tb927.2.380	retrotransposon hot spot protein 2 (RHS2), putative
	Tb927.2.830	retrotransposon hot spot protein (RHS, pseudogene), putative, retrotransposon hot spot protein 1 (RHS1), interrupted
28	Tb927.1.2580	RNasePH-like protein, exosome-associated protein 1 (EAP1)
	Tb927.10.5840	translation elongation factor 1-beta, putative
	Tb927.10.7450	exosome complex exonuclease RRP41A, Ribosomal RNA processing protein 41A (RRP41A)

	Tb927.11.16600	exosome-associated protein 2 (EAP2)
	Tb927.11.630	RNA polymerase I second largest subunit (RPA135)
	Tb927.3.1150	Conserved hypothetical ATP binding protein, putative
	Tb927.3.1300	hypothetical protein, conserved
	Tb927.5.1200	exosome component CSL4 (CSL4)
	Tb927.6.670	ribosomal RNA processing protein 45, exosome complex exonuclease (RRP45)
	Tb927.7.4670	ribosomal RNA processing protein 4, exosome complex exonuclease (RRP4)
	Tb927.7.710	heat shock 70 kDa protein, putative (HSP70)
	Tb927.9.7070	exosome complex exonuclease RRP40 (RRP40)
29	Tb927.11.12680	prefoldin subunit 2, putative
	Tb927.5.580	prefoldin subunit, putative
	Tb927.7.2590	prefoldin, putative
	Tb927.7.570	prefoldin, putative
30	Tb927.10.1550	proteasome regulatory non-ATP-ase subunit 5 (RPN5)
	Tb927.10.15720	proteasome regulatory non-ATP-ase subunit 9 (RPN9)
	Tb927.10.3030	proteasome regulatory non-ATPase subunit 11 (RPN11)
	Tb927.10.3520	protease regulatory ATPase subunit 4 (RPT4)
	Tb927.10.9740	Regulatory particle triple-A ATPase subunit 6, 19S proteasome regulatory subunit (RPT6)
	Tb927.11.14430	proteasome regulatory non-ATP-ase subunit
	Tb927.11.16030	proteasome regulatory non-ATP-ase subunit 7 (RPN7)
	Tb927.11.3740	proteasome regulatory ATPase subunit 2 (RPT2)
	Tb927.11.8310	class I transcription factor A, subunit 4 (CITFA-4)
	Tb927.2.2440	proteasome regulatory non-ATPase subunit 6 (RPN6)
	Tb927.6.1090	proteasome regulatory ATPase subunit 3 (RPT3)
	Tb927.7.2500	proteasome regulatory ATPase subunit 1
	Tb927.7.2550	proteasome regulatory ATPase subunit 5 (RPT5)
	Tb927.8.570	proteasome regulatory non-ATP-ase subunit 10
31	Tb11.v5.0196	Proteasome subunit A N-terminal signature/Proteasome subunit, putative
	Tb927.10.230	proteasome subunit alpha type-5, putative
	Tb927.10.290	proteasome alpha 2 subunit, putative
	Tb927.10.4710	20S proteasome subunit, proteasome subunit beta type-2, putative (PSB4)
	Tb927.10.6080	proteasome subunit beta type-5, putative, proteasome subunit beta type-5, putative
	Tb927.11.7020	proteasome alpha 7 subunit, putative (PSA4)
	Tb927.11.7270	proteasome beta 3 subunit, putative (PSB3)
	Tb927.3.780	proteasome alpha 7 subunit (TbPSA7)
	Tb927.4.430	proteasome beta 7 subunit
	Tb927.6.1260	proteasome beta-1 subunit, putative (PSB1)
	Tb927.7.4420	proteasome alpha 3 subunit, putative
	Tb927.7.4790	proteasome beta 6 subunit, 20S proteasome beta 6 subunit, putative (BETA6)

	Tb927.9.11310	unspecified product
	Tb927.9.9670	proteasome alpha 1 subunit, putative, 20S proteasome subunit alpha-6, (putative) (TbPSA6)
32	Tb927.10.3280	60S ribosomal proteins L38, putative
	Tb927.2.6090	60S ribosomal protein L44 (RPL44)
33	Tb927.10.3380	60S acidic ribosomal protein P2, putative
	Tb927.9.5690	60S acidic ribosomal protein, putative
34	Tb927.10.2200	hypothetical protein, conserved
	Tb927.10.4220	hypothetical protein, conserved
35	Tb11.02.5380b	exosome complex exonuclease RRP44p homologue, putative
	Tb927.10.7630	transportin2- like protein
36	Tb927.10.9250	adenylyl cyclase-associated protein, putative
	Tb927.7.880	RNA-binding protein, putative (RPB25)
37	Tb927.11.14190	Staphylococcal nuclease homologue/Tudor domain containing protein, putative
	Tb927.11.2650	heat shock protein 84, putative
38	Tb927.11.11680	2-oxoglutarate dehydrogenase E2 component, putative
	Tb927.11.1450	2-oxoglutarate dehydrogenase E1 component, putative
	Tb927.11.9980	2-oxoglutarate dehydrogenase E1 component, putative
39	Tb927.11.3660	Dynein light chain Tctex-type, putative
	Tb927.7.4820	Trm112p-like protein, putative
40	Tb927.11.3980	mitochondrial processing peptidase alpha subunit, putative, metallo-peptidase, Clan ME, Family M16
	Tb927.5.1060	mitochondrial processing peptidase, beta subunit, putative, metallo-peptidase, Clan ME, Family M16
41	Tb927.11.4360	Protein of unknown function (DUF1014), putative
	Tb927.3.1920	NOT5 protein (NOT5)
42	Tb927.11.4480	radial spoke protein RSP4/6, putative
	Tb927.3.2890	radial spoke protein RSP10, putative
43	Tb927.11.4920	hypothetical protein, conserved
	Tb927.5.1900	hypothetical protein, conserved
44	Tb927.11.9610	eukaryotic translation initiation factor 3 subunit 2, putative (eIF-3 beta)
	Tb927.5.2570	translation initiation factor, putative (EIF3B)
45	Tb927.2.2230	hypothetical protein, conserved
	Tb927.2.2390	hypothetical protein, conserved
46	Tb927.10.1890	cysteine peptidase, Clan CA, family C2, putative
	Tb927.2.5810	Holliday-junction resolvase-like of SPT6/SH2 domain containing protein, putative
47	Tb927.3.1210	protein transport protein Sec24 (SEC24.1)
	Tb927.3.3890	hypothetical protein, conserved
48	Tb927.3.2660	TatD related DNase, putative
	Tb927.9.8200	Pescadillo N-terminus/BRCA1 C Terminus (BRCT) domain containing protein, putative
49	Tb927.3.3630	Elongation factor Ts, mitochondrial, putative (EF-Ts)
	Tb927.7.1340	10 kDa heat shock protein, putative (HSP10)

50	Tb927.3.5310	paraflagellar rod protein
	Tb927.5.940	NADH-dependent fumarate reductase, putative
51	Tb927.3.5370	hypothetical protein, conserved
	Tb927.6.1870	eukaryotic translation initiation factor 4e, putative
52	Tb927.11.3850	AMP deaminase, putative
	Tb927.4.1680	ZFP family member, putative (ZC3H10)
53	Tb927.4.2630	ATP-dependent DEAD/H RNA helicase, putative
	Tb927.4.4160	mitochondrial RNA binding protein (MRB4160)
54	Tb927.6.1200	hypothetical protein, conserved
	Tb927.7.930	zinc finger CCCH domain containing protein 17 (ZC3H17)
55	Tb927.5.3900	Galactose oxidase, central domain/Domain of unknown function (DUF4110), putative
	Tb927.6.1990	hypothetical protein, conserved
56	Tb927.11.4160	predicted C2 domain protein
	Tb927.6.3310	calpain-like cysteine peptidase, putative, cysteine peptidase, Clan CA, family C2, putative
57	Tb927.6.2170	co-chaperone GrpE, putative
	Tb927.6.4000	small glutamine-rich tetratricopeptide repeat protein, putative, (SGT)
58	Tb927.10.600	hypothetical protein, conserved
	Tb927.6.4200	hypothetical protein, conserved
59	Tb927.11.2640	ras-like small GTPase, putative (TbNST)
	Tb927.6.4750	hypothetical protein, conserved
60	Tb927.7.1360	hypothetical protein, conserved
	Tb927.7.7140	Vta1 like, putative
61	Tb927.11.10330	Regulator of chromosome condensation (RCC1) repeat, putative
	Tb927.7.970	NMD3 family, putative
62	Tb927.8.1420	acyl-CoA dehydrogenase, mitochondrial precursor, putative
	Tb927.8.1740	hypothetical protein, conserved
63	Tb927.11.10760	kinesin-like protein, putative
	Tb927.8.2630	kinesin, putative
64	Tb927.8.4400	hypothetical protein, conserved
	Tb927.9.8820	hypothetical protein, conserved
65	Tb927.8.1090	NPAPL (NPAPL)
	Tb927.8.4870	DIGIT
66	Tb927.10.15410	glycosomal malate dehydrogenase (gMDH)
	Tb927.9.11940	replication factor A protein 3, putative
67	Tb927.7.4500	PX domain containing protein, putative
	Tb927.9.13380	phosphoinositide-binding protein, putative
68	Tb927.9.5040	cAMP-specific phosphodiesterase (PDEB1)
	Tb927.9.5100	cAMP-specific phosphodiesterase (PDEB2)
69	Tb927.11.1840	hypothetical protein, conserved
	Tb927.6.5070	hypothetical protein, conserved
70	Tb927.10.8940	flagellum targeting protein kharon1, putative (KH1)

	Tb927.3.3750	paraflagellar rod component, putative (PFC7)
71	Tb927.10.4330	2-oxoisovalerate dehydrogenase beta subunit, mitochondrial precursor, putative
	Tb927.10.660	2-oxoisovalerate dehydrogenase alpha subunit, putative
72	Tb927.11.6740	pumilio/PUF RNA binding protein 10, putative
	Tb927.7.2170	hypothetical protein, conserved
73	Tb927.1.1000	developmentally regulated phosphoprotein
	Tb927.11.4780	pyruvate dehydrogenase (lipoamide) kinase, putative
74	Tb927.10.15520	signal recognition particle protein, putative
	Tb927.4.1850	hypothetical protein, conserved
	Tb927.5.3800	glutamine hydrolysing (not ammonia-dependent) carbomoyl phosphate synthase, putative
75	Tb927.10.14150	nuclear segregation protein, putative
	Tb927.11.6440	hypothetical protein, conserved
	Tb927.6.4770	protein mkt1, putative (MKT1)
76	Tb927.10.170	pseudouridine synthase, Cbf5p
	Tb927.4.470	snoRNP protein GAR1, putative
	Tb927.4.750	50S ribosomal protein L7Ae, putative
77	Tb927.10.2720	hypothetical protein, conserved
	Tb927.10.6680	member of the NOL1/NOP2/sun family of proteins
	Tb927.10.9050	pseudouridine synthase TruD, putative, tRNA pseudouridine synthase TruD, putative (pus7)
78	Tb927.11.13740	receptor-type adenylate cyclase GRESAG 4, putative
	Tb927.11.17040	expression site-associated gene 4 (ESAG4) protein, putative, receptor-type adenylate cyclase, putative
	Tb927.7.7470	receptor-type adenylate cyclase GRESAG 4, putative
79	Tb927.10.3150	N-acetyltransferase, putative
	Tb927.10.5670	N-acetyltransferase subunit Nat1, putative (NAT1)
	Tb927.11.4530	N-acetyltransferase subunit ARD1 (ARD1)
80	Tb927.11.11330	heat shock protein 70
	Tb927.11.7510	glucose-regulated protein 78, putative, luminal binding protein 1 (BiP), putative (BiP)
	Tb927.6.3800	heat shock 70 kDa protein, mitochondrial precursor, putative
81	Tb927.2.4540	Small nuclear ribonucleoprotein-associated protein B (snRNP-B) (Sm protein B) (Sm-B) (SmB), putative (TbSmB)
	Tb927.2.5850	small nuclear ribonucleoprotein SmD2 (Sm-D2)
	Tb927.4.890	small nuclear ribonucleoprotein SmD3, putative (SmD3)
82	Tb927.3.2280	vacuolar sorting protein 33 , putative
	Tb927.8.3370	Ran-binding protein, putative
	Tb927.8.6270	hypothetical protein, conserved
83	Tb927.11.5840	Protein translation factor SUI1 homolog, putative
	Tb927.3.3300	hypothetical protein, conserved
	Tb927.8.6240	STOP axonemal protein
84	Tb927.11.4460	ALBA-Domain Protein (ALBA1)
	Tb927.4.2030	ALBA-Domain Protein (ALBA4)

	Tb927.4.2040	ALBA-Domain Protein (ALBA3)
85	Tb927.10.8030	hypothetical protein, conserved
	Tb927.11.6250	hypothetical protein, conserved
	Tb927.5.1780	hypothetical protein, conserved
86	Tb927.11.13960	U6 snRNA-associated Sm-like protein LSm4p (TbLSm4)
	Tb927.5.4030	U6 snRNA-associated Sm-like protein LSm7p (TbLSm7)
	Tb927.8.2850	Poly(A)-specific ribonuclease PARN-1
87	Tb927.10.10590	histone H2B, putative
	Tb927.5.4260	histone H4, putative
	Tb927.7.2940	histone H2A, putative
88	Tb927.10.10010	60S acidic ribosomal protein, putative
	Tb927.10.1900	DNA topoisomerase IA, putative
	Tb927.5.4420	nucleolar RNA helicase II, putative, nucleolar RNA helicase Gu, putative
89	Tb927.7.6280	Domain of unknown function (DUF3508), putative
	Tb927.8.1560	hypothetical protein, conserved
90	Tb11.v5.0394	hypothetical protein, conserved
	Tb927.6.2010	AMP-binding enzyme, putative
	Tb927.6.740	ATP-dependent DEAH-box RNA helicase, putative
91	Tb927.3.1010	hypothetical protein, conserved
	Tb927.7.2640	hypothetical protein, conserved
	Tb927.9.10400	hypothetical protein, conserved
92	Tb927.10.3580	hypothetical protein, conserved
	Tb927.11.2530	Mitochondrial SSU ribosomal protein, putative, mitochondrial RNA binding complex 1 subunit
	Tb927.7.3050	hypothetical protein, conserved
93	Tb927.11.1680	vesicular-fusion protein SEC18, putative
	Tb927.7.1100	hypothetical protein, conserved
	Tb927.7.3810	'Cold-shock' DNA-binding domain containing protein, putative
94	Tb927.2.4230	NUP-1 protein, putative
	Tb927.7.3330	hypothetical protein, conserved
	Tb927.8.3870	SRP40, C-terminal domain containing protein, putative
95	Tb927.11.3830	hypothetical protein, conserved
	Tb927.8.5090	DNA-directed RNA polymerase I largest subunit (RPA190)
	Tb927.9.2120	hypothetical protein, conserved
96	Tb927.10.7230	hypothetical protein, conserved
	Tb927.3.3520	hypothetical protein, conserved (POMP25)
	Tb927.9.3400	endo-beta-N-acetylglucosaminidase, putative
97	Tb927.11.16200	cytoskeleton-associated protein 17, corset-associated protein 17 (CAP17)
	Tb927.8.6440	RNA-binding protein, putative (RPB20)
	Tb927.9.9060	Lsm12 protein, putative
98	Tb927.3.1600	Tim10/DDP family zinc finger, putative
	Tb927.9.1350	hypothetical protein, conserved

99	Tb927.10.5840	translation elongation factor 1-beta, putative
	Tb927.10.7940	methyltransferase, putative
	Tb927.11.13190	elongation factor 1 gamma, putative
	Tb927.11.16490	hypothetical protein, conserved
	Tb927.4.3590	translation elongation factor 1-beta, putative
100	Tb927.11.12670	Methyltransferase TYW3, putative
	Tb927.11.7080	acidocalcisomal pyrophosphatase
	Tb927.8.2050	GDP-mannose pyrophosphorylase
	Tb927.9.1880	WD domain, G-beta repeat, putative
101	Tb927.1.1030	leucine-rich repeat protein (LRRP), putative
	Tb927.10.2570	lysosomal alpha-mannosidase precursor, putative
	Tb927.11.6670	hypothetical protein, conserved
	Tb927.4.1520	Pestivirus Npro endopeptidase C53, putative
102	Tb927.11.1050	Ribosome production factor 1, putative
	Tb927.11.3120	nucleolar GTP-binding protein 1 (NOG1)
	Tb927.11.6790	predicted WD40 repeat protein
	Tb927.6.3790	valosin-containing protein homolog, putative, AAA ATPase
103	Tb927.10.11170	hypothetical protein, conserved
	Tb927.11.10840	hypothetical protein, conserved
	Tb927.11.7770	oxidoreductase-like protein
	Tb927.9.12780	Leucine carboxyl methyltransferase/Cupin-like domain containing protein, putative
104	Tb927.5.1840	hypothetical protein, conserved
	Tb927.8.6900	transport protein particle (TRAPP) component, putative
105	Tb927.10.6120	Peptidase M76 family, putative
	Tb927.3.2300	DNL zinc finger, putative
	Tb927.6.2990	Putative papain-like cysteine peptidase (DUF1796), putative
106	Tb927.4.360	1,2-Dihydroxy-3-keto-5-methylthiopentene dioxygenase, putative
	Tb927.6.3600	hypothetical protein, conserved
107	Tb927.6.4530	RNA-binding protein, putative (RBP17)
	Tb927.7.5760	nuclear transport factor 2 protein, putative, mRNA transport regulator MTR2, putative, nuclear transport factor, nuclear transport factor 2 protein, putative (MTR2)
	Tb927.9.9450	zinc finger protein family member, putative (ZC3H28)
108	Tb927.10.13270	Periodic tryptophan protein 2 homolog, putative
	Tb927.11.10480	PQQ-like domain/WD domain, G-beta repeat/Utp21 specific WD40 associated putative domain containing protein, putative
	Tb927.11.460	predicted WD40 repeat protein
	Tb927.7.4220	WD domain, G-beta repeat/Dip2/Utp12 Family, putative
109	Tb927.10.13720	RNA-binding protein 29, putative (RBP29)
	Tb927.11.10540	hypothetical protein, conserved
110	Tb927.3.2490	hypothetical protein, conserved
	Tb927.6.1770	kinesin, putative
	Tb927.9.15470	kinesin, putative

	Tb927.9.9730	hypothetical protein, conserved
111	Tb927.6.4300	glyceraldehyde 3-phosphate dehydrogenase, glycosomal (GAPDH)
	Tb927.8.2000	cyclophilin, putative (NCP1)
	Tb927.8.3530	glycerol-3-phosphate dehydrogenase [NAD+], glycosomal
	Tb927.9.8720	fructose-1,6-bisphosphatase (FBPase)
112	Tb927.10.15800	actin related protein 2, putative
	Tb927.10.4540	ARP2/3 complex subunit, putative
	Tb927.2.2900	ARP2/3 complex subunit, putative
	Tb927.8.4410	ARP2/3 complex subunit, putative
	Tb927.9.5350	actin related protein 3, putative
113	Tb927.11.6630	3-methylcrotonoyl-CoA carboxylase beta subunit, putative
	Tb927.3.4390	dihydrolipoamide dehydrogenase, putative (GCVL-1)
	Tb927.4.5040	dihydrolipoamide dehydrogenase, putative
	Tb927.8.6970	3-methylcrotonyl-CoA carboxylase alpha subunit, putative
	Tb927.8.7380	dihydrolipoamide dehydrogenase, point mutation, acetoin dehydrogenase e3 component, putative
114	Tb927.1.3180	40S ribosomal protein S11, putative
	Tb927.10.14930	Zinc finger CCCH domain-containing protein 39 (ZC3H39)
	Tb927.11.6300	40S ribosomal protein S5, putative
	Tb927.7.4910	hypothetical protein, conserved
115	Tb927.11.8030	hypothetical protein, conserved
	Tb927.3.3030	START domain containing protein, putative
	Tb927.5.780	hypothetical protein, conserved
	Tb927.7.3370	hypothetical protein, conserved
116	Tb927.5.2320	Mak10 subunit, NatC N(alpha)-terminal acetyltransferase, putative
	Tb927.7.2080	methyltransferase, putative, mRNA cap methyltransferase-like protein
	Tb927.7.2360	N-acetyltransferase, putative
	Tb927.8.960	hypothetical protein, conserved
	Tb927.9.4500	heat shock protein, putative, HSP70-like protein
117	Tb927.11.11080	Nucleoporin (TbNup149)
	Tb927.11.6170	protein transport protein SEC31, putative
	Tb927.11.7900	mitochondrial RNA binding protein 16 (RBP16)
	Tb927.7.320	hypothetical protein, conserved (TbRBP8)
	Tb927.9.3760	poly(A) export protein, putative (TbGLE2)
118	Tb927.11.8270	hypothetical protein, conserved
	Tb927.5.2530	Domain of unknown function (DUF1726)/Helicase/GNAT acetyltransferase 2/Possible tRNA binding domain containing protein, putative
	Tb927.8.1980	UTP15 C terminal, putative
119	Tb927.5.4270	ATP-dependent DEAD/H RNA helicase, putative
	Tb927.8.4820	eukaryotic translation initation factor 4 gamma, putative (eIF4G3)
120	Tb927.10.2770	eukaryotic translation initiation factor 5, putative
	Tb927.11.11010	hypothetical protein, conserved

121	Tb927.10.11210	hypothetical protein, conserved
	Tb927.11.12230	ATP-dependent protease ATPase subunit HslU2 (HslU2)
	Tb927.2.3800	MRB1-associated protein, guide RNA associated protein 1 (GAP1)
	Tb927.3.2900	eukaryotic initiation factor 2a, putative
	Tb927.7.2570	guide RNA associated protein, GAP2, mitochondrial RNA binding protein 1
122	Tb927.10.14700	hypothetical protein, conserved
	Tb927.3.3610	peroxisomal targeting signal 2 receptor, putative, Peroxin-7, putative, PTS2 receptor, putative (PEX7)
	Tb927.8.5780	phosphatase of regenerating liver-type phosphatase, putative
123	Tb927.10.5860	hypothetical protein, conserved
	Tb927.11.16220	hypothetical protein, conserved
	Tb927.4.1080	V-type ATPase, A subunit, putative
124	Tb927.11.8100	Ankyrin repeats (3 copies), putative
	Tb927.5.1280	alanine racemase, putative
125	Tb927.11.14490	RNA polymerase subunit, putative (RPB7)
	Tb927.8.6580	succinate dehydrogenase flavoprotein, putative
	Tb927.8.6920	hypothetical protein, conserved
126	Tb927.1.3050	tRNA (Uracil-5-)-methyltransferase, putative
	Tb927.10.3630	hypothetical protein, conserved
	Tb927.11.13280	mitochondrial RNA binding protein 2 (GBP25)
	Tb927.11.1550	XRN 5'-3' exonuclease N-terminus, putative
	Tb927.11.1710	mitochondrial RNA binding protein 1, guide RNA-binding protein of 21 kDa (gBP21)
	Tb927.3.1590	mitochondrial RNA binding complex 1 subunit (MRB1590)
	Tb927.3.3230	NOL1/NOP2/sun family, putative
127	Tb927.10.6220	5'-3' exoribonuclease D (XRND)
	Tb927.3.1040	unspecified product
	Tb927.5.2790	mitochondrial DNA polymerase beta-PAK (Pol beta-PAK)
	Tb927.5.4040	hypothetical protein, conserved
	Tb927.7.6800	Alpha/beta hydrolase family, putative
	Tb927.9.5590	DNA topoisomerase ii (TOP2)
128	Tb927.10.5770	Transitional endoplasmic reticulum ATPase, putative, Valosin-containing protein, Cell division control protein 48 (VCP)
	Tb927.10.9430	phosphoribosylpyrophosphate synthetase, putative (PRS)
	Tb927.11.3030	phosphoribosylpyrophosphate synthetase, putative (PRS)
	Tb927.5.2960	phosphoribosylpyrophosphate synthetase, putative (PRS)
	Tb927.5.3170	ribose-phosphate pyrophosphokinase, putative (PRPS5)
	Tb927.7.6910	Phosphatidylinositol 4-phosphate 5-kinase, putative, MORN repeat-containing protein
	Tb927.8.7100	acetyl-CoA carboxylase
129	Tb927.10.1060	T-complex protein 1, delta subunit, putative (TCP-1-delta)
	Tb927.10.8190	T-complex protein 1, theta subunit, putative, CCT-theta, putative
	Tb927.11.14250	T-complex protein 1, epsilon subunit, putative (TCP-1-epsilon)
	Tb927.11.16760	T-complex protein 1, alpha subunit, putative (TCP-1-alpha)

	Tb927.11.1900	T-complex protein 1, beta subunit, putative
	Tb927.11.3240	T-complex protein 1, zeta subunit, putative (TCP-1-zeta)
	Tb927.8.3150	T-complex protein 1, gamma subunit, putative (TCP-1-gamma)
	Tb927.9.11270	T-complex protein 1, eta subunit, putative, t- complex protein 1 (eta subunit), putative (TCP-1-eta)
130	Tb927.10.180	ATP synthase F1 subunit gamma protein, putative
	Tb927.10.5050	Mitochondrial ATP synthase epsilon chain, putative
	Tb927.11.13070	O-phosphoseryl-tRNA(Sec) selenium transferase, putative
	Tb927.3.1380	ATP synthase subunit beta, mitochondrial, ATP synthase F1, beta subunit (ATPB)
	Tb927.3.3410	aspartyl aminopeptidase, putative, metallo-peptidase, Clan MH, Family M20
	Tb927.5.1710	ribonucleoprotein p18, mitochondrial precursor, putative
	Tb927.6.4990	ATP synthase, epsilon chain, putative
	Tb927.7.7430	ATP synthase alpha chain, mitochondrial precursor, ATP synthase F1, alpha subunit
131	Tb927.3.1680	hypothetical protein, conserved
	Tb927.3.5470	hypothetical protein, conserved
	Tb927.4.3560	protein phoshatase 1, putative
	Tb927.8.4220	hypothetical protein, conserved
132	Tb927.10.5860	hypothetical protein, conserved
	Tb927.5.3970	adenylate kinase, putative (ADKE)
	Tb927.6.2790	L-threonine 3-dehydrogenase, putative
	Tb927.7.5160	deoxyuridine triphosphatase, putative, dUTP diphosphatase
133	Tb927.10.13950	tubulin-specific chaperone, putative
	Tb927.11.8770	ATP-dependent RNA helicase FAL1, putative
	Tb927.8.1990	peroxidoxin (TRYP2)
134	Tb927.1.1200	SSU ribosomal protein, mitochondrial (MRPS15)
	Tb927.10.7380	hypothetical protein, conserved
	Tb927.11.11630	hypothetical protein, conserved
	Tb927.11.1250	Mitochondrial SSU ribosomal protein, putative
	Tb927.11.6000	ribosomal protein L4/L1 family, putative
	Tb927.11.870	hypothetical protein, conserved
	Tb927.2.4890	ribosomal protein L11, putative
	Tb927.3.5610	ribosomal protein L3 mitochondrial, putative
	Tb927.5.3980	hypothetical protein, conserved
	Tb927.5.4120	hypothetical protein, conserved
	Tb927.6.2080	hypothetical protein, conserved
	Tb927.6.4560	hypothetical protein, conserved
	Tb927.7.2760	ribosomal protein L22p/L17e, putative
	Tb927.7.3030	hypothetical protein, conserved
	Tb927.7.4140	ribosomal protein L21, putative
	Tb927.8.5200	hypothetical protein, conserved
	Tb927.9.7170	Mitochondrial 39-S ribosomal protein L47 (MRP-L47), putative

	Tb927.9.8290	hypothetical protein, conserved
135	Tb927.11.7170	seryl-tRNA synthetase
	Tb927.5.4560	guanine deaminase, putative, guanase, putative, guanine aminase, putative, guanine aminohydrolase, putative
	Tb927.8.1020	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase, putative
	Tb927.9.7550	adenylosuccinate lyase, putative (ADSL)
136	Tb927.10.8880	hypothetical protein
	Tb927.9.12910	unspecified product
137	Tb927.10.2350	pyruvate dehydrogenase complex E3 binding protein, putative
	Tb927.9.7550	adenylosuccinate lyase, putative (ADSL)
138	Tb927.10.10140	paraflagellar rod component, putative (PFC19)
	Tb927.10.3710	proteasome activator protein PA26 (pa26)
	Tb927.10.8060	SET domain containing protein, putative
	Tb927.10.8360	Poly(A)-specific ribonuclease PARN-2
	Tb927.11.2630	diphthamide biosynthesis enzyme Dph1/Dph2 domain containing protein, putative
139	Tb927.10.1930	hypothetical protein, conserved
	Tb927.6.1830	WD domain, G-beta repeat, putative
	Tb927.6.2380	hypothetical protein, conserved
	Tb927.7.2450	hypothetical protein, conserved
	Tb927.7.2520	Uncharacterised protein family UPF0066, putative
	Tb927.7.4970	glutamine synthetase, putative (GS)
140	Tb927.10.7550	hypothetical protein, conserved
	Tb927.11.7520	hypothetical protein, conserved
	Tb927.3.3330	heat shock protein 20, putative
	Tb927.8.2020	agmatinase, putative
141	Tb927.2.1890	E2-like ubiquitin-conjugation enzyme (ATG3)
	Tb927.6.4950	mago nashi-like protein, putative
142	Tb927.11.14120	phenylalanyl-tRNA synthetase alpha chain, putative
	Tb927.4.1270	ruvB-like DNA helicase, putative
	Tb927.4.2000	ruvB-like DNA helicase, putative (RUVBL)
	Tb927.6.4670	MORN repeat-containing protein 1 (MORN1)
	Tb927.7.5290	hypothetical protein, conserved
	Tb927.9.5900	glutamate dehydrogenase (GDH)
143	Tb927.1.3000	amidohydrolase, putative
	Tb927.10.14140	pyruvate kinase 1 (PYK1)
	Tb927.11.6590	aminopeptidase, putative, metallo-peptidase, Clan MF, Family M17
	Tb927.6.2970	pseudouridine synthase A-like protein, putative
	Tb927.7.3770	YjeF family N-terminal domain/YjeF family C-terminal domain containing protein, putative
	Tb927.9.7540	calpain-like cysteine peptidase, putative, cysteine peptidase, Clan CA, family C2, putative
144	Tb927.10.14680	ribosome biogenesis protein, putative
	Tb927.10.8200	Ribosomal protein L1p/L10e family, putative

	Tb927.11.1390	class I transcription factor A, subunit 1 (CITFA-1)
	Tb927.11.14020	RNA-binding protein (NRBD2)
	Tb927.11.1410	class I transcription factor A, subunit 3 (CITFA-3)
	Tb927.11.14960	pumilio/PUF RNA binding protein 7, putative (PUF7)
	Tb927.11.1670	cysteine desulfurase
	Tb927.3.1350	hypothetical protein, conserved
	Tb927.3.2830	brix domain containing protein, putative
	Tb927.5.2080	guanosine monophosphate reductase, putative
	Tb927.5.840	KRI1-like family/KRI1-like family C-terminal, putative
	Tb927.5.970	class I transcription factor A, subunit 6 (CITFA-6)
	Tb927.6.2050	ribosome biogenesis regulatory protein (RRS1), putative
	Tb927.7.270	ribosome biogenesis protein, putative
	Tb927.8.5040	EMG1/NEP1 methyltransferase, putative
	Tb927.9.15060	rRNA processing protein, putative
145	Tb927.10.11130	SET domain containing protein, putative
	Tb927.10.12710	heat shock protein 110, putative
	Tb927.11.6370	leucine-rich repeat protein (LRRP), putative
146	Tb927.11.16020	RNA-binding protein, putative (DRBD10)
	Tb927.11.16400	kinetoplast-associated protein 3, putative (KAP3)
	Tb927.11.4690	mitochondrial DNA polymerase I protein B (POLIB)
147	Tb927.10.1930	hypothetical protein, conserved
	Tb927.10.2490	glucose-6-phosphate 1-dehydrogenase (G6PD)
	Tb927.10.7420	bromodomain factor 2 protein, putative
148	Tb11.v5.0325	retrotransposon hot spot (RHS) protein, putative
	Tb11.v5.0713	retrotransposon hot spot (RHS) protein, putative
	Tb927.11.2630	diphthamide biosynthesis enzyme Dph1/Dph2 domain containing protein, putative
	Tb927.3.4040	Ankyrin repeats (3 copies), putative
149	Tb927.11.15280	tRNA-sepcific adenosine deaminase (ADAT3)
	Tb927.8.4180	tRNA-specific adenosine deaminase (ADAT2)
150	Tb927.5.1270	hypothetical protein, conserved
	Tb927.9.10200	hypothetical protein, conserved
	Tb927.9.10880	hypothetical protein, conserved
151	Tb927.10.7140	membrane-bound acid phosphatase 2 (MBAP2)
	Tb927.8.5720	Met-10+ like-protein, putative
152	Tb927.1.1220	RWD domain-containing protein
	Tb927.10.13650	ARF-like 2-binding protein, putative
	Tb927.11.13690.1	unspecified product
	Tb927.6.2860	WD domain, G-beta repeat, putative
	Tb927.6.4420	tRNA (Guanine-1)-methyltransferase, putative
	Tb927.9.9020	ribosome-interacting GTPase 2, putative (RBG2)
153	Tb927.11.1300	UBA/TS-N domain containing protein, putative
	Tb927.3.4340	diphthamide synthesis protein, putative

	Tb927.6.2940	phosphopantothenoylcysteine decarboxylase, putative (PPCDC)
	Tb927.9.12830	hypothetical protein, conserved
154	Tb927.10.2990	nuclear cap binding complex subunit CBP110 (CBP110)
	Tb927.7.6270	peptidase t, putative, aminotripeptidase, putative
155	Tb927.10.8060	SET domain containing protein, putative
	Tb927.11.11290	heat shock protein 70, putative
	Tb927.7.6620	Probable N6-adenine methyltransferase, putative
156	Tb927.10.3950	hypothetical protein, conserved
	Tb927.5.1460	Possible lysine decarboxylase, putative
	Tb927.5.3520	queuine tRNA-ribosyltransferase, putative
	Tb927.6.3130	queuine tRNA-ribosyltransferase, putative
	Tb927.9.5190	proliferative cell nuclear antigen (PCNA), putative
157	Tb927.11.16510	hypothetical protein, conserved
	Tb927.6.4210	aldehyde dehydrogenase, putative (ALDH)
	Tb927.9.9710	Histidine phosphatase superfamily (branch 1), putative
158	Tb927.10.8290	eukaryotic translation initiation factor 3 subunit 8, putative
	Tb927.7.6090	hypothetical protein, conserved
159	Tb927.6.4920	S-adenosylmethionine synthetase, putative (METK1)
	Tb927.7.6450	tRNA-dihydrouridine synthase 3, putative
160	Tb927.10.12710	heat shock protein 110, putative
	Tb927.6.670	ribosomal RNA processing protein 45, exosome complex exonuclease (RRP45)
	Tb927.7.710	heat shock 70 kDa protein, putative (HSP70)
161	Tb927.11.10240	hslVU complex proteolytic subunit, threonine peptidase, Clan T(1), family T1B, ATP-dependent protease subunit HslV (HsIV)
	Tb927.11.14120	phenylalanyl-tRNA synthetase alpha chain, putative
	Tb927.11.2360	phenylalanyl-tRNA synthetase (beta subunit), putative
162	Tb927.10.7110	inositol-3-phosphate synthase, putative
	Tb927.11.16770	glucosamine-6-phosphate isomerase, putative
163	Tb927.11.7060	acidocalcisomal pyrophosphatase
	Tb927.2.450	retrotransposon hot spot protein 4 (RHS4), putative
	Tb927.3.1550	Rab3 GTPase-activating protein catalytic subunit, putative
	Tb927.5.3430	ubiquitin-activating enzyme E1, putative
	Tb927.9.7540	calpain-like cysteine peptidase, putative, cysteine peptidase, Clan CA, family C2, putative
164	Tb927.10.11310	intraflagellar transport protein 57/55 (IFT57/55)
	Tb927.10.13860	GPI-anchor transamidase subunit 8 (GPI8)
	Tb927.10.4040	3-keto-dihydrosphingosine reductase
	Tb927.10.4610	dolicholphosphate-mannose synthase, putative (DPMS)
	Tb927.11.13820	hypothetical protein, conserved
	Tb927.11.15760	GPI transamidase subunit Tta1 (TTA1)
	Tb927.2.1810	transcription silencer (ISWI)
	Tb927.5.1930	signal peptidase subunit, putative

	Tb927.8.5760	Ankyrin repeats (many copies)/Alpha/beta hydrolase family, putative
165	Tb927.1.4690	arginine N-methyltransferase (PRMT1)
	Tb927.10.12980	Multisite-specific tRNA:(cytosine-C(5))-methyltransferase, putative
	Tb927.10.14750	fibrillarin, putative
	Tb927.10.1960	hypothetical protein, conserved
	Tb927.10.3560	arginine N-methyltransferase, putative
	Tb927.10.7500	fibrillarin (NOP1)
	Tb927.8.3750	Nucleolar protein 56, putative (NOP56)
	Tb927.8.900	splicing factor TSR1 (TSR1)
	Tb927.9.5320	nucleolar RNA binding protein, putative
	Tb927.9.6870	RNA-binding protein, putative (RBSR1)
166	Tb927.3.640	hypothetical protein, conserved
	Tb927.5.1020	disulfide isomerase, putative
	Tb927.6.1140	dolichyl-P-Man:GDP-Man5GlcNAc2-PP-dolichyl alpha-1,2-mannosyltransferase, putative (ALG9)
	Tb927.6.1810	Alpha/beta hydrolase family, putative
	Tb927.9.3770	hypothetical protein, conserved
167	Tb927.10.2940	Soluble NSF attachment protein, SNAP, putative
	Tb927.7.2260	SEP domain containing protein, putative
168	Tb927.10.390	DUF2407 ubiquitin-like domain containing protein, putative
	Tb927.7.6850	trans-sialidase (TS)
169	Tb927.11.10960	2OG-Fe(II) oxygenase superfamily, putative
	Tb927.8.4100	hypothetical protein, conserved
170	Tb927.5.930	NADH-dependent fumarate reductase (FRDg)
	Tb927.7.2440	pyrroline-5-carboxylate reductase, putative (P5CR)
171	Tb927.11.11380	hypothetical protein, conserved
	Tb927.6.650	hypothetical protein, conserved
172	Tb927.11.9700	nascent polypeptide associated complex alpha subunit, putative
	Tb927.7.5490	arginine N-methyltransferase, type III (PRMT7)
173	Tb927.1.2750	hypothetical protein, conserved
	Tb927.7.5680	deoxyribose-phosphate aldolase, putative
174	Tb927.10.8720	CCR4-NOT transcription complex subunit 10, putative
	Tb927.8.1960	hypothetical protein, conserved
175	Tb927.8.1590	ubiquitin-protein ligase, putative (upl3)
	Tb927.9.13610	helicase, putative
176	Tb927.1.120	retrotransposon hot spot protein 4 (RHS4), putative
	Tb927.9.13990	RNA-binding protein, putative (DRBD2)
177	Tb927.9.15150	unspecified product
	Tb927.9.4200	fatty acyl CoA synthetase 2 (ACS2)
178	Tb927.4.1250	peroxisome biogenesis factor 1, putative
	Tb927.5.3920	peroxisome assembly protein, putative
179	Tb11.v5.0746	tatD related deoxyribonuclease, putative

	Tb927.10.4640	eukaryotic translation initiation factor 3 subunit L, putative (EIF3L)
	Tb927.11.15420	COP9 signalosome, subunit CSN8, putative
180	Tb927.11.15430	U5 small nuclear ribonucleoprotein component, putative, U5 snrnp-specific protein, putative
	Tb927.5.2290	ATP-dependent RNA helicase, putative
	Tb927.9.11110	PRP8 protein homologue, U5 snRNA-associated splicing factor
181	Tb927.11.14200	ubiquitin-conjugating enzyme E2, putative
	Tb927.7.1020	hypothetical protein, conserved
	Tb927.8.5880	eukaryotic translation initiation factor 1A, putative
182	Tb927.11.2340	hypothetical protein, conserved
	Tb927.11.2370	mRNA export factor MEX67, Nuclear RNA export factor (MEX67)
	Tb927.5.4380	kinetoplastid-specific phospho-protein phosphatase, putative
	Tb927.8.1840	Sec7 domain containing protein, putative
183	Tb927.3.4850	enoyl-CoA hydratase, mitochondrial precursor, putative
	Tb927.6.4540	3-hydroxy-3-methylglutaryl-CoA reductase, putative
	Tb927.7.1080	hypothetical protein, conserved
	Tb927.8.5510	apurinic/apyrimidinic endonuclease, putative
184	Tb927.10.14180	protein transport protein SEC13, putative
	Tb927.11.3770	Dpy-30 motif containing protein, putative
	Tb927.8.4790	Putative snoRNA binding domain containing protein, putative
	Tb927.9.5410	hypothetical protein, conserved
185	Tb927.10.1490	Temperature dependent protein affecting M2 dsRNA replication, putative
	Tb927.6.2640	importin alpha subunit, putative (TbKap60)
186	Tb927.11.230	cleavage and polyadenylation specificity factor, putative
	Tb927.4.1340	cleavage and polyadenylation specificity factor subunit, putative (CPSF3)
187	Tb927.10.11300	paraflagellar rod component, putative (PFC16)
	Tb927.6.4140	paraflagellar rod component, putative (PFC4)
188	Tb927.11.9880	Protein of unknown function (DUF2009), putative
	Tb927.9.10690	Protein of unknown function (DUF2009), putative
189	Tb927.1.2570	coatomer beta subunit (beta-coP)
	Tb927.10.7060	nucleoporin interacting component (NUP93), putative
	Tb927.11.11900	Coatomer subunit gamma (COPG)
	Tb927.11.14970	hypothetical protein, conserved
	Tb927.11.5400	signal recognition particle 54 kDa (SRP54)
	Tb927.7.1920	paraflagellar rod component, putative (PFC5)
190	Tb927.1.3830	glucose-6-phosphate isomerase, glycosomal (PGI)
	Tb927.10.13430	citrate synthase, putative
	Tb927.9.12110	6-phosphogluconate dehydrogenase, decarboxylating (gnD)
191	Tb927.4.3950	cytoskeleton-associated protein CAP5.5, putative, cysteine peptidase, Clan CA, family C2, putative, Calpain-like protein 1 (CAP5.5)
	Tb927.9.2470	nucleolar protein (NOP86)

192	Tb927.4.2670	Cysteine peptidase, Clan CF, family C15, pyroglutamyl-peptidase I, putative (PPI)
	Tb927.5.950	monothiol glutaredoxin, putative
	Tb927.7.6890	methyltransferase domain containing protein, putative
	Tb927.8.4930	hypothetical protein, conserved
	Tb927.9.7000	methyltransferase domain containing protein, putative
193	Tb927.3.860	Acyl carrier protein, mitochondrial, NADH-ubiquinone oxidoreductase complex I subunit, putative (ACP)
	Tb927.9.9840	lipoic acid containing carrier protein, putative (GCVH)
194	Tb927.11.180	electron transfer flavoprotein, putative
	Tb927.11.7540	electron-transfer-flavoprotein, alpha polypeptide, putative
195	Tb927.11.14870	NAD+ synthase, putative
	Tb927.4.2700	Hydroxymethylglutaryl-CoA lyase, mitochondrial, putative
196	Tb927.4.3350	N2227-like protein, putative
	Tb927.9.1510	D-ala D-ala ligase C-terminus/SET domain containing protein, putative
197	Tb927.5.2690	inositol-1(or 4)-monophosphatase 1, putative (IMPase 1)
	Tb927.9.6350	inositol-1(or 4)-monophosphatase, putative (IMPase)
198	Tb927.5.1000	ubiquitin-conjugating enzyme E2, putative, ubiquitin carrier protein, putative, ubiquitin-protein ligase, putative
	Tb927.9.1520	hypothetical protein, conserved
199	Tb927.7.4360	hypothetical protein, conserved
	Tb927.8.6930	serine/threonine-protein kinase NrkA (NRKB)
200	Tb927.6.2330	RGG protein (RGG1)
	Tb927.7.1790	Adenine phosphoribosyltransferase, putative
201	Tb927.6.2200	DJ-1 family protein, putative
	Tb927.8.1440	maoC-like dehydratase, putative
202	Tb927.10.8920	ras-like small GTPase, putative (TbGRP)
	Tb927.11.3320	ras-like small GTPase, putative (TbGTR)
203	Tb927.11.3310	ubiquitin-conjugating enzyme, putative
	Tb927.9.8000	ubiquitin-conjugating enzyme E2, putative, ubiquitin carrier protein, putative, ubiquitin-protein ligase, putative
204	Tb927.3.3450	ADP-ribosylation factor-like protein 3, putative (arl3)
	Tb927.3.3780	tryparedoxin 1a, putative (TXN1a)
205	Tb927.10.8360	Poly(A)-specific ribonuclease PARN-2
	Tb927.3.4490	protein farnesyltransferase alpha subunit, putative
	Tb927.7.460	protein farnesyltransferase beta subunit (TbPFT)
206	Tb927.3.4040	Ankyrin repeats (3 copies), putative
	Tb927.7.4000	glutathione synthetase, putative
207	Tb927.10.7060	nucleoporin interacting component (NUP93), putative
	Tb927.10.8170	nuclear pore complex protein (NUP155), putative, nucleoporin, putative
	Tb927.9.2320	methyltransferase domain containing protein, putative (POMP1)
208	Tb927.11.7840	ribonucleoside-diphosphate reductase large chain (RNR1)
	Tb927.3.610	N-acetyltransferase complex ARD1 subunit, putative

	Tb927.6.1950	N-acetyltransferase B complex (NatB) non catalytic subunit, putative
209	Tb927.7.1120	trypanothione/tryparedoxin dependent peroxidase 1, cytosolic, glutathione peroxidase-like protein 1 (TDPX1)
	Tb927.7.1130	trypanothione/tryparedoxin dependent peroxidase 2, glutathione peroxidase-like 2 (TDPX2)
210	Tb927.11.7100	cytoplasmic translation machinery associated protein, putative
	Tb927.8.4330	small GTP-binding protein Rab11 (RAB11)
211	Tb927.11.2090	choline kinase (EK1)
	Tb927.6.1530	Glutamine amidotransferase class-I, putative
212	Tb927.10.3260	Long-chain-fatty-acid--CoA ligase 5 (EC 6.2.1.3) (Long-chain acyl-CoA synthetase 5) (LACS 5), putative
	Tb927.2.4130	enoyl-CoA hydratase/Enoyl-CoA isomerase/3-hydroxyacyl-CoA dehydrogenase, putative
213	Tb927.1.1270	homocysteine S-methyltransferase, putative
	Tb927.11.13730	ornithine decarboxylase (ODC)
	Tb927.4.1740	hypothetical protein, conserved
	Tb927.6.1780	mitogen-activated protein kinase, putative, protein kinase, putative
	Tb927.8.2690	SET domain containing protein, putative
214	Tb927.11.5440	malic enzyme
	Tb927.11.5450	malic enzyme
215	Tb927.10.7930	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (PGAM)
	Tb927.11.8970	ribose 5-phosphate isomerase, putative
	Tb927.4.3320	uracil phosphoribosyltransferase, putative
	Tb927.9.7770	spermidine synthase (SpSyn)
216	Tb927.6.4480	valyl-tRNA synthetase, putative (ValRS)
	Tb927.6.4920	S-adenosylmethionine synthetase, putative (METK1)
217	Tb927.10.6970	dipeptidyl-peptidase 8-like serine peptidase, serine peptidase, Clan SC, Family S9B
	Tb927.10.8060	SET domain containing protein, putative
	Tb927.11.7150	NGG1 interacting factor 3-like
	Tb927.7.1910	pyridoxal phosphate containing glycine decarboxylase, putative (GCVP)
	Tb927.9.9820	glyceraldehyde-3-phosphate dehydrogenase, putative
218	Tb927.11.6210	sterol 14-alpha-demethylase (CYP51)
	Tb927.7.210	proline dehydrogenase
	Tb927.9.4190	fatty acyl CoA syntetase 1 (ACS1)
219	Tb927.2.4110	metallo-peptidase, Clan ME, Family M16, Mitochondrial-processing peptidase subunit alpha (MPPA)
	Tb927.4.600	Alpha/beta hydrolase family, putative
	Tb927.9.4520	metallo-peptidase, Clan ME, Family M16, Mitochondrial-processing peptidase subunit beta (MPPB)
220	Tb927.3.2090	aminopeptidase P1, putative, metallo-peptidase, Clan MG, Family M24, Xaa-Pro aminopeptidase, putative
	Tb927.8.2640	ubiquitin-activating enzyme E1, putative (UBA1)
221	Tb927.1.1380	serine/threonine protein phosphatase 2A regulatory subunit, putative
	Tb927.6.4760	T-complex protein 11, putative

222	Tb927.1.3950	alanine aminotransferase (ALAT)
	Tb927.11.3570	aminopeptidase, putative, metallo-peptidase, Clan MA(E) Family M1
	Tb927.11.5090	aspartate aminotransferase, mitochondrial
223	Tb927.11.11250	cytosolic malate dehydrogenase (cMDH)
	Tb927.8.4430	uridine phosphorylase
224	Tb927.10.4000	methylglutaconyl-CoA hydratase, mitochondrial precursor, putative
	Tb927.2.4590	branched-chain amino acid aminotransferase, putative
	Tb927.7.2100	GMP synthase, putative, glutamine amidotransferase, putative
	Tb927.8.980	phosphoacetylglucosamine mutase, putative, acetylglucosaminephosphomutase, putative, N-acetylglucosamine-phosphate mutase, putative
225	Tb927.10.13130	UTP-glucose-1-phosphate uridylyltransferase
	Tb927.10.14780	mitogen-activated protein kinase kinase kinase, putative (CBPK1)
	Tb927.9.1960	nitrilase, putative
226	Tb927.11.16480	enoyl-CoA hydratase/isomerase family protein, putative
	Tb927.6.2360	adenosine kinase, putative
	Tb927.8.5600	transaldolase, putative
227	Tb927.11.1560	1,2-Dihydroxy-3-keto-5-methylthiopentene dioxygenase, putative
	Tb927.11.15910	iron superoxide dismutase
	Tb927.7.1780	Adenine phosphoribosyltransferase, putative
228	Tb927.9.13490	aminopeptidase P, putative, metallo-peptidase, Clan MG, Family M24
	Tb927.9.2010	kynureninase, putative
229	Tb927.7.5210	Putative Phosphatase/Protein of unknown function DUF89, putative
	Tb927.8.1860	pitrilysin-like metalloprotease, metallo-peptidase, Clan ME, Family M16C
230	Tb927.10.10390	trypanothione reductase
	Tb927.10.11970	glutamine aminotransferase (GlnAT) (GlnAT)
231	Tb927.3.2960	inosine-adenosine-guanosine-nucleosidehydrolase, IAG-nucleoside hydrolase (IAGNH)
	Tb927.5.3830	dihydroorotate oxidase
	Tb927.6.2740	pyridoxal kinase (pdxK)
232	Tb927.11.6870	14-3-3 protein
	Tb927.11.9530	14-3-3-I protein
233	Tb927.10.14840	Mitochondrial ADP/ATP carrier protein 5a, putative (MCP5a)
	Tb927.2.2520	voltage-dependent anion-selective channel 2, Mitochondrial outer membrane protein porin 2 (VDAC2)
234	Tb927.10.11120	hypothetical protein, conserved
	Tb927.11.1600	tatD related deoxyribonuclease, putative
	Tb927.8.3550	mitogen-activated protein kinase 3, putative

Supplementary Table 3: List of suggested names for machine learning predicted protein complexes. Complex number matches those described in Table S2. Proportion of proteins within the complex annotated as 'hypothetical' or classed as 'essential' in any life-cycle stage *(Alsford et al., 2011)* are also displayed. The description highlights the proteins identified within the complex and denoting where there is orthogonal information for protein interaction. Complexes annotated as 'Complex mixture' contain more than two proteins with a number of different proposed functions or sub-cellular localisations.

Complex no.	Name	Proportion 'hypothetical'	Proportion 'essential'	Description
1	Glycerophospholipid metabolism complex	0/4	1/4	2 of 4 are components of glycerophospholipid metabolism - glycerophosphoryl diester phosphodiesterase; glycerol-3-phosphate dehydrogenase
2	DNA topoisomerase IB and hypothetical	1/2	1/2	Hypothetical described as localised to FAZ
3	AMPK	1/5	1/5	AMPK beta and gamma (Clemmens et al., 2009), two predicted AMPK alpha subunits and a hypothetical protein
4	Mitochondrial ribosome complex 1	3/6	2/6	5 of 6 have been identified in multiple mitochondrial ribosome Ips, including MRPL13 and KRIT2 (Zikova et al., 2008), and one remaining is a DEAD-box helicase, with homolofy to RRP3
5	Complex mixture 1	0/6	3/6	GAPDH, HSP70, quenine tRNA ribosyltransferase, METK1, NGG1 interacting factor 3-like protein and PCNA
6	Membrane protein complex	1/15	5/15	Membrane transporters of multiple categories including 3 multi-drug resistance transporters and two genes linked to pentamidine efficacy and resistance (Alsford et al., 2012)
7	Complex mixture 2	0/6	2/6	RNA binder, fructose-bisphosphate aldolase, ribonuclease, RHS, DUF2779 and tubulin tyrosine ligase
8	Complex mixture 3	2/6	0/6	Pex19, a dynein light chain, tryptophanyl tRNA synthetase, inosine-guanine hydrolase and two hypothetical proteins
9	Deoxyhypusine synthase	0/2	0/2	Interaction demonstrated in (Nguyen et al., 2013)
10	RuvB helicase complex and mixture 1	1/8	5/8	Two ruvB helicases with homology to RUVBL1 and 2, thought to interact in a dodecamer. Other proteins sinclude a SET domain, glutamate dehydrogenase, carnitine O-acetyltransferase, hypotheitcal, cysteinyl tRNA synthase, hsIVU protease
11	Threonyl tRNA synthase and	2/4	1/4	Two hypothetical proteins (consistent co-elution), ARM domain protein and threonyl tRNA synthase

	hypotheticals complex			
12	Heat shock protein 70/90 complex	1/5	1/5	HSP70 and HSP90, a S/T phosphatase 5 (interacting with HSP90 in (Jones et al., 2008)) a WD domain and hypothetical protein
13	Exon junction complex	0/3	0/3	Y14 and mago-nashi protein, shown to interact in META domain (DUF1935)
14	Hypothetical complex 1	3/3	2/3	3 hypothetical proteins.
15	tRNA synthetase complex	2/5	2/5	3 putative tRNA synthetases (tyrosyl/methionyl, glutaminyl and prolyl), and two hypothetical proteins
16	Pyruvate dehydrogenase complex	0/4	3/4	PDH E1 alpha and beta, dihydrolipoamide acetyltransferase E2 and threonyl-tRNA transferase
17	Nucelosome assembly complex	0/4	0/4	2 nucleosome assembly proteins, ARM like and leucine-repeat protein
18	E2 ubiquitin conjugating enzyme and CS domain complex	0/2	1/2	E2 like enzyme (ATG3) + CS domain containing protein
19	Major vault protein complex	1/4	0/4	three major vault proteins + hypothetical
20	Trafficking protein particle complex	0/2	0/2	Transport protein particle complex putative subunits 5 and 2
21	Proteasome activator and ankyrin repeat complex	0/2	0/2	Proteasome activator and ankyrin repeat complex
22	Ribosomal complex 1	0/7	3/7	L35a, L7a, L7, L13 and L6 components of 60S + L14 of 40s and ribosomal protein S7
23	SET domain protein and 60S ribosome complex	0/2	1/2	SET domain protein and putative RPL10A 60S ribosome protein
24	Heat shock protein and coronin (CRN12) complex	0/2	1/2	HSP and CRN12
25	Ribosomal complex 2	0/13	6/13	S19, L18 ribo proteins and L34, L10, L19, L13a, RPL27a/L27a/L28, L21e, L23, L9 of 60S and S16, S4, S8 of 40S
26	Vacuolar ATP synthase	0/4	0/4	4 vacuolar type ATP synthases, all linked to isometamidum resistance (Baker et al., 2015)
27	tRNA methyltransferase complex	0/5	0/5	Gcd10p and conserved protein (putative tRNA adenine methyltransferases). NOL1/NOP2/sun family - methyltransferase BLAST + 2 RHS proteins
28	Exosome complex	2/12	5/12	EAP1, EAP2, RRP4, RRP45, RRP41A, RRP40, CSL4 (Estevez 2001 and 2003)+ hypothetical, HSP70, translation elongation factor, RNApolI subunit (RPA135), conserved ATP binding

29	Prefoldin complex	0/4	0/4	4 prefoldin components
30	Proteasome nonATPase regulatory complex	0/14	11/14	RPT5, RPT4, RPN5, RPN6, RPN7, RPT2, RPT3, RPT6, RPN11, RPN9 + three more unnumbered + CIFTA-4 transcription factor
31	Proteasome core complex	0/14	11/14	14 subunits including one termed unspecified product that BLASTs as core proteasome subunit
32	Ribosomal complex 3	0/2	1/2	L38 and L44 of 60S ribosome
33	Ribosomal complex 4	0/2	0/2	2 acidic ribosomal proteins
34	Hypothetical complex 2	2/2	0/2	2 hypothetical proteins
35	Transportin and exosome protein	0/2	1/2	Transportin and exosome (RRP44p) proteins
36	Adenylyl cyclase associated and RNA binding protein	0/2	0/2	Adenyly cyclase associated protein and RBP
37	Tudor domain containing protein HSP84	0/2	1/2	Tudor domain protein and HSP84
38	2-oxoglutarate dehydrogenase complex	0/3	2/3	2 E1 and 1 E2 2-oxoglutarate dehydrogenase components
39	Dynein light chain and putative tRNA methyltransferase	0/2	2/2	Dynein light chain and tRNA methyltransferase
40	Mitochondrial processing peptidase complex	0/2	0/2	alpha and beta subunits
41	NOT5 and DUF protein	0/2	0/2	NOT5 and DUF protein
42	Radial spoke protein complex	0/2	2/2	RSP4/6 and RSP10
43	Hypothetical complex 3	2/2	1/2	Two hypothetical proteins
44	eIF3 complex	0/2	1/2	eIF3I and eIF3B
45	Hypothetical complex 4	2/2	1/2	Two hypothetical proteins
46	Resolvase domain containing protein and cysteine peptidase	0/2	2/2	Resolvase domain containing protein and cysteine peptidase
47	SEC24.1 and hypothetical protein	1/2	1/2	SEC24.1 and hypothetical protein
48	TatD related DNase and pescadillo domain	0/2	1/2	TatD related DNase and pescadillo domain containing protein

	containing protein			
49	Mitochondrial elongation factor and HSP10	0/2	1/2	Mitochondrial elongation factor and HSP10
50	Paraflagellar rod protein and NADH fumarate reductase	0/2	1/2	Paraflagellar rod protein and NADH fumarate reductase
51	eIF4e and hypothetical protein	1/2	1/2	eIF4e and hypothetical protein
52	AMP deaminase and ZFP family member ZC3H10	0/2	1/2	AMP deaminase and ZFP family member ZC3H10
53	mitochondrial RNA binding protein and ATP-dependent DEAD/H RNA helicase	0/2	1/2	mitochondrial RNA binding protein and ATP-dependent DEAD/H RNA helicase
54	polyA binding complex protein and zinc finger protein	1/2	1/2	polyA binding complex protein and zinc finger protein
55	Hypothetical and galactose oxidase domain	1/2	0/2	Hypothetical and galactose oxidase domain
56	Cysteine peptidase complex	0/2	0/2	two C2 domain containing proteins complex
57	GrpE and small glutamine tetratricopeptide repeat protein	0/2	1/2	GrpE and small glutamine tetratricopeptide repeat protein
58	Mitochondrial ribosome complex 2	2/2	0/2	2 hypothetical proteins, both identified in (Zikova et al., 2008), one MRPL29, and other also annotated as MAPK5
59	Hypothetical protein and ras like GTPase (NST)	1/2	0/2	Hypothetical protein and ras like GTPase (NST)
60	Putative transport protein complex	1/2	1/2	Vta1 (similar to vacuolar protein sorting associated) and hypothetical (blast match to USX1, intracellular transport protein)
61	NMD3 and regulator of chrosome condensation RCC1	0/2	1/2	RCC1 (annotated as ISWI complex, E3 ligase or GTPase) and NMD3 is ribosomal export protein
62	AcylCoA dehydrogenase and hypothetical mitochondrial import protein	1/2	2/2	Hypothetical protein annotated as Tim62, a mitochondrial membrane transport complex involved in tRNA import. Both proteins identified in a pulldown of Tim17 (Singha et al., 2012)
63	Kinesin complex	0/2	1/2	Two kinsesins, kin-C and kin-D, shown to interact in (Hu et al., 2012)

64	Hypothetical complex 5	2/2	1/2	Both annotated with BLAST hit - trichohyalin. One also as myosin heavy chain and other also inner centromere protein
65	DIGIT and NPAPL	0/2	2/2	A polyA polymerase protein and flagellar protein
66	Replication factor and glycosomal malate deydrogenase	0/2	1/2	Replication factor and glycosomal malate deydrogenase
67	ATG24 phosphoinositide binding complex	0/2	0/2	Both proteins contain PX domains and linked to suramin efficacy and resistance (Alsford et al., 2012), both classed as SNX proteins previously with one classed as ATG24 (Brennand et al., 2015)
68	cAMP specific phosphodiesterase complex	0/2	1/2	PDEB1 and 2
69	Hypothetical complex 6	2/2	2/2	Both BLASTs match Rad50 ATPase
70	Flagellar protein complex	0/2	1/2	KH1 (flagellum targetting protein) and PFC7 (paraflagellar protein)
71	2-oxoisovalerate dehydrogenase complex	0/2	0/2	alpha and beta subunits
72	PUF10 complex	1/2	1/2	PUF10 and hypothetical protein
73	Pyruvate dehydrogenase kinase complex	0/2	0/2	One named developmental regulated phospho-protein - BLASTs also as PDH kinase
74	SRP68-72 complex	1/3	3/3	SRP72 and 68 and glutamine hydrolysing carbomoyl phosphate synthase. Interaction between SRPs are identified previously in (Lustig et al., 2005)
75	Translation associated complex	1/3	2/3	BFR (associated with yeast polysomal mRNP, also chromosome segregation), hypothetical protein (BLAST to Bromodomain chromosome ass./RNAPII deg factor/splicing factor) and MKT1 (translation elongation in yeast)
76	H/ACA ribonucleoprotein complex	0/3	0/3	Cbf5p, GAR1 and putative Nhp2
77	Pseudouridine synthase complex	0/3	0/3	2 pseudouridine synthases and a putative tRNA methyltransferase
78	GRESAG complex	0/3	0/3	Three receptor type adenylate cyclase
79	N-acetyltransferase complex	0/3	2/3	Three N-acetyltransferase subunit
80	Heat shock protein 70 complex	0/3	1/3	2 HSP70s and BiP (also with HSP70 homology)
81	snRNP complex	0/3	2/3	SmD2, SmB, SmD3
82	Vaculolar sorting, Ran binding and hypothetical protein	1/3	1/3	Vaculolar sorting, Ran binding and hypothetical protein

83	Translation initiation factor, flagellar protein and hypothetical flagellar protein	1/3	0/3	Translation initiation factor, flagellar protein and hypothetical flagellar protein
84	ALBA complex	0/3	2/3	ALBA1, 3 and 4
85	F1/F0 ATP synthase complex 1	3/3	1/3	2 mitochondrial ATP synthase components identified in (Zikova et al., 2009) and a hypothetical protein
86	U6 spliceosome complex	0/3	0/3	LSm7p, LSm4p and PARN1. LSm proteins identified interacting in (Tkacz et al 2008)
87	Histone complex	0/3	0/3	H2A, H2B, H4
88	Complex mixture 4	0/3	2/3	Ribosome biogenesis, nucleolar RNA helicase and DNA topoisomerase proteins
89	Hypothetical + DUF proteins	1/2	1/2	Hypothetical + DUF proteins
90	Complex mixture 5	1/3	0/3	AMP binding protein thought to be part of NADH-ubiquinone oxidoreductase complex
91	Hypothetical complex 7	3/3	1/3	3 hypothetical proteins, one associated with cytoskeleton and one with ISWI domain homology.
92	Mitochondrial ribosome complex 3	2/3	0/3	All identified in (Zikova et al., 2008)
93	Complex mixture 6	1/3	0/3	Vesicular fusion protein SEC18, hypothetical protein and cold-shock DNA binding domain protein
94	Complex mixture 7	1/3	2/3	NUP-1 protein with CEP250 domain, hypothetical protein annotated as FAZ10 also with predicted CEP250 domain and SRP40 C-terminal domain protein
95	Complex mixture 8	2/3	1/3	DNA directed RNA polymerase and two hypotheticals
96	Complex mixture 9	2/3	0/3	2 hypotheticals annotated as flagellar protein and POMP respectively and an endo-beta-N-acetylglucosaminadase
97	Complex mixture 10	0/3	2/3	Lsm12 protein, cytoskeleton associated protein 17 and RNA binding protein 20
98	Mitochondrial inner membrane Tim10 and hypothetical protein	1/2	0/2	Mitochondrial inner membrane Tim10 and hypothetical protein
99	mRNA methytransferase cap and elongation factor complex	1/5	1/5	3 elongation factors, cap methyltransferase and hypothetical shown to associate in (Zamudio et al., 2009)
100	Complex mixture 11	0/4	0/4	Twy3 homolog, tRNA methyltransferase, two pyrophosphorylases and a WD domain containing protein
101	Complex mixture 12	1/4	1/4	Pestivirus endopeptidase, a hypothetical protein, lysosmal alpha-mannosidase and leucine rich repeat protein

102	Complex mixture 13	0/4	2/4	valsolin containing protein, WD40 ribosome biogensis protein, nucleolar GTP binding protein NOG1 and ribosome production factor protein
103	Complex mixture 14	2/4	0/4	2 hypothetical proteins, an oxidoreductase protein and leucine carboxymethyltransferase
104	Hypothetical and transport protein	1/2	0/2	Hypothetical protien and transport protein particle complex component
105	Peptidase complex	0/3	1/3	M76 and papain cysteine peptidase and DNL Zn finger type protein with Tim15 domain homology. M76 peptidase also has homology to ATP23 a mitochondrial inner membrane protein
106	1,2-diOH-3-keto-5-methylthiopentene deoxygenase and hypothetical	1/2	1/2	Hypothetical identified in (Zikova et al., 2008) as mitochondrial ribosome subunit
107	Complex mixture 15	0/3	2/3	RNA binding protein 17, mRNA export receptor Mtr2 and Zn finger containing protein
108	U3 ribonucleoprotein sub-complex	0/4	1/4	Pwp2 yeast homolog and three proteins containing Utp21, Utp13 and Utp12 domains indicitative of U3 ribonucleoprotein complex
109	Hypothetical protein and RNA binder (RBP29)	1/2	0/2	Hypothetical protein and RNA binder (RBP29)
110	Kinesin complex II	2/4	2/4	One hypothetical with flagellar pocket localisation, two kinesins and another hypothetical protein
111	Complex mixture 16	0/4	1/4	GAPDH, cyclophilin, glycerol-3-phosphate dehydrogenase and fructose 1-6-biphosphatase
112	ARP2/3 complex	0/5	0/5	5 ARP2/3 components
113	3-methylcrotonyl-coA carboxylase and dihydrolipoamide dehydrogenase complexes	0/5	0/5	3-methylcrotonyl-coA carboxylase alpha and beta subunits and three dihydrolipoamide dehydrogenase subunits
114	Complex mixture 17	1/4	2/4	2 predicted 40S ribosomal components, a Zn finger protein and a hypothetical protein
115	Hypothetical complex 8	3/4	1/4	Intraflagellar transport hypothetical protein, a START domain protein and a hypothetical protein with motile cillium GO annotation
116	Complex mixture 18	1/5	0/5	Mak10 N-acetyltransferase, Cgm1 mRNA N-methyltransferase, a third N-acetyltransferase, a hypothetical protein and a predicted heat shock protein
117	Complex mixture 19	1/5	3/5	Nup149, SEC31, RBP16, a hypothetical annotated as RBP8 and poly(A) export protein GLE2
118	Complex mixture 20	1/3	1/3	A hypothetical protein, a KRE33 ribosomal biogenesis protein, and UTP15 C-terminal domain putative U3 ribonucleoprotein component

119	eIF4G3 and ATP dependent DEAD/H helicase	0/2	1/2	eIF4G3 and ATP dependent DEAD/H helicase
120	eIF5 and hypothetical protein	1/2	1/2	eIF5 and hypothetical protein
121	Guide RNA associated complex	1/5	2/5	MRB1 associated proteins - GAP1 and 2 (Hashimi et al., 2009), and HslU2 protease, eIF2a and hypothetical protein
122	Complex mixture 21	1/3	2/3	Pex7, hypothetical protein and phosphatase
123	Complex mixture 22	2/3	1/3	2 hypothetical proteins, one annotated as mitochondrial cytochrome bc1 component and a V-type ATPase
124	Ankyrin repeat and alanine racemase protein	0/2	1/2	Ankyrin repeat and alanine racemase protein
125	Complex mixture 23	1/3	1/3	RPB7 RNA polymerase subunit, succinate dehydrogenase and hypothetical protein
126	Mitochondrial RNA binding complex	1/7	3/7	GBP25, GBP21, MRB1590 + tRNA uracil methyltransferase, NOL1/NOP2/sun domain, an XRN exonuclease and a hypothetical protein
127	Complex mixture 24	1/6	0/6	Hypothetical and alpha/beta hydrolase were identified as mitochondrial ribosome components in (Zikova et al., 2008),XRN exoribonuclease, a unspecified product (cAMP response gene), a mitochondrial DNA polymerase subunit and a DNA topoisomerase
128	Phosphoribosylt ransferase complex	0/8	3/7	3 PRS subunits and PRPS5 kinase, acetyl coA carboxylasem transitional ER ATPase (VCP) and phosphatidylinositol 4-phosphate 5-kinase
129	T-complex	0/8	7/8	All 8 chaperonin T-complex subunits
130	F1/F0 ATP synthase complex 2	0/8	4/8	6 ATP synthase subunits identified in (Zikova et al., 2009) and aspartyl aminopeptidase and O-phosphoryl-rRNA selenium transferase
131	Protein phosphatase complex	3/4	1/4	Protein phosphatase 1 and 3 hypothetical proteins (1 phosphatase inhibitor, 1 translating ribosome interactor, and 1 eIF3)
132	Complex mixture 25	1/4	2/4	A hypothetical protein, an adenylate kinase (ADKE), threonine dehydrogenase and deoxyuridine triphosphatase
133	Complex mixture 26	0/3	1/3	Peroxidoxin, tubulin-specific chaperone and an ATP-dependent RNA helicase
134	Mitochondrial ribosome 4	10/18	4/18	All proteins identified in (Zikova et al., 2008)
135	Complex mixture 27	0/4	2/4	seryl tRNA synthase, guanine deaminase, PFK/FBP and adenylosuccinate lyase
136	Hypothetical protein and unspecified product	1/2	0/2	RNA interference factor and iron sulfur cluster assembly protein
137	Pyruvate dehydrogenase	0/2	1/2	PDH E3 binding protien and adenoylsuccinate lyase

	binding protein and adenylosuccinate lyase			
138	Complex mixture 28	0/5	2/5	Paraflagellar rod component, proteasome activator 26, poly(A) ribonuclease 2 and dipthamide biosynthesis enzyme
139	Hypothetical complex 9	3/6	2/6	Glutamine synthase, polyadenylation factor, 3 hypothetical proteins and an unspecified product
140	Complex mixture 29	2/4	0/4	2 hypothetical proteins, an HSP20 annotated as a putative tRNA import complex and agmatinase
141	E2 ubiquitin conjugation enzyme and mago nashi protein	0/2	1/2	Exon junction component and ATG autophagy associated E2 ubiquitin conjugation enzyme
142	RuvB helicase complex and mixture 2	1/6	4/6	2 ruvB like DNA helicase proteins, phenylalanyl tRNA synthetase, MORN1, a hypothetical protein and glutamate dehydrogenase
143	Complex mixture 30	0/6	2/6	amidohydrolase, pyruvate kinase, aminopeptidase, pseudouridine synthase, YjeF N-terminal domain protein and calpain like cysteine peptidase
144	Ribosome biogenesis complex	1/16	9/16	CIFTA 1,3 and 6, PUF7, and mostly ribosomal biogenesis or rRNA associated proteins
145	Complex mixture 31	0/3	1/3	HSP110, leucine rich repeat protein and SET domain protein
146	Complex mixture 32	0/3	2/3	Kinetoplast associated protein 3, DRBD10 and mitochondrial polymerase subunit
147	Complex mixture 33	1/3	0/3	Bromodomain protein, glucose-6-phosphate 1-dehydrogenase and hypothetical protein
148	Complex mixture 34	0/4	0/4	2 retrotransposon hotspot proteins, dipthamide biosynthesis enzyme and ankyrin repeat protein
149	tRNA specific adenosine deaminase	0/2	0/2	ADAT2 and 3
150	Hypothetical complex 10	3/3	3/3	3 hypothetical proteins
151	Mitochondrial tRNA methylase and membrane bound acid phosphatase	0/2	1/2	Mitochondrial tRNA methylase and membrane bound acid phosphatase
152	Complex mixture 35	0/6	1/6	RWD domain protein, ARF like 2 binding protein, radial spoke protein 11, WD domain protein, tRNA guanine methyltransferase and ribosome interacting GTPase
153	UBA domain, diphthamide synthesis, phospanttocysteine decar, hypo	1/4	0/4	Ubiquitin like protein dipthamide biosynthesis enzyme, phosphopantothenoylysteine decarboxylase and a hypothetical protein

154	Nuclear cap binding protein and peptidase	0/2	0/2	CBP110
155	Complex mixture 36	0/3	2/3	SET domain (putative lysine metransferase), N6-adenine methyltransferase and HSP70
156	Queine tRNA ribosytransferase complex and mixture	1/5	4/5	2 queine tRNA ribosyltransferase, a lysine decarboxylase, a hypothetical protein and PCNA
157	Complex mixture 37	1/3	0/3	A hypothetical protein, aldehyde dehdrogenase and a histidine phopsphatase
158	eIF3 complex	1/2	1/2	eIF3C and hypothetical identified as eIF3A in (Rezende et al., 2014)
159	tRNA dihydrouridine synthase 3 and METK1	0/2	0/2	tRNA dihydrouridine synthase 3 and METK1
160	Heat shock protein complex	0/3	2/3	HSP110, HSP70 and RRP45(exosome)
161	Phenylalanyl tRNA synthetase complex	0/3	2/3	Phenylalanyl tRNA synthetse alpha and beta and hslVU peptidase component
162	inositol-3-phosphate synthase and glucosamine-6-phosphate isomerase	0/2	0/2	inositol-3-phosphate synthase and glucosamine-6-phosphate isomerase
163	Complex mixture 38	0/5	1/5	calpain peptidase, Rab GAP, Ubiquitin E1 enzyme, pyrophophatase and RHS4
164	GPI associated complex and mixture	0/9	4/9	GPI anchor transamidase (GPI8), signal peptidase, GPI transamidase (TTA1), dolicholphosphate-mannose synthase, Ankyrin repeat protein, ISWI, IFT57/55, 3-keto-dihydrosphingosine reductase and hypothetical protein
165	Nucleolar associated complex	1/10	3/10	splicing factor (TSR1), RSBR1 RNA binding protein, 2 arginine methyltransferases (functional association from personal communication with Laurie Read), NOP1 (fibrillarin) NOP56, nucleolar RNA bind, cytosine tRNA metransferase, fibrillarin, a hypothetical protein
166	Complex mixture 39	2/5	0/5	Serine peptidase, 2 hypothetical proteins, disulfide isomerase, and ALG9 mannosyltransferase
167	Putative Golgi/ER biogenesis complex	0/2	0/2	NSF attachment protein (SNARE type) and SEP containing protein with NSFL1 cofactor Blast
168	Trans-sialidase and ubiquitin like domain protein	0/2	0/2	Trans-sialidase and ubiquitin like domain protein
169	FE(II) oxygenase and	1/2	0/2	Hypothetical protein annotate as FLA1BP

	hypothetical protein			
170	NADH fumarate reductase and pyrroline-5-carboxylate reductase	0/2	1/2	NADH fumarate reductase and pyrroline-5-carboxylate reductase
171	Hypothetical complex 11	2/2	1/2	2 hypothetical proteins
172	Arginine N-metrans (PRMT7) and nascent polypeptide associated complex protein	0/2	1/2	PRMT7 and dubious annotation of nascent polypeptide associated complex protein
173	Hypothetical protein and deoxyribose-phosphate aldolase	1/2	0/2	Hypothetical protein and deoxyribose-phosphate aldolase
174	CNOT 10 and 11 complex	1/2	2/2	Hypothetical protein, human homolog CNOT11 and CNOT11 protein. CNOT10 and 11 known to interact in human (Mauxion et al., 2013)
175	Helicase and ubiquitin ligase	0/2	0/2	Ubiquitin ligase also annotated as yeast TOM1 homolog with role in nuclear mRNA export
176	DRBD2 and RHS4	0/2	0/2	DRBD2 and RHS4
177	Unspecified product and fatty acyl CoA synthetase 2	0/2	1/2	Unspecified product annotated as ribosomal protein
178	Peroxisome biogenesis associated complex	0/2	0/2	Peroxisome biogenesis protein and assembly protein PEX1 and 6
179	eIF3L and COP9 signalosome (eIF3K) and tatD deoxyribonuclease	0/3	1/3	COP9 signalsome protein annotated as eIF3K (Rezende et al., 2014)
180	U5 spliceosome complex	0/3	2/3	U5 spliceosome protein, PRP8 and ATP dependent RNA helicase with hoology to U5 spliceosome component. All identified in pull down experimetn of U5 spliceosome in (Silva et al., 2011)
181	Complex mixture 40	1/3	0/3	E2 ubiquitin conjugating enzyme, hypothetical protein and eIF1A
182	Complex mixture 41	1/4	3/4	hypo, MEX67 mRNA export, kinetoplastid phosphatase, SEC7
183	Complex mixture 42	1/4	0/4	hypothetical protein, 3-hydroxy-3methyl glutaryl coA reductase, endonuclease, enoyl CoA hydratase
184	Complex mixture 43	1/4	2/4	hypothetical protein, snoRNA binder, Dpy-30 domain protein and SEC13

185	Importin alpha and M2 dsRNA replication associated protein	0/2	2/2	Importin annotated as Kap60 and M2 dsRNA protein is annotated as putative splicing protein associated with nuclear speckles
186	Cleavage and poly(A) specificity complex	0/2	1/2	2 cleavage and poly(A) specficity factor proteins, CPSF2 and 3
187	Paraflagellar rod complex	0/2	0/2	PFC4 and 16
188	DUF complex	0/2	0/2	two DUF proteins with same domain (DUF2009)
189	Coatomer complex and mixture	1/6	4/6	Coatomer beta and gamma, together with PFC5, SRP54, hypothetical protein and NUP93
190	Complex mixture 44	0/3	1/3	PGI, 6-phosphogluconate dehydrogenase, citrate synthase
191	NOP86 and cytoskeleon associated protein	0/2	2/2	NOP86 and cytoskeleon associated protein
192	Complex mixture 45	1/5	0/5	2 methyltranferase domain containing proteins, a hypothetical protein, monthiol glutaredoxin and a cysteine peptidase
193	Acyl carrier protein and lipoic acid carrier protein	0/2	0/2	ACP, member of NADHubiquinone oxidoreductase complex and lipoic acid carrier protein, GCVH
194	Electron transfer flavoprotein complex	0/2	0/2	2 electron transfer flavoproteins
195	Hydroxy-meglutaryl-coA lyase and NAD+ synthase	0/2	1/2	Hydroxy-meglutaryl-coA lyase and NAD+ synthase
196	D-alanine D-alanine ligase and N2227 protein	0/2	1/2	D-alanine D-alanine ligase and N2227 protein
197	Inositol-1 monophosphatase complex	0/2	0/2	2 IMPases
198	Ubiquitin E2 conjugation enzyme and hypothetical protein	1/2	2/2	Ubiquitin E2 conjugation enzyme and hypothetical protein
199	Hypothetical protein and S/T kinase NrkA	1/2	1/2	Hypothetical protein and S/T kinase NrkA
200	RGG domain protein and adenine phosphoribosyltransferase	0/2	1/2	RGG domain protein and adenine phosphoribosyltransferase
201	maoC dehydratase and DJ-1 family protein	0/2	1/2	maoC dehydratase and DJ-1 family protein

202	ras-like GTPase complex	0/2	0/2	2 ras GTPases, GTR and GTP
203	Ubiquitin conjugation complex	0/2	0/2	2 E2 ubiqutin conjugation enzymes
204	ADP ribosylation factor and tryparedoxin 1a	0/2	1/2	ADP ribosylation factor and tryparedoxin 1a
205	Farnesyltransferase complex	0/3	2/3	Farnesyltranfserase alpha and beta subunits and PARN2
206	Glutathione synthetase and ankyrin repeat protein	0/2	0/2	Glutathione synthetase and ankyrin repeat protein
207	Nuclear pore complex	0/3	2/3	NUP155, Nup93 and POMP1 a mitochondrial outer membrane protein
208	N-acetyltransferase complex	0/3	1/3	N-acetyltransferases ARD1 and NatB and ribonucleoside-diphosphate reductase RNR1
209	Tryparedoxin-peroxidase complex	0/2	1/2	TDPX1 and 2
210	RAB11 and translation machinery associated protein	0/2	1/2	RAB11 and translation machinery associated protein
211	Glutamine amidotransferase and choline kinase	0/2	1/2	Glutamine amidotransferase and choline kinase
212	Fatty acid CoA ligase and enoyl CoA hydratase	0/2	0/2	Fatty acid CoA ligase and enoyl CoA hydratase
213	Complex mixture 46	1/5	1/5	Hypothetical protein, SET domain protein, MAPK4, homocysteineSmethyltransferase and ornithine decarboxylase
214	malic enzymes	0/2	0/2	2 malic enzyme proteins
215	Complex mixture 47	0/4	3/4	ribose 5-phosphate isomerase, spermidine synthase, uracil phosphoribosyltransferase, PGAM
216	Valyl tRNA synthetase and METK1	0/2	0/2	Valyl tRNA synthetase and METK1
217	glyceraldehyde-3-ph DH, NGG1, SET dom, serine peptidase and PLP G decarboxylase	0/5	1/5	Glyceraldehyde-3-phosphate dehydrogenase, NGG1 interacting factor like protein, SET domain protein, serine peptidase and pyridoxal phosphate containing glycine decarboxylase
218	Complex mixture 48	0/3	1/3	Fatty acyl coA synthase (ACS1), proline dehydrogenase and sterol alpha demethylase
219	Mitochondrial processing peptidase complex	0/3	1/3	Mitochondrial processing peptidase alpha and beta subunits shown to interact in (Desy et al., 2012) and an alpha beta hydrolase

220	Metallopeptidase and ubiquitin activating enzyme	0/2	1/2	Metallopeptidase and E1 ubiqutin activating enzyme UBA1
221	S/T protein phosase 2A and T complex like protein	0/2	1/2	S/T protein phosase 2A and T complex like protein
222	Complex mixture 49	0/3	0/3	aminopeptidase, alanine aminotransferase and aspartate aminotransferase
223	cytosolic malate dehydrogenase and uridine phosphorylase	0/2	0/2	cytosolic malate dehydrogenase and uridine phosphorylase
224	Complex mixture 50	0/4	0/4	GMP synthase, branched chain amino acid aminotransferase, phosphoacetylglucosamine mutase, methylglutaconyl-coA hydratase
225	Complex mixture 51	0/3	1/3	UTP-glucose-1-phosphate uridyltransferase, nitrilase, MAPKKK
226	Complex mixture 52	0/3	2/3	adenosine kinase, transaldolase and enoyl CoA hydratase
227	adenine phribosyltrans, diO2ase, iron superoxide dismutase	0/3	0/3	adenine phosphoribosyltransferase, dioxygenase, iron superoxide dismutase
228	Kynureninase and aminopeptidase	0/2	1/2	Kynureninase and aminopeptidase
229	Phosphatase and metalloprotease	0/2	0/2	Phosphatase and metalloprotease
230	Trypanothione reductase and glutamine aminotransferase	0/2	0/2	Trypanothione reductase and glutamine aminotransferase
231	Complex mixture 53	0/3	0/3	Pyridoxal kinase, nucleosidehydrolase, dihydroorotate oxidase
232	14-3-3 protein complex	0/2	2/2	14-3-3 subunits I and II, identified in (Inoue et al., 2005)
233	Mito ADP/ATP carrier protein and voltage-dependent anion-selective channel	0/2	0/2	MCP5a and VDAC2
234	Complex mixture 54	1/3	1/3	Hypothetical protein, tatD deoxyribonuclease and MAPK3

Supplementary Table 4: Proteins classified in cell-cycle regulated clusters. Table displays the cluster a protein was classified into along with the maximum fold change detected across the time-course of cell division.

Gene ID	Gene description	Cluster	Max fc
Tb927.10.6070	universal minicircle sequence binding protein (UMSBP), putative,predicted zinc finger protein	G2&M	2.2
Tb927.7.2160	hypothetical protein, conserved	G2&M	2.0
Tb927.8.3220	exonuclease, putative	G2&M	2.0
Tb927.11.12070	hypothetical protein, conserved	G2&M	2.0
Tb927.10.12920	hypothetical protein, conserved	G2&M	2.0
Tb927.7.4850	kinetoplastid kinetochore protein 5 (kkt5)	G2&M	1.9
Tb927.10.13100	hypothetical protein, conserved	G2&M	1.9
Tb927.10.15400	kinesin, putative	G2&M	1.9
Tb927.5.440	trans-sialidase, putative	G2&M	1.9
Tb927.10.4990	cell division related protein kinase 2, putative,CDC2-related protein kinase (CRK3)	G2&M	1.9
Tb927.6.570	hypothetical protein, conserved	G2&M	1.8
Tb927.4.3430	hypothetical protein, conserved	G2&M	1.8
Tb927.11.15800	hypothetical protein, conserved	G2&M	1.8
Tb927.3.2140	transcription activator, putative	G2&M	1.8
Tb927.4.4440	adenylyl cyclase (GRESAG 4.4)	G2&M	1.8
Tb927.11.16390	cyclin dependent kinases regulatory subunit, putative (CKS1)	G2&M	1.7
Tb927.8.3560	DNA repair and recombination helicase protein PIF5 (PIF5)	G2&M	1.7
Tb927.9.13140	hypothetical protein, conserved	G2&M	1.7
Tb927.11.15780	hypothetical protein, conserved	G2&M	1.7
Tb11.v5.0609	DNA-directed RNA polymerase III largest subunit, putative	G2&M	1.7
Tb927.10.3380	60S acidic ribosomal protein P2, putative	G2&M	1.7
Tb927.8.7120	squalene synthase, putative	G2&M	1.7
Tb927.2.3720	ubiquitin-conjugating enzyme, putative	G2&M	1.7
Tb927.6.3760	kinetoplastid kinetochore protein 15 (kkt15)	G2&M	1.6
Tb927.11.7050	hypothetical protein, conserved	G2&M	1.6
Tb927.9.14290	hypothetical protein, conserved	G2&M	1.6
Tb927.11.11240	hypothetical protein, conserved	G2&M	1.6
Tb927.9.13070	predicted heat shock factor binding protein	G2&M	1.6
Tb927.8.960	hypothetical protein, conserved	G2&M	1.6
Tb927.11.16380	hypothetical protein, conserved	G2&M	1.6
Tb927.10.8710	centrin, putative	G2&M	1.6
Tb927.6.1540	hypothetical protein, conserved	G2&M	1.6
Tb927.3.860	acyl carrier protein,acyl carrier protein, mitochondrial precursor, putative (ACP)	G2&M	1.6
Tb927.9.9520	zinc finger protein family member, putative (ZC3H29)	G2&M	1.6
Tb927.7.4680	hypothetical protein, conserved	G2&M	1.6
Tb927.3.3240	hypothetical protein, conserved	G2&M	1.5
Tb927.10.4050	serine palmitoyltransferase, putative	G2&M	1.5

Tb927.11.16820	hypothetical protein, conserved	G2&M	1.5
Tb927.10.6060	universal minicircle sequence binding protein (UMSBP), putative,DNA-binding protein HEXBP, putative,zinc finger protein	G2&M	1.5
Tb927.10.6410	mismatch repair protein MSH8, putative,mismatch repair protein	G2&M	1.5
Tb927.11.13100	hypothetical protein, conserved	G2&M	1.5
Tb927.10.890	kinesin, putative	G2&M	1.5
Tb927.10.4750	hypothetical protein, conserved	G2&M	1.5
Tb927.8.4190	hypothetical protein, conserved	G2&M	1.5
Tb927.11.14410	hypothetical protein, conserved	G2&M	1.5
Tb927.11.7600	hypothetical protein, conserved	G2&M	1.5
Tb927.9.10250	small nuclear ribonucleoprotein Sm-F (Sm-F)	G2&M	1.5
Tb927.8.3460	hypothetical protein, conserved	G2&M	1.5
Tb927.5.3310	hypothetical protein, conserved	G2&M	1.5
Tb927.11.12820	ribonucleoside-diphosphate reductase small chain (RNR2)	G2&M	1.5
Tb927.11.4770	retrotransposon hot spot protein (RHS, pseudogene), putative	G2&M	1.5
Tb927.6.200	receptor-type adenylate cyclase GRESAG 4, putative	G2&M	1.5
Tb927.10.11270	RNA-binding protein, putative (RBP23)	G2&M	1.5
Tb927.10.14460	leucine-rich repeat protein (LRRP), putative	G2&M	1.5
Tb927.10.8240	hypothetical protein, conserved	G2&M	1.5
Tb927.8.3780	hypothetical protein, conserved	G2&M	1.5
Tb927.7.2290	hypothetical protein, conserved	G2&M	1.5
Tb927.11.640	hypothetical protein, conserved	G2&M	1.4
Tb927.11.13690.1	radial spoke protein RSP11, putative	G2&M	1.4
Tb927.8.7340	trans-sialidase, putative,neuraminidase, putative	G2&M	1.4
Tb927.11.3090	LisH domain-containing protein FOPNL, putative,FOP-related protein of 20 kDa, putative (fopnl)	G2&M	1.4
Tb927.10.10000	hypothetical protein, conserved	G2&M	1.4
Tb927.1.4700	hypothetical protein, conserved	G2&M	1.4
Tb927.11.13980	hypothetical protein, conserved	G2&M	1.4
Tb927.9.12680	hypothetical protein, conserved	G2&M	1.4
Tb927.10.10780	hypothetical protein, conserved	G2&M	1.4
Tb927.8.6750	translationally controlled tumor protein (TCTP), putative	G2&M	1.4
Tb927.5.1960	hypothetical protein, conserved	G2&M	1.4
Tb927.11.5260	SNF2 DNA repair protein, putative,SNF2 family protein	G2&M	1.4
Tb927.10.11580	predicted WD40 repeat protein	G2&M	1.4
Tb927.3.2330	kinetoplastid kinetochore protein 17 (kkt17)	G2&M	1.4
Tb927.11.1300	hypothetical protein, conserved	G2&M	1.4
Tb927.6.3360	hypothetical protein, conserved	G2&M	1.4
Tb927.3.4950	hypothetical protein, conserved	G2&M	1.4
Tb927.6.3290	intraflagellar transport protein 20, putative (IFT20)	G2&M	1.4
Tb927.8.910	hypothetical protein, conserved	G2&M	1.4
Tb927.10.3470	hypothetical protein, conserved	G2&M	1.4
Tb927.11.6340	hypothetical protein, conserved	G2&M	1.4
Tb927.4.4620	cytochrome oxidase subunit VIII (COXVIII)	G2&M	1.4

Tb927.10.9080	pteridine transporter, putative	G2&M	1.4
Tb927.7.1270	hypothetical protein, conserved	G2&M	1.4
Tb927.11.13150	hypothetical protein, conserved	G2&M	1.4
Tb927.11.14710	hypothetical protein, conserved	G2&M	1.4
Tb927.11.10520	kinetoplastid kinetochore protein 2,serine/threonine-protein kinase, putative,serine/threonine-protein kinase, putative (kkt2)	G2&M	1.4
Tb927.10.7430	GTP-binding protein, putative	G2&M	1.4
Tb927.9.2760	hypothetical protein, conserved	G2&M	1.4
Tb927.4.1750	hypothetical protein, conserved	G2&M	1.4
Tb927.11.13520	hypothetical protein, conserved	G2&M	1.4
Tb927.9.11350	hypothetical protein, conserved	G2&M	1.4
Tb11.v5.0971	mitochondrial DNA polymerase I protein A, putative	G2&M	1.4
Tb927.5.3630	hypothetical protein, conserved	G2&M	1.4
Tb927.11.16750	hypothetical protein, conserved	G2&M	1.4
Tb927.10.9230	adaptin complex 4 subunit, putative,beta-adaptin 4 protein, putative	G2&M	1.4
Tb927.4.2190	GTP-binding protein, putative	G2&M	1.4
Tb927.11.9860	hypothetical protein, conserved	G2&M	1.4
Tb927.10.3300	hypothetical protein, conserved	G2&M	1.4
Tb927.11.10100	hypothetical protein, conserved	G2&M	1.3
Tb927.7.4610	hypothetical protein, conserved	G2&M	1.3
Tb927.10.5750	hypothetical protein, conserved,protocatechuate 4,5-dioxygenase-like protein	G2&M	1.3
Tb927.1.1270	homocysteine S-methyltransferase, putative	G2&M	1.3
Tb927.8.2220	hypothetical protein, conserved	G2&M	1.3
Tb927.11.6560	hypothetical protein, conserved	G2&M	1.3
Tb927.8.2150	hypothetical protein, conserved	G2&M	1.3
Tb927.8.2690	hypothetical protein, conserved	G2&M	1.3
Tb927.6.5130	60S acidic ribosomal protein P2, putative	G2&M	1.3
Tb927.3.3950	hypothetical protein, conserved	G2&M	1.3
Tb927.7.5730	hypothetical protein, conserved	G2&M	1.3
Tb927.10.14330	hypothetical protein, conserved	G2&M	1.3
Tb927.10.4620	peptidyl-prolyl cis-trans isomerase, putative (PPIase)	G2&M	1.3
Tb927.10.5860	hypothetical protein, conserved	G2&M	1.3
Tb927.11.1240	hypothetical protein, conserved	G2&M	1.3
Tb927.9.1850	60S ribosomal protein L35, putative	G2&M	1.3
Tb927.4.3350	hypothetical protein, conserved	G2&M	1.3
Tb927.9.7580	hypothetical protein, conserved	G2&M	1.3
Tb927.5.4320	zinc finger protein family member, putative (FIP1)	G2&M	1.3
Tb927.11.12520	hypothetical protein, conserved	G2&M	1.3
Tb927.3.4270	hypothetical protein, conserved	G2&M	1.3
Tb927.3.1890	cytochrome c oxidase assembly protein, putative	G2&M	1.3
Tb927.8.2910	mannosyl-oligosaccharide 1,2-alpha-mannosidase IB, putative	G2&M	1.3
Tb927.11.14950	zinc finger protein 2 (ZFP2)	G2&M	1.3
Tb927.4.4970	myosin heavy chain kinase A, putative	G2&M	1.3

Tb927.11.7670	hypothetical protein	G2&M	1.3
Tb927.2.4140	hypothetical protein, conserved	G2&M	1.3
Tb927.7.1630	hypothetical protein, conserved	G2&M	1.3
Tb927.1.4300	chaperone protein DNAj, putative	G2&M	1.3
Tb927.10.11900	hypothetical protein, conserved	G2&M	1.3
Tb927.10.9160	hypothetical protein, conserved	G2&M	1.3
Tb927.10.7620	mitochondrial ATP-dependent zinc metallopeptidase, putative,metallo-peptidase, Clan MA(E) Family M41	G2&M	1.3
Tb927.10.4930	protein phosphatase 2C, putative	G2&M	1.3
Tb927.11.20300	variant surface glycoprotein (VSG), putative	G2&M	1.3
Tb927.7.5050	hypothetical protein, conserved	G2&M	1.3
Tb927.11.4970	predicted WD40 repeat protein	G2&M	1.3
Tb927.10.7300	hypothetical protein, conserved	G2&M	1.3
Tb927.11.7930	RNA polymerase B subunit RPB8, putative (RPB8)	G2&M	1.3
Tb11.1650	variant surface glycoprotein (VSG, pseudogene), putative,chrXI additional, unordered contigs	G2&M	1.3
Tb927.10.7780	mitogen-activated protein kinase 1, putative	G2&M	1.3
Tb927.8.4170	RNA-binding protein, putative (MRD1)	G2&M	1.3
Tb927.4.670	conserved protein, unknown function	G2&M	1.3
Tb927.11.5850	RNA-binding protein, putative (RBP38)	G2&M	1.3
Tb927.8.1750	hypothetical protein, conserved	G2&M	1.3
Tb927.8.1130	protein phosphatase with EF-Hand domains, putative	G2&M	1.3
Tb927.8.2870	conserved protein	G2&M	1.3
Tb927.1.3450	hypothetical protein, conserved	G2&M	1.3
Tb927.5.3200	hypothetical protein, conserved	G2&M	1.3
Tb927.10.6960	hypothetical protein, conserved	G2&M	1.3
Tb927.11.2350	hypothetical protein, conserved	G2&M	1.3
Tb927.11.8100	hypothetical protein, conserved	G2&M	1.3
Tb927.6.3710	hypothetical protein, conserved	G2&M	1.3
Tb927.11.6440	hypothetical protein, conserved	G2&M	1.3
Tb927.10.9870	hypothetical protein, conserved	G2&M	1.3
Tb927.1.1210	conserved protein, unknown function	G2&M	1.3
Tb927.4.5180	tousled-like kinase I (TLK1)	G2&M	1.3
Tb927.10.8580	ADP-ribosylation factor 3, putative,ADP-ribosylation factor-like protein, putative (ARL3)	G2&M	1.3
Tb927.10.10710	hypothetical protein, conserved	G2&M	1.3
Tb927.10.11730	zinc finger protein family member, putative (ZC3H33)	G2&M	1.3
Tb927.11.650	hypothetical protein, conserved	G2&M	1.3
Tb927.7.2060	DNA repair protein, putative	G2&M	1.3
Tb927.11.7110	hypothetical protein, conserved	G2&M	1.3
Tb927.11.3080	vesicular protein trafficking mediator, putative	G2&M	1.3
Tb927.6.300	receptor-type adenylate cyclase GRESAG 4, putative	G2&M	1.3
Tb927.4.2910	hypothetical protein, conserved	G2&M	1.3
Tb927.11.12930	DEAD-box helicase, putative	G2&M	1.3
Tb927.11.14460	ADP-ribosylation factor GTPase activating protein 1, putative	G2&M	1.3
Tb927.9.7470	purine nucleoside transporter (NT10)	G2&M	1.3

Tb927.10.6540	kinetoplastid-specific phospho-protein phosphatase, putative	G2&M	1.3
Tb927.10.12220	hypothetical protein, conserved	G2&M	1.3
Tb927.1.990	Enkuring domain-containig protein	G2&M	1.3
Tb927.9.11820	hypothetical protein, conserved	G2&M	1.3
Tb927.9.15050	trypanin-related protein, putative	G2&M	1.3
Tb927.8.4640	flagellar protofilament ribbon protein, putative	G2&M	1.3
Tb927.4.4610	hypothetical protein, conserved	G2&M	1.2
Tb927.10.5270	mitogen-activated protein kinase kinase 5, putative (MKK5)	G2&M	1.2
Tb927.10.1200	hypothetical protein, conserved	G2&M	1.2
Tb927.10.1450	hypothetical protein, conserved	G2&M	1.2
Tb927.8.6910	cyclophilin, putative	G2&M	1.2
Tb927.4.4370	hypothetical protein, conserved	G2&M	1.2
Tb927.10.14560	hypothetical protein, conserved	G2&M	1.2
Tb927.2.5950	hypothetical protein, conserved	G2&M	1.2
Tb927.10.10330	hypothetical protein, conserved	G2&M	1.2
Tb11.02.5105b	hypothetical protein, conserved	G2&M	1.2
Tb927.10.8610	hypothetical protein, conserved	G2&M	1.2
Tb927.3.4230	subtilisin-like serine peptidase,serine peptidase, clan SB, family S8-like protein	G2&M	1.2
Tb927.2.6080	hypothetical protein, conserved	G2&M	1.2
Tb927.9.5380	hypothetical protein, conserved	G2&M	1.2
Tb927.4.720	hypothetical protein, conserved	G2&M	1.2
Tb927.8.2800	hypothetical protein, conserved	G2&M	1.2
Tb927.10.13650	ARF-like 2-binding protein, putative,BART	G2&M	1.2
Tb927.7.4780	hypothetical protein, conserved	G2&M	1.2
Tb927.9.13350	hypothetical protein, conserved	G2&M	1.2
Tb927.11.3930	hypothetical protein, conserved	G2&M	1.2
Tb927.9.8260	rhomboid-like protein,serine peptidase, Clan S- , family S54, putative	G2&M	1.2
Tb927.4.1150	hypothetical protein, conserved	G2&M	1.2
Tb927.11.7280	DNA-directed RNA polymerase II, subunit 9, putative (RPB9)	G2&M	1.2
Tb927.7.3000	kinesin, putative	G2&M	1.2
Tb927.8.2590	hypothetical protein, conserved	G2&M	1.2
Tb927.8.1370	hypothetical protein, conserved	G2&M	1.2
Tb927.7.3740	hypothetical protein, conserved	G2&M	1.2
Tb927.3.2040	kinesin, putative	G2&M	1.2
Tb927.9.15020	hypothetical protein, conserved	G2&M	1.2
Tb927.3.2370	hypothetical protein, conserved	G2&M	1.2
Tb927.11.2020	hypothetical protein, conserved	G2&M	1.2
Tb927.1.1630	transcription elongation factor 1 homolog (ELOF1)	G2&M	1.2
Tb927.8.2300	hypothetical protein, conserved	G2&M	1.2
Tb927.4.1170	ankyrin, putative	G2&M	1.2
Tb927.4.3130	Spef1,flagellar component	G2&M	1.2
Tb927.11.9750	hypothetical protein, conserved	G2&M	1.2
Tb927.11.8890	DNA-dependent RNA polymerases, putative (RPC19)	G2&M	1.2

Tb927.1.690	hypothetical protein, conserved	G2&M	1.2
Tb927.10.3160	hypothetical protein, conserved	G2&M	1.2
Tb927.7.4630	hypothetical protein, conserved	G2&M	1.2
Tb927.8.7540	hypothetical protein	G1/early S	2.8
Tb927.7.3410	centrin, putative	G1/early S	2.3
Tb927.2.5000	hypothetical protein, conserved	G1/early S	2.0
Tb927.10.12440	kinesin, putative	G1/early S	1.9
Tb927.9.15230	hypothetical protein, conserved	G1/early S	1.9
Tb927.11.4510	hypothetical protein, conserved	G1/early S	1.8
Tb927.10.11340	hypothetical protein, conserved	G1/early S	1.8
Tb927.7.6900	double-strand-break repair protein rad21 homolog, putative	G1/early S	1.8
Tb927.10.7730	hypothetical protein, conserved	G1/early S	1.8
Tb927.8.2120	hypothetical protein, conserved	G1/early S	1.7
Tb927.11.8780	hypothetical protein, conserved	G1/early S	1.7
Tb927.7.7360	cdc2-related kinase 2,cell division control protein 2 homolog 2 (CRK2)	G1/early S	1.7
Tb927.8.6000	fatty acid desaturase, putative	G1/early S	1.7
Tb927.6.3990	hypothetical protein, conserved	G1/early S	1.6
Tb927.4.1910	hypothetical protein, conserved	G1/early S	1.6
Tb927.9.5710	general transcription factor IIB (tf2b)	G1/early S	1.6
Tb927.10.6010	hypothetical protein, conserved	G1/early S	1.6
Tb927.2.4900	hypothetical protein, conserved	G1/early S	1.6
Tb927.7.1610	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase, putative	G1/early S	1.6
Tb927.11.6420	hypothetical protein, conserved	G1/early S	1.5
Tb11.v5.0452	GTPase activating protein, putative	G1/early S	1.5
Tb927.8.6530	hypothetical protein, conserved	G1/early S	1.5
Tb927.8.7190	hypothetical protein	G1/early S	1.5
Tb927.10.11810	hypothetical protein, conserved	G1/early S	1.5
Tb927.10.4280	hypothetical protein, conserved	G1/early S	1.5
Tb927.10.5070	ribonuclease H, putative	G1/early S	1.5
Tb927.7.5330	hypothetical protein, conserved	G1/early S	1.5

Tb927.7.1410	hypothetical protein, conserved	G1/early S	1.5
Tb927.10.11990	RNA-binding protein	G1/early S	1.5
Tb927.11.14490	RNA polymerase subunit, putative (RPB7)	G1/early S	1.5
Tb927.3.4000	clathrin coat assembly protein ap19, putative	G1/early S	1.5
Tb927.1.1130	glycerol-3-phosphate dehydrogenase (FAD-dependent), putative	G1/early S	1.5
Tb927.5.1620	hypothetical protein, conserved	G1/early S	1.5
Tb927.9.10470	hypothetical protein, conserved	G1/early S	1.5
Tb927.11.5130	hypothetical protein, conserved	G1/early S	1.4
Tb927.11.13740	receptor-type adenylate cyclase GRESAG 4, putative	G1/early S	1.4
Tb927.5.1600	hypothetical protein, conserved	G1/early S	1.4
Tb927.10.910	DNA repair and recombination helicase protein PIF6 (TbPIF6)	G1/early S	1.4
Tb927.10.2680	hypothetical protein, conserved	G1/early S	1.4
Tb927.3.4500	fumarate hydratase, class I (FHc)	G1/early S	1.4
Tb927.8.4920	hypothetical protein, conserved	G1/early S	1.4
Tb927.9.13540	hypothetical protein, conserved	G1/early S	1.4
Tb927.7.4440	hypothetical protein, conserved	G1/early S	1.4
Tb927.7.6790	hypothetical protein, conserved	G1/early S	1.4
Tb927.4.4400	hypothetical protein, conserved	G1/early S	1.4
Tb927.7.5320	hypothetical protein, conserved	G1/early S	1.4
Tb927.6.2900	dendritic cell-derived IFNG-induced protein,SAM domain and HD domain-containing protein 1	G1/early S	1.4
Tb927.3.2820	TFIIF-stimulated CTD phosphatase, putative	G1/early S	1.4
Tb927.4.5120	hypothetical protein, conserved	G1/early S	1.4
Tb11.1810b	retrotransposon hot spot protein (RHS, pseudogene), putative	G1/early S	1.4
Tb927.7.7500	thymine-7-hydroxylase, putative (TLP7)	G1/early S	1.4
Tb927.10.2660	hypothetical protein, conserved	G1/early S	1.4
Tb927.8.4240	hypothetical protein, conserved	G1/early S	1.4
Tb927.11.7850	hypothetical protein, conserved	G1/early S	1.4
Tb927.10.11840	hypothetical protein	G1/early S	1.4
Tb927.8.3530	glycerol-3-phosphate dehydrogenase [NAD+], glycosomal	G1/early S	1.4

Tb927.11.7180	hypothetical protein, conserved	G1/early S	1.4
Tb927.7.4390	threonine synthase, putative	G1/early S	1.4
Tb927.11.15240	Ras-related protein Rab2 (RAB2)	G1/early S	1.4
Tb927.9.5240	mismatch repair protein MSH3, putative,mismatch repair protein (MSH3)	G1/early S	1.4
Tb927.11.12500	hypothetical protein, conserved	G1/early S	1.4
Tb927.6.4180	hypothetical protein, conserved	G1/early S	1.4
Tb927.11.14900	coatomer epsilon subunit, putative,cytosolic coat protein, putative	G1/early S	1.4
Tb927.3.1970	hypothetical protein, conserved	G1/early S	1.4
Tb927.11.11810	hypothetical protein, conserved	G1/early S	1.4
Tb927.2.1820	protein kinase, putative	G1/early S	1.4
Tb927.11.3320	ras-like small GTPase, putative (TbGTR)	G1/early S	1.4
Tb927.7.1190	regulator of chromosome condensation, putative	G1/early S	1.4
Tb927.11.3540	hypothetical protein, conserved	G1/early S	1.4
Tb927.8.840	hypothetical protein, conserved	G1/early S	1.4
Tb927.10.11020	DNA mismatch repair protein MSH2, putative,MutS protein homolog 2, putative (MSH2)	G1/early S	1.4
Tb927.11.2260	eukaryotic translation initiation factor 4e, putative (eIF4E)	G1/early S	1.4
Tb927.11.8920	hypothetical protein, conserved	G1/early S	1.4
Tb927.5.2690	inositol-1(or 4)-monophosphatase 1, putative (IMPase 1)	G1/early S	1.4
Tb927.10.2010	hexokinase (HK1)	G1/early S	1.4
Tb927.3.3270	ATP-dependent phosphofructokinase (PFK)	G1/early S	1.4
Tb927.10.7090	alternative oxidase (AOX)	G1/early S	1.4
Tb927.8.1780	protein kinase, putative	G1/early S	1.3
Tb927.6.3920	hypothetical protein, conserved	G1/early S	1.3
Tb927.11.5520	triosephosphate isomerase (TIM)	G1/early S	1.3
Tb927.7.2980	hypothetical protein, conserved	G1/early S	1.3
Tb927.11.12280	hypothetical protein, conserved	G1/early S	1.3
Tb927.8.1420	acyl-CoA dehydrogenase, mitochondrial precursor, putative	G1/early S	1.3
Tb927.10.1800	hypothetical protein, conserved	G1/early S	1.3
Tb927.10.13020	Vacuolar protein sorting-associated protein 16 homolog, putative	G1/early S	1.3

Tb927.7.4430	hypothetical protein, conserved	G1/early S	1.3
Tb927.4.4950	hypothetical protein, conserved	G1/early S	1.3
Tb927.6.3690	pre-mRNA cleavage complex II Clp1-like, conserved	G1/early S	1.3
Tb927.3.4710	hypothetical protein, conserved	G1/early S	1.3
Tb927.4.2400	hypothetical protein, conserved	G1/early S	1.3
Tb927.10.2000	actin-like protein, putative	G1/early S	1.3
Tb927.10.6710	hypothetical protein, conserved	G1/early S	1.3
Tb927.8.2380	ABC transporter, putative	G1/early S	1.3
Tb927.6.3220	hypothetical protein, conserved	G1/early S	1.3
Tb927.7.1790	Adenine phosphoribosyltransferase, putative	G1/early S	1.3
Tb927.7.7470	receptor-type adenylate cyclase GRESAG 4, putative	G1/early S	1.3
Tb927.8.5460	flagellar calcium-binding protein,44 kDa calflagin,44 kDa calcimedin (Tb-44)	G1/early S	1.3
Tb927.10.4040	3-keto-dihydrosphingosine reductase	G1/early S	1.3
Tb927.1.4420	ABC transporter, putative	G1/early S	1.3
Tb927.11.11520	glycosomal membrane protein (PEX11)	G1/early S	1.3
Tb927.1.2260	calpain-like protein fragment, putative	G1/early S	1.3
Tb927.8.560	hypothetical protein, conserved	G1/early S	1.3
Tb927.6.1910	hypothetical protein, conserved	G1/early S	1.3
Tb927.9.4200	fatty acyl CoA synthetase 2 (ACS2)	G1/early S	1.3
Tb927.11.5580	hypothetical protein, conserved	G1/early S	1.3
Tb927.10.5780	amino acid tansporter, putative	G1/early S	1.3
Tb927.8.3540	hypothetical protein, conserved	G1/early S	1.3
Tb927.11.14820	hypothetical protein, conserved	G1/early S	1.3
Tb927.6.2760	hypothetical protein, conserved	G1/early S	1.2
Tb927.11.4840	hypothetical protein, conserved	G1/early S	1.2
Tb927.10.640	arginine N-methyltransferase, type II (PRMT5)	early G1	2.2
Tb927.4.3920	hypothetical protein, conserved	early G1	2.0
Tb927.7.2440	pyrroline-5-carboxylate reductase, putative (P5CR)	early G1	2.0
Tb927.11.16790	mitogen-activated protein kinase (ECK1)	early G1	2.0
Tb927.6.2540	DREV methyltransferase, putative	early G1	1.8
Tb927.4.3800	hypothetical protein, conserved	early G1	1.8

Tb927.1.190	retrotransposon hot spot protein (RHS, pseudogene), putative	early G1	1.7
Tb927.10.14280	mitochondrial carrier protein (MCP20)	early G1	1.7
Tb927.1.1390	hypothetical protein, conserved	early G1	1.7
Tb927.10.3280	60S ribosomal proteins L38, putative	early G1	1.7
Tb927.1.1640	hypothetical protein, conserved	early G1	1.7
Tb927.6.4420	hypothetical protein, conserved	early G1	1.7
Tb927.11.3620	nucleobase/nucleoside transporter 8.1 (NT8.1)	early G1	1.7
Tb927.11.5590	hypothetical protein, conserved	early G1	1.7
Tb927.2.5850	small nuclear ribonucleoprotein SmD2 (Sm-D2)	early G1	1.7
Tb927.8.6490	protein kinase, putative	early G1	1.7
Tb927.2.2670	histone H4 variant	early G1	1.7
Tb927.7.4170	fatty acid elongase, putative	early G1	1.7
Tb927.11.9620.1	hypothetical protein, conserved	early G1	1.7
Tb927.4.3290	hypothetical protein, conserved	early G1	1.7
Tb927.10.14360	U2 small nuclear ribonucleoprotein 40K (U2_40K)	early G1	1.7
Tb927.11.6040	hypothetical protein, conserved	early G1	1.6
Tb927.7.3520	mitochondrial pyruvate carrier protein 2, putative	early G1	1.6
Tb927.11.9770	hypothetical protein, conserved	early G1	1.6
Tb927.6.2940	phosphopantothenoylcysteine decarboxylase, putative (PPCDC)	early G1	1.6
Tb927.10.8410	hypothetical protein, conserved	early G1	1.6
Tb927.11.16930	ABC transporter of the mitochondrion, putative,Iron-sulfur clusters transporter, putative (ABCB7)	early G1	1.6
Tb927.10.7740	protein transport protein SEC23 (SEC23.2)	early G1	1.6
Tb927.3.3480	U2 small nuclear ribonucleoprotein B, putative (RBP36 U2SNRNPB)	early G1	1.6
Tb927.7.7170	CYC2-like cyclin, putative (CYC4)	early G1	1.6
Tb927.5.930	NADH-dependent fumarate reductase (FRDg)	early G1	1.6
Tb927.11.7560	predicted WD40 repeat protein	early G1	1.6
Tb927.10.15940	cation transporter, putative	early G1	1.6
Tb927.11.7780	hypothetical protein, conserved	early G1	1.6
Tb927.2.4930	esterase, putative	early G1	1.6
Tb927.7.4050	hypothetical protein, conserved	early G1	1.6
Tb927.8.7250	hypothetical protein, conserved	early G1	1.6
Tb927.4.3560	protein phoshatase 1, putative	early G1	1.6
Tb927.2.4540	Small nuclear ribonucleoprotein-associated protein B (snRNP-B) (Sm protein B) (Sm-B) (SmB), putative (TbSmB)	early G1	1.5
Tb927.7.6930	ATPase, putative	early G1	1.5
Tb927.8.3270	hypothetical protein	early G1	1.5
Tb927.11.4760	hypothetical protein	early G1	1.5
Tb927.3.1840	3-oxo-5-alpha-steroid 4-dehydrogenase, putative,polyprenol reductase	early G1	1.5
Tb927.4.3500	hypothetical protein, conserved	early G1	1.5
Tb927.4.5060	hypothetical protein	early G1	1.5
Tb927.9.12630	glycerol kinase, glycosomal (glk1)	early G1	1.5
Tb11.v5.0553	ubiquitin hydrolase, putative	early G1	1.4

Tb927.3.2280	vacuolar sorting protein 33 , putative	early G1	1.4
Tb927.3.3120	hypothetical protein, conserved	early G1	1.4
Tb927.11.11750	membrane-bound acid phosphatase, putative	early G1	1.4
Tb927.4.1370	hypothetical protein, conserved	early G1	1.3
Tb927.8.4770	small GTP-binding protein Rab18 (TbRAB18)	early G1	1.3
Tb927.11.1030	kinetoplastid kinetochore protein 7 (kkt7)	high G2&M	6.0
Tb927.1.2600	pumilio/PUF RNA binding protein 9 (PUF9)	high G2&M	5.3
Tb927.9.6110	hypothetical protein, conserved	high G2&M	4.7
Tb927.9.1340	Myosin-like protein,Nucleoporin (TbMlp-2)	high G2&M	4.7
Tb927.6.4820	chromosomal passenger protein (CPC1)	high G2&M	4.7
Tb927.10.6330	kinetoplastid kinetochore protein 1 (kkt1)	high G2&M	4.5
Tb927.5.4520	hypothetical protein, conserved	high G2&M	4.0
Tb927.11.10940	hypothetical protein, conserved	high G2&M	3.8
Tb927.11.12410	kinetoplastid kinetochore protein 10 (kkt10)	high G2&M	3.6
Tb927.11.8370	hypothetical protein, conserved	high G2&M	3.6
Tb927.9.3650	kinesin (Kif-13-1)	high G2&M	3.4
Tb927.11.8220	aurora B kinase (AUK1)	high G2&M	3.3
Tb927.11.2880	kinesin, putative (KIN-A)	high G2&M	3.2
Tb927.3.2020	kinesin, putative	high G2&M	3.2
Tb927.6.1210	kinetoplastid kinetochore protein 6 (kkt6)	high G2&M	3.2
Tb927.10.870	hypothetical protein, conserved	high G2&M	3.1
Tb927.10.10990	predicted WD40 repeat protein	high G2&M	2.7
Tb927.9.13920	kinetoplastid membrane protein KMP-11 (KMP-11)	high G2&M	2.2
Tb927.7.6770	hypothetical protein, conserved	high G2&M	1.9
Tb927.11.1130	calpain-like cysteine peptidase, putative,antigen, putative	high G2&M	1.9
Tb927.6.2850	ESAG8-associated protein, putative (PIE8)	high S	4.3
Tb927.3.4960	kinesin, putative	high S	4.2
Tb927.11.14750	hypothetical protein, conserved	high S	4.1
Tb927.10.8330	S. cerevisiae PSP1 homologue, putative	high S	3.5
Tb927.11.1340	hypothetical protein, conserved	high S	3.4
Tb927.6.4470	prozyme,S-adenosylmethionine decarboxylaseregulator	high S	2.8
Tb927.11.13110	hypothetical protein, conserved	high S	2.7
Tb927.10.3970	hypothetical protein, conserved	high S	2.6
Tb927.4.530	hypothetical protein, conserved	high S	2.6

Tb927.10.880	thymidine kinase, putative	high S	2.5
Tb927.11.5410	hypothetical protein, conserved	high S	2.3
Tb927.1.4480	hypothetical protein, conserved	high S	2.2
Tb927.1.4010	primase 2	high S	2.1
Tb927.8.1900	hypothetical protein, conserved	high S	2.1
Tb927.8.2550	mitochondrial DNA primase (PRI1)	high S	2.0
Tb927.7.4370	hypothetical protein, conserved	high S	1.8
Tb927.9.11580	Gim5A protein,glycosomal membrane protein (gim5A)	high early G1	3.4
Tb927.9.11600	unspecified product (gim5B)	high early G1	3.2
Tb927.5.320	receptor-type adenylate cyclase GRESAG 4, putative,adenylyl cyclase, putative	high early G1	3.1
Tb927.6.3240	hypothetical protein, conserved	high early G1	3.1
Tb927.10.13230	hypothetical protein, conserved	high early G1	3.0
Tb927.9.2680	hypothetical protein, unlikely	high early G1	2.9
Tb927.6.820	pumilio RNA binding protein, putative (PUF4)	high early G1	2.9
Tb927.3.960	protein transport protein Sec61 gamma subunit, putative	high early G1	2.8
Tb927.7.2620	hypothetical protein, conserved	high early G1	2.5
Tb927.8.730	nucleolar RNA-binding protein, putative	high early G1	2.4
Tb927.1.570	Histone-lysine N-methyltransferase, H3 lysine-76 specific,Histone H3-K76 methyltransferase (DOT1B)	high early G1	2.3
Tb927.11.5940	receptor-type adenylate cyclase GRESAG 4, putative	high early G1	2.1
Tb927.10.12800	Zinc finger CCCH domain-containing protein 38 (ZC3H38)	high early G1	2.0
Tb927.10.15540	predicted VPS53-like domain protein	high G1	5.6
Tb927.11.950	RNA-binding protein, putative (DRBD14)	high G1	4.5
Tb927.11.14830	hypothetical protein, conserved	high early S	3.5
Tb927.11.6890	DNA repair and recombination helicase protein PIF1 (PIF1)	high early S	3.5
Tb927.11.4180	hypothetical protein, conserved	high early S	3.3
Tb927.6.5100	serine/threonine-protein kinase, putative	high early S	2.4
Tb927.8.710	RNA-binding protein, putative (DRBD17)	early G1/late G2&M	2.8

Supplementary Table 5: List of proteins detected as cell-cycle regulated in both *T. brucei* proteomic and transcriptomic *(Archer et al., 2011)* analysis.

Gene ID	Product Description
Tb927.1.1270	homocysteine S-methyltransferase, putative
Tb927.1.2600	pumilio/PUF RNA binding protein 9 (PUF9)
Tb927.1.3450	hypothetical protein, conserved
Tb927.1.4480	WD repeat and HMG-box DNA-binding protein, putative
Tb927.2.2670	histone H4 variant
Tb927.2.5000	hypothetical protein, conserved
Tb927.3.2020	kinesin, putative
Tb927.3.4500	Fumarate hydratase class I, cytosolic (FHc)
Tb927.3.4710	hypothetical protein, conserved
Tb927.3.4960	kinesin, putative
Tb927.4.1910	hypothetical protein, conserved
Tb927.4.2400	hypothetical protein, conserved
Tb927.4.4370	hypothetical protein, conserved
Tb927.4.5120	hypothetical protein, conserved
Tb927.5.4520	hypothetical protein, conserved
Tb927.6.570	conserved protein
Tb927.6.2850	ESAG8-associated protein, putative (PIE8)
Tb927.6.2900	dendritic cell-derived IFNG-induced protein
Tb927.6.3220	hypothetical protein, conserved
Tb927.6.3240	hypothetical protein, conserved
Tb927.6.4820	chromosomal passenger protein (CPC1)
Tb927.6.5130	60S acidic ribosomal protein P2, putative
Tb927.7.1410	hypothetical protein, conserved
Tb927.7.2290	hypothetical protein, conserved
Tb927.7.3410	centrin-4 (Centrin4)
Tb927.7.4390	threonine synthase, putative
Tb927.7.5320	hypothetical protein, conserved
Tb927.7.5730	hypothetical protein, conserved
Tb927.7.6900	double-strand-break repair protein rad21 homolog, putative
Tb927.7.7470	procyclic-enriched flagellar receptor adenylate cyclase 3 (ACP3)
Tb927.8.710	RNA-binding protein, putative (DRBD17)
Tb927.8.960	hypothetical protein, conserved
Tb927.8.1900	hypothetical protein, conserved
Tb927.8.2550	mitochondrial DNA primase (PRI1)
Tb927.8.3540	hypothetical protein, conserved
Tb927.8.3560	DNA repair and recombination helicase protein PIF5 (PIF5)
Tb927.8.4240	hypothetical protein, conserved
Tb927.8.7340	trans-sialidase, putative

Tb927.8.7540	hypothetical protein
Tb927.9.1340	Myosin-like protein (TbMlp-2)
Tb927.9.5240	mismatch repair protein MSH3, putative (MSH3)
Tb927.9.6110	hypothetical protein, conserved
Tb927.9.13140	hypothetical protein, conserved
Tb927.9.14290	hypothetical protein, conserved
Tb927.9.15050	trypanin-related protein, putative
Tb927.9.15230	hypothetical protein, conserved
Tb927.10.870	hypothetical protein, conserved
Tb927.10.880	thymidine kinase, putative
Tb927.10.890	kinesin, putative
Tb927.10.2660	hypothetical protein, conserved
Tb927.10.3970	hypothetical protein, conserved
Tb927.10.4990	cdc2-related kinase 3, putative (CRK3)
Tb927.10.5070	ribonuclease H, putative
Tb927.10.6010	STAG domain containing protein, putative
Tb927.10.6330	kinetoplastid kinetochore protein 1 (kkt1)
Tb927.10.7730	Domain of unknown function (DUF4496), putative
Tb927.10.8330	S. cerevisiae PSP1 homologue, putative
Tb927.10.9870	Protein of unknown function (DUF2870), putative
Tb927.10.11020	DNA mismatch repair protein MSH2, putative (MSH2)
Tb927.10.11340	hypothetical protein, conserved
Tb927.10.11990	RNA-binding protein
Tb927.10.12920	hypothetical protein, conserved
Tb927.10.13100	hypothetical protein, conserved
Tb927.10.15400	kinesin, putative
Tb927.11.950	RNA-binding protein, putative (DRBD14)
Tb927.11.1030	kinetoplastid kinetochore protein 7 (kkt7)
Tb927.11.1130	calpain-like cysteine peptidase, putative
Tb927.11.1340	hypothetical protein, conserved
Tb927.11.4180	PSP1 C-terminal conserved region, putative
Tb927.11.4510	Sister chromatid cohesion C-terminus, putative
Tb927.11.4840	hypothetical protein, conserved
Tb927.11.5410	hypothetical protein, conserved
Tb927.11.6040	Nodulin-like, putative
Tb927.11.6560	hypothetical protein, conserved
Tb927.11.7600	hypothetical protein, conserved
Tb927.11.7930	RNA polymerase B subunit RPB8, putative (RPB8)
Tb927.11.8220	aurora B kinase (AUK1)
Tb927.11.8920	hypothetical protein, conserved
Tb927.11.10520	kinetoplastid kinetochore protein 2 (kkt2)
Tb927.11.10940	hypothetical protein, conserved

Tb927.11.13110	hypothetical protein, conserved
Tb927.11.14750	PSP1 C-terminal conserved region, putative
Tb927.11.14830	hypothetical protein, conserved
Tb927.11.15800	hypothetical protein, conserved
Tb927.11.16390	cyclin dependent kinases regulatory subunit, putative (CKS1)

Supplementary Table 6: List of proteins identified in both transcriptomic and proteomic analysis of the cell-cycle, and only classified as regulated from proteomic data.

Gene ID	Product Description
Tb927.1.1130	glycerol-3-phosphate dehydrogenase (FAD-dependent), putative
Tb927.1.1210	conserved protein, unknown function
Tb927.1.1390	hypothetical protein, conserved
Tb927.1.1630	transcription elongation factor 1 homolog (ELOF1)
Tb927.1.1640	hypothetical protein, conserved
Tb927.1.190	retrotransposon hot spot protein (RHS, pseudogene), putative
Tb927.1.2260	calpain-like protein fragment, putative
Tb927.1.4010	primase 2
Tb927.1.4300	chaperone protein DNAj, putative
Tb927.1.4420	ABC transporter, putative
Tb927.1.4700	hypothetical protein, conserved
Tb927.1.570	Histone-lysine N-methyltransferase, H3 lysine-76 specific (DOT1B)
Tb927.1.690	hypothetical protein, conserved
Tb927.1.990	Enkuring domain-containig protein
Tb927.10.10000	hypothetical protein, conserved
Tb927.10.10330	Anaphase-promoting complex, cyclosome, subunit 3/TPR repeat, putative
Tb927.10.10710	hypothetical protein, conserved
Tb927.10.10780	hypothetical protein, conserved
Tb927.10.10990	predicted WD40 repeat protein
Tb927.10.11270	RNA-binding protein, putative (RBP23)
Tb927.10.11580	predicted WD40 repeat protein
Tb927.10.11730	zinc finger protein family member, putative (ZC3H33)
Tb927.10.11810	hypothetical protein, conserved
Tb927.10.11900	thioredoxin, putative
Tb927.10.1200	Programmed cell death protein 2, C-terminal putative domain containing protein, putative
Tb927.10.12220	hypothetical protein, conserved
Tb927.10.12440	kinesin, putative
Tb927.10.13020	Vacuolar protein sorting-associated protein 16 homolog, putative
Tb927.10.13230	hypothetical protein, conserved
Tb927.10.14280	mitochondrial carrier protein (MCP20)

Tb927.10.14330	Utp14 protein, putative
Tb927.10.14360	U2 small nuclear ribonucleoprotein 40K (U2_40K)
Tb927.10.14460	leucine-rich repeat protein (LRRP), putative
Tb927.10.1450	plectin, putative
Tb927.10.14560	TPR repeat, putative
Tb927.10.15540	predicted VPS53-like domain protein
Tb927.10.15940	cation transporter, putative
Tb927.10.1800	methyltransferase domain containing protein, putative
Tb927.10.2000	actin-like protein, putative
Tb927.10.2680	pyridine nucleotide-disulphide oxidoreductase, putative
Tb927.10.3160	hypothetical protein, conserved
Tb927.10.3280	60S ribosomal proteins L38, putative
Tb927.10.3300	Domain of unknown function (DUF4206), putative
Tb927.10.3380	60S acidic ribosomal protein P2, putative
Tb927.10.3470	Ankyrin repeats (3 copies), putative
Tb927.10.4040	3-keto-dihydrosphingosine reductase
Tb927.10.4050	serine palmitoyltransferase, putative
Tb927.10.4280	ubiquinol-cytochrome c reductase complex 14kD subunit, putative
Tb927.10.4620	peptidyl-prolyl cis-trans isomerase, putative (PPIase)
Tb927.10.4750	hypothetical protein, conserved
Tb927.10.4930	protein phosphatase 2C, putative
Tb927.10.5270	mitogen-activated protein kinase kinase 5, putative (MKK5)
Tb927.10.5750	Extradiol ring-cleavage dioxygenase, putative
Tb927.10.5780	amino acid tansporter, putative
Tb927.10.5860	hypothetical protein, conserved
Tb927.10.6060	universal minicircle sequence binding protein 2 (UMSBP2)
Tb927.10.6070	universal minicircle sequence binding protein 1 (UMSBP1)
Tb927.10.640	arginine N-methyltransferase, type II (PRMT5)
Tb927.10.6410	DNA mismatch repair protein MSH6, putative (MSH6)
Tb927.10.6540	kinetoplastid-specific phospho-protein phosphatase, putative
Tb927.10.6710	hypothetical protein, conserved
Tb927.10.6960	hypothetical protein, conserved
Tb927.10.7090	Alternative oxidase, mitochondrial (AOX)
Tb927.10.7430	GTP-binding protein, putative
Tb927.10.7620	mitochondrial ATP-dependent zinc metallopeptidase, putative
Tb927.10.7740	protein transport protein SEC23 (SEC23.2)
Tb927.10.7780	mitogen-activated protein kinase 1, putative
Tb927.10.8240	hypothetical protein, conserved
Tb927.10.8410	hypothetical protein, conserved
Tb927.10.8580	ADP-ribosylation factor-like protein
Tb927.10.8610	hypothetical protein, conserved
Tb927.10.8710	centrin, putative

Tb927.10.9080	pteridine transporter, putative
Tb927.10.910	DNA repair and recombination helicase protein PIF6 (TbPIF6)
Tb927.10.9160	hypothetical protein, conserved
Tb927.10.9230	beta adaptin, putative
Tb927.11.10100	hypothetical protein, conserved
Tb927.11.11240	hypothetical protein, conserved
Tb927.11.11520	glycosomal membrane protein (PEX11)
Tb927.11.11810	hypothetical protein, conserved
Tb927.11.12280	hypothetical protein, conserved
Tb927.11.1240	hypothetical protein, conserved
Tb927.11.12410	kinetoplastid kinetochore protein 10 (kkt10)
Tb927.11.12500	hypothetical protein, conserved
Tb927.11.12520	hypothetical protein, conserved
Tb927.11.12930	DEAD-box helicase, putative
Tb927.11.1300	UBA/TS-N domain containing protein, putative
Tb927.11.13100	hypothetical protein, conserved
Tb927.11.13150	UEV domain/Zinc finger, C3HC4 type (RING finger), putative
Tb927.11.13520	hypothetical protein, conserved
Tb927.11.13740	procyclic-enriched flagellar receptor adenylate cyclase 5 (ACP5)
Tb927.11.13980	Outer row dynein-assembly protein 7 (ODA7)
Tb927.11.14410	Ankyrin repeats (3 copies), putative
Tb927.11.14460	ADP-ribosylation factor GTPase activating protein 1, putative
Tb927.11.14490	RNA polymerase subunit, putative (RPB7)
Tb927.11.14710	rRNA processing, putative
Tb927.11.14820	hypothetical protein, conserved
Tb927.11.14900	coatomer subunit epsilon (COPE)
Tb927.11.14950	zinc finger protein 2 (ZFP2)
Tb927.11.15240	Ras-related protein RAB2B, putative (RAB2B)
Tb927.11.15780	hypothetical protein, conserved
Tb927.11.16380	TPR repeat, putative
Tb927.11.16750	hypothetical protein, conserved
Tb927.11.16820	hypothetical protein, conserved
Tb927.11.16930	ABC transporter of the mitochondrion, putative (ABCB7)
Tb927.11.2020	hypothetical protein, conserved
Tb927.11.2260	eukaryotic translation initiation factor 4E, putative (eIF4E)
Tb927.11.2880	kinesin, putative (KIN-A)
Tb927.11.3080	vesicular protein trafficking mediator, putative
Tb927.11.3090	LisH domain-containing protein FOPNL, putative (fopnl)
Tb927.11.3320	ras-like small GTPase, putative (TbGTR)
Tb927.11.3540	hypothetical protein, conserved
Tb927.11.3930	hypothetical protein, conserved
Tb927.11.4760	hypothetical protein

Tb927.11.4770	retrotransposon hot spot protein (RHS, pseudogene), putative
Tb927.11.4970	predicted WD40 repeat protein
Tb927.11.5130	hypothetical protein, conserved
Tb927.11.5260	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein, putative
Tb927.11.5520	triosephosphate isomerase (TIM)
Tb927.11.5580	hypothetical protein, conserved
Tb927.11.5590	hypothetical protein, conserved
Tb927.11.5850	RNA-binding protein 38, putative (RBP38)
Tb927.11.5940	receptor-type adenylate cyclase GRESAG 4, putative
Tb927.11.6340	PUB domain containing protein, putative
Tb927.11.640	pentatricopeptide repeat domain containing protein, putative
Tb927.11.6420	hypothetical protein, conserved
Tb927.11.6440	hypothetical protein, conserved
Tb927.11.650	hypothetical protein, conserved
Tb927.11.6890	DNA repair and recombination helicase protein PIF1 (PIF1)
Tb927.11.7050	PA domain containing protein, putative (ESP7)
Tb927.11.7110	myotubularin-associated protein, putative
Tb927.11.7180	MORN repeat, putative
Tb927.11.7280	DNA-direcetd RNA polymerase II, subunit 9, putative (RPB9)
Tb927.11.7560	predicted WD40 repeat protein
Tb927.11.7780	Mitochondrial import receptor subunit ATOM46 (ATOM46)
Tb927.11.7850	hypothetical protein, conserved
Tb927.11.8100	Ankyrin repeats (3 copies), putative
Tb927.11.8370	hypothetical protein, conserved
Tb927.11.8780	hypothetical protein, conserved
Tb927.11.8890	DNA-dependent RNA polymerases, putative (RPC19)
Tb927.11.9750	Protein of unknown function (DUF498/DUF598), putative
Tb927.11.9770	hypothetical protein, conserved
Tb927.11.9860	EF-hand domain pair, putative
Tb927.2.4140	hypothetical protein, conserved
Tb927.2.4540	Small nuclear ribonucleoprotein-associated protein B (snRNP-B) (Sm protein B) (Sm-B) (SmB), putative (TbSmB)
Tb927.2.4930	esterase, putative
Tb927.2.5850	small nuclear ribonucleoprotein SmD2 (Sm-D2)
Tb927.2.5950	hypothetical protein, conserved
Tb927.2.6080	hypothetical protein, conserved
Tb927.3.1840	3-oxo-5-alpha-steroid 4-dehydrogenase, putative
Tb927.3.1890	cytochrome c oxidase assembly protein, putative
Tb927.3.1970	DNA / pantothenate metabolism flavoprotein, putative
Tb927.3.2040	kinesin, putative
Tb927.3.2140	transcription activator, putative
Tb927.3.2280	vacuolar sorting protein 33 , putative

Tb927.3.2330	kinetoplastid kinetochore protein 17 (kkt17)
Tb927.3.2370	hypothetical protein, conserved
Tb927.3.2820	TFIIF-stimulated CTD phosphatase, putative
Tb927.3.3120	hypothetical protein, conserved
Tb927.3.3240	hypothetical protein, conserved
Tb927.3.3270	ATP-dependent 6-phosphofructokinase, glycosomal (PFK)
Tb927.3.3480	U2 small nuclear ribonucleoprotein B, putative (RBP36 U2SNRNPB)
Tb927.3.3950	hypothetical protein, conserved
Tb927.3.4000	clathrin coat assembly protein ap19, putative
Tb927.3.4230	subtilisin-like serine peptidase (CMF33)
Tb927.3.4270	hypothetical protein, conserved
Tb927.3.4950	hypothetical protein, conserved
Tb927.3.860	Acyl carrier protein, mitochondrial (ACP)
Tb927.3.960	protein transport protein Sec61 gamma subunit, putative
Tb927.4.1150	hypothetical protein, conserved
Tb927.4.1170	ankyrin, putative
Tb927.4.1370	hypothetical protein, conserved
Tb927.4.1750	hypothetical protein, conserved
Tb927.4.2190	GTP-binding protein, putative
Tb927.4.2910	hypothetical protein, conserved
Tb927.4.3130	Spef1
Tb927.4.3350	N2227-like protein, putative
Tb927.4.3430	hypothetical protein, conserved
Tb927.4.3500	Amastin surface glycoprotein, putative
Tb927.4.3560	protein phoshatase 1, putative
Tb927.4.3800	hypothetical protein, conserved
Tb927.4.3920	CRAL/TRIO domain containing protein, putative
Tb927.4.4440	adenylyl cyclase (GRESAG 4.4)
Tb927.4.4610	hypothetical protein, conserved
Tb927.4.4620	cytochrome oxidase subunit VIII (COXVIII)
Tb927.4.4970	myosin heavy chain kinase A, putative
Tb927.4.5060	hypothetical protein
Tb927.4.5180	tousled-like kinase I (TLK1)
Tb927.4.530	hypothetical protein, conserved
Tb927.4.670	conserved protein, unknown function
Tb927.4.720	hypothetical protein, conserved
Tb927.5.1600	hypothetical protein, conserved
Tb927.5.1620	hypothetical protein, conserved
Tb927.5.1960	hypothetical protein, conserved
Tb927.5.2690	inositol-1(or 4)-monophosphatase 1, putative (IMPase 1)
Tb927.5.320	receptor-type adenylate cyclase GRESAG 4, putative
Tb927.5.3200	hypothetical protein, conserved

Tb927.5.3310	hypothetical protein, conserved
Tb927.5.3630	hypothetical protein, conserved
Tb927.5.4320	zinc finger protein family member, putative (FIP1)
Tb927.5.440	trans-sialidase, putative
Tb927.5.930	NADH-dependent fumarate reductase (FRDg)
Tb927.6.1210	kinetoplastid kinetochore protein 6 (kkt6)
Tb927.6.1540	Yip1 domain containing protein, putative
Tb927.6.1910	hypothetical protein, conserved
Tb927.6.200	receptor-type adenylate cyclase GRESAG 4, putative
Tb927.6.2540	DREV methyltransferase, putative
Tb927.6.2760	hypothetical protein, conserved
Tb927.6.2940	phosphopantothenoylcysteine decarboxylase, putative (PPCDC)
Tb927.6.3290	intraflagellar transport protein 20 (IFT20)
Tb927.6.3690	pre-mRNA cleavage complex II Clp1-like, conserved
Tb927.6.3710	hypothetical protein, conserved
Tb927.6.3760	kinetoplastid kinetochore protein 15 (kkt15)
Tb927.6.3920	hypothetical protein, conserved
Tb927.6.3990	hypothetical protein, conserved
Tb927.6.4180	FUN14 family, putative (POMP32)
Tb927.6.4420	tRNA (Guanine-1)-methyltransferase, putative
Tb927.6.4470	prozyme
Tb927.6.5100	serine/threonine-protein kinase, putative
Tb927.7.1190	regulator of chromosome condensation, putative
Tb927.7.1610	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase, putative
Tb927.7.1790	Adenine phosphoribosyltransferase, putative
Tb927.7.2060	DNA repair protein, putative
Tb927.7.2160	hypothetical protein, conserved
Tb927.7.2440	pyrroline-5-carboxylate reductase, putative (P5CR)
Tb927.7.2620	hypothetical protein, conserved
Tb927.7.2980	Nitroreductase family, putative
Tb927.7.3000	kinesin, putative
Tb927.7.3520	mitochondrial pyruvate carrier protein 2, putative
Tb927.7.3740	hypothetical protein, conserved
Tb927.7.4050	hypothetical protein, conserved
Tb927.7.4170	fatty acid elongase, putative
Tb927.7.4370	hypothetical protein, conserved
Tb927.7.4430	hypothetical protein, conserved
Tb927.7.4440	NAD dependent epimerase/dehydratase family, putative
Tb927.7.4610	hypothetical protein, conserved
Tb927.7.4630	Miro-like protein, putative
Tb927.7.4680	hypothetical protein, conserved
Tb927.7.4780	conserved hypothetical protein, putative

Tb927.7.4850	kinetoplastid kinetochore protein 5 (kkt5)
Tb927.7.5050	hypothetical protein, conserved
Tb927.7.5330	hypothetical protein, conserved
Tb927.7.6770	Acyl CoA binding protein, putative
Tb927.7.6790	hypothetical protein, conserved
Tb927.7.6930	ATPase, putative
Tb927.7.7170	CYC2-like cyclin, putative (CYC4)
Tb927.7.7360	cdc2-related kinase 2 (CRK2)
Tb927.7.7500	thymine-7-hydroxylase, putative (TLP7)
Tb927.8.1130	protein phosphatase with EF-Hand domains, putative
Tb927.8.1420	acyl-CoA dehydrogenase, mitochondrial precursor, putative
Tb927.8.1750	Cytokine-induced anti-apoptosis inhibitor 1, Fe-S biogenesis, putative
Tb927.8.1780	protein kinase, putative
Tb927.8.2120	hypothetical protein, conserved
Tb927.8.2150	hypothetical protein, conserved
Tb927.8.2220	hypothetical protein, conserved
Tb927.8.2300	hypothetical protein, conserved
Tb927.8.2380	ABC transporter, putative
Tb927.8.2590	hypothetical protein, conserved
Tb927.8.2690	SET domain containing protein, putative
Tb927.8.2800	hypothetical protein, conserved
Tb927.8.2870	conserved protein
Tb927.8.2910	mannosyl-oligosaccharide 1,2-alpha-mannosidase IB, putative
Tb927.8.3220	exonuclease, putative
Tb927.8.3270	hypothetical protein
Tb927.8.3460	Ring finger domain containing protein, putative
Tb927.8.3530	glycerol-3-phosphate dehydrogenase [NAD+], glycosomal
Tb927.8.3780	MORN repeat, putative
Tb927.8.4170	RNA-binding protein, putative (MRD1)
Tb927.8.4190	hypothetical protein, conserved
Tb927.8.4640	flagellar protofilament ribbon protein, putative
Tb927.8.4920	hypothetical protein, conserved
Tb927.8.5460	Flagellar calcium-binding protein 44 (Tb-44)
Tb927.8.560	hypothetical protein, conserved
Tb927.8.6000	fatty acid desaturase, putative
Tb927.8.6490	protein kinase, putative
Tb927.8.6530	SET domain containing protein, putative
Tb927.8.6750	translationally controlled tumor protein (TCTP), putative
Tb927.8.7120	squalene synthase, putative
Tb927.8.7190	hypothetical protein
Tb927.8.7250	Maf1 regulator, putative
Tb927.8.730	nucleolar RNA-binding protein, putative

Tb927.8.840	hypothetical protein, conserved
Tb927.9.10250	small nuclear ribonucleoprotein Sm-F (Sm-F)
Tb927.9.10470	hypothetical protein, conserved
Tb927.9.11350	hypothetical protein, conserved
Tb927.9.11580	Gim5A protein (gim5A)
Tb927.9.11600	Gim5B protein (gim5B)
Tb927.9.11820	CRAL/TRIO domain containing protein, putative
Tb927.9.12680	hypothetical protein, conserved
Tb927.9.13070	predicted heat shock factor binding protein
Tb927.9.13350	hypothetical protein, conserved
Tb927.9.13540	hypothetical protein, conserved
Tb927.9.15020	hypothetical protein, conserved
Tb927.9.2760	EB1-like C-terminal motif containing protein, putative
Tb927.9.3650	Kinesin-13 1 (KIN13-1)
Tb927.9.4200	fatty acyl CoA synthetase 2 (ACS2)
Tb927.9.5380	Rab-GTPase-TBC domain containing protein, putative
Tb927.9.5710	general transcription factor IIB (tf2b)
Tb927.9.7470	purine nucleoside transporter (NT10)
Tb927.9.7580	hypothetical protein, conserved
Tb927.9.8260	rhomboid-like protein
Tb927.9.9520	zinc finger protein family member, putative (ZC3H29)

Supplementary Table 7: List of proteins identified in both transcriptomic and proteomic analysis of the cell-cycle, and only classified as regulated from transcriptomic data.

Gene ID	Product Description
Tb927.1.1890	Replication Factor C Subunit 1-related protein
Tb927.1.2670	paralyzed flagella 16 (PF16)
Tb927.1.2730	hypothetical protein, conserved
Tb927.1.3310	hypothetical protein, conserved
Tb927.1.3390	Eukaryotic protein of unknown function (DUF866), putative
Tb927.1.4180	leucine-rich repeat protein (LRRP), putative
Tb927.1.4310	hypothetical protein, conserved
Tb927.1.5030	leucine-rich repeat protein (LRRP), putative
Tb927.1.880	Midasin, putative (MDN1)
Tb927.10.10150	hypothetical protein, conserved
Tb927.10.10370	hypothetical protein, conserved
Tb927.10.1140	epsilon tubulin (TUBE1)
Tb927.10.11780	hypothetical protein, conserved
Tb927.10.11850	WD domain, G-beta repeat, putative
Tb927.10.12030	hypothetical protein, conserved
Tb927.10.12130	Ubiquitin family, putative

Tb927.10.12180	hypothetical protein, conserved
Tb927.10.12860	pre-RNA processing PIH1/Nop17, putative
Tb927.10.13290	ethanolamine phosphotransferase (EPT)
Tb927.10.13430	citrate synthase, putative
Tb927.10.13970	uracil-DNA glycosylase, putative
Tb927.10.14000	aconitase (ACO)
Tb927.10.14010	tubulin cofactor C domain-containing protein RP2 (rp2)
Tb927.10.14390	Histone chaperone Rttp106-like, putative
Tb927.10.14490	hypothetical protein, conserved
Tb927.10.14500	hypothetical protein, conserved
Tb927.10.14550	ATP-dependent DEAD/H RNA helicase, putative
Tb927.10.14710	40S ribosomal protein S2, putative (RPS2)
Tb927.10.15330	CYTH domain containing protein, putative
Tb927.10.15660	hypothetical protein, conserved
Tb927.10.15730	SPRY domain/Ankyrin repeats (3 copies), putative
Tb927.10.1900	DNA topoisomerase IA, putative
Tb927.10.2020	hexokinase (HK2)
Tb927.10.2200	hypothetical protein, conserved
Tb927.10.3110	DNA primase large subunit, putative (pril)
Tb927.10.350	protein kinase, putative
Tb927.10.3920	hypothetical protein, conserved
Tb927.10.4220	hypothetical protein, conserved
Tb927.10.4450	stress-inducible protein STI1-like, putative
Tb927.10.4740	nucleolar RNA-binding protein, putative
Tb927.10.5020	Eukaryotic translation initiation factor 4E type 5 (eif4e5)
Tb927.10.5260	C2 domain containing protein, putative
Tb927.10.5370	40S ribosomal protein S10, putative
Tb927.10.5400	hypothetical protein, conserved
Tb927.10.5540	hypothetical protein, conserved
Tb927.10.5620	fructose-bisphosphate aldolase, glycosomal (ALD)
Tb927.10.5880	Proteophosphoglycan, putative
Tb927.10.5890	Galactose oxidase, central domain containing protein, putative
Tb927.10.6110	hypothetical protein, conserved
Tb927.10.700	hypothetical protein, conserved
Tb927.10.7120	hypothetical protein, conserved
Tb927.10.7230	Flagellar Member 1 (FLAM1)
Tb927.10.7270	hypothetical protein, conserved
Tb927.10.8290	eukaryotic translation initiation factor 3 subunit 8, putative
Tb927.10.830	adenylate kinase, putative
Tb927.10.8820	bilobe protein, conserved
Tb927.10.8930	paraflagellar rod component, putative (PFC18)
Tb927.10.8980	hypothetical protein, conserved

Tb927.10.9000	hypothetical protein, conserved
Tb927.10.9240	Tir chaperone protein (CesT) family, putative
Tb927.10.9570	paraflagellar rod component, putative (PFC14)
Tb927.10.9840	chaperone protein DNAj, putative
Tb927.11.10300	hypothetical protein, conserved
Tb927.11.10540	hypothetical protein, conserved
Tb927.11.10660	hypothetical protein, conserved
Tb927.11.1090	calpain-like protein, putative
Tb927.11.10930	tubulin delta chain (TUBD1)
Tb927.11.11010	hypothetical protein, conserved
Tb927.11.11210	paraflagellar rod component, putative (PFC17)
Tb927.11.11220	dynein heavy chain, putative
Tb927.11.11580	hypothetical protein, conserved
Tb927.11.12030	hypothetical protein, conserved
Tb927.11.12150	flagellar protein essential for flagellar pocket biogenesis (BILBO1)
Tb927.11.1220	hypothetical protein, conserved
Tb927.11.12210	Domain of unknown function (DUF4586), putative
Tb927.11.12840	Domain of unknown function (DUF4486), putative
Tb927.11.13270	hypothetical protein, conserved
Tb927.11.13500	par1
Tb927.11.13650	cytochrome b5, putative (CYB5)
Tb927.11.13700	hypothetical protein, conserved
Tb927.11.13780	profilin
Tb927.11.13930	U3 small nucleolar ribonucleoprotein protein IMP3, putative
Tb927.11.14300	dynein intermediate chain IC70, putative
Tb927.11.14650	hypothetical protein, conserved
Tb927.11.14690	Microtubule-binding protein MIP-T3, putative
Tb927.11.14880	Paraflagellar Rod Proteome Component 9, putative (PFC9)
Tb927.11.14890	DNA polymerase alpha/epsilon subunit B, putative
Tb927.11.15070	hypothetical protein, conserved
Tb927.11.15100	Tb5.20
Tb927.11.15650	poly(A) polymerase, putative
Tb927.11.16120	hypothetical protein, conserved
Tb927.11.16400	kinetoplast-associated protein 3, putative (KAP3)
Tb927.11.16640	Uncharacterised protein family UPF0564, putative
Tb927.11.2060	60S acidic ribosomal protein P0, putative
Tb927.11.20730	glycerophosphoryl diester phosphodiesterase, putative
Tb927.11.2500	metallo-peptidase, Clan MA(E) Family M32
Tb927.11.2540	hypothetical protein, conserved
Tb927.11.2570	hypothetical protein, conserved
Tb927.11.270	mitochondrial carrier protein (MCP10)
Tb927.11.3250	dynein heavy chain, putative

Tb927.11.3260	mitochondrial DNA polymerase I protein D, putative (POLID)
Tb927.11.3290	hypothetical protein, conserved (p166)
Tb927.11.350	RNA-binding protein, putative
Tb927.11.3660	dynein light chain Tctex-type, putative
Tb927.11.3990	hypothetical protein, conserved
Tb927.11.4210	U-box domain containing protein, putative
Tb927.11.4690	mitochondrial DNA polymerase I protein B (POLIB)
Tb927.11.5030	hypothetical protein, conserved
Tb927.11.5570	DNA replication licensing factor MCM5, putative
Tb927.11.5640	hypothetical protein, conserved
Tb927.11.6050	Flagellar-associated PapD-like, putative
Tb927.11.6530	hypothetical protein, conserved
Tb927.11.6710	predicted tetratricopeptide repeat (TPR) protein
Tb927.11.6780	regulator of chromatin condensation, putative
Tb927.11.7130	hypothetical protein, conserved
Tb927.11.7840	ribonucleoside-diphosphate reductase large chain (RNR1)
Tb927.11.7860	DNA replication licensing factor MCM2, putative
Tb927.11.8060	SNARE associated Golgi protein, putative
Tb927.11.8210	RAD50 DNA repair-like protein
Tb927.11.8230	hypothetical protein, conserved
Tb927.11.8390	hypothetical protein, conserved
Tb927.11.8810	hypothetical protein, conserved
Tb927.11.8950	leucine rich repeat (TbLRRP1)
Tb927.11.900	isocitrate dehydrogenase, putative (IDH)
Tb927.11.9100	Domain of unknown function (DUF4586), putative
Tb927.11.9540	hypothetical protein, conserved
Tb927.2.4810	hypothetical protein, conserved
Tb927.2.5270	dynein heavy chain, putative
Tb927.2.5280	Enriched in surface-labeled proteome protein 1 (ESP1)
Tb927.2.5660	adenylate kinase, putative (ADKA)
Tb927.2.5760	Flagellar Member 8 (FLAM8)
Tb927.2.5970	hypothetical protein, conserved
Tb927.2.5980	heat shock protein, putative
Tb927.2.6100	hypothetical protein, conserved
Tb927.3.1040	unspecified product
Tb927.3.1060	cAMP response protein, putative (CARP4)
Tb927.3.1700	diacylglycerol acyltransferase, putative
Tb927.3.1900	hypothetical protein, conserved
Tb927.3.2050	hypothetical protein, conserved
Tb927.3.2310	PACRGA
Tb927.3.2670	hypothetical protein, conserved
Tb927.3.2810	hypothetical protein, conserved

Tb927.3.2880	hypothetical protein, conserved
Tb927.3.2890	radial spoke protein RSP10, putative
Tb927.3.3200	Domain of unknown function (DUF4586), putative
Tb927.3.3440	phytanoyl-CoA dioxygenase (PhyH), putative
Tb927.3.3490	high mobility group protein, putative (TDP1)
Tb927.3.3790	hypothetical protein, conserved
Tb927.3.3940	RNA-binding protein, putative (DRBD11)
Tb927.3.5350	hypothetical protein, conserved
Tb927.3.5370	hypothetical protein, conserved
Tb927.3.5430	hypothetical protein, conserved
Tb927.3.5620	Metallopeptidase family M24/FACT complex subunit (SPT16/CDC68)/Histone chaperone Rttp106-like, putative
Tb927.3.690	protein kinase, putative
Tb927.3.700	hypothetical protein, conserved
Tb927.3.930	dynein heavy chain, putative
Tb927.3.950	2OG-Fe(II) oxygenase superfamily, putative
Tb927.3.990	hypothetical protein, conserved
Tb927.4.1330	DNA topoisomerase IB, large subunit
Tb927.4.2450	thioredoxin, putative
Tb927.4.2600	hypothetical protein, conserved
Tb927.4.310	SPRY domain/HECT-domain (ubiquitin-transferase), putative
Tb927.4.3120	bilobe protein, conserved
Tb927.4.3320	uracil phosphoribosyltransferase, putative
Tb927.4.3700	hypothetical protein, conserved
Tb927.4.4690	hypothetical protein, conserved
Tb927.4.4700	hypothetical protein, conserved
Tb927.4.4940	hypothetical protein, conserved
Tb927.4.870	dynein heavy chain, putative
Tb927.5.1120	Phage tail fibre repeat, putative
Tb927.5.1180	hypothetical protein, conserved
Tb927.5.1270	hypothetical protein, conserved
Tb927.5.1360	nucleoside 2-deoxyribosyltransferase (NDRT)
Tb927.5.1490	Eukaryotic translation initiation factor 4 gamma type 1 (eif4g1)
Tb927.5.1700	replication Factor A 28 kDa subunit, putative
Tb927.5.1880	inhibitor of serine peptidase (ISP), putative (ISP2)
Tb927.5.1900	hypothetical protein, conserved
Tb927.5.1940	hypothetical protein, conserved
Tb927.5.2780	mitochondrial DNA polymerase beta
Tb927.5.2790	mitochondrial DNA polymerase beta-PAK (Pol beta-PAK)
Tb927.5.2950	Repeat of unknown function (DUF1126), putative
Tb927.5.3260	WD domain, G-beta repeat, putative
Tb927.5.3510	structural maintenance of chromosome 3 , putative (SMC3)
Tb927.5.3730	hypothetical protein, conserved

Tb927.5.4160	hypothetical protein, conserved
Tb927.5.4390	hypothetical protein, conserved
Tb927.5.4440	dynein light chain, putative
Tb927.5.4470	hypothetical protein, conserved
Tb927.5.4480	paraflagellar rod component par4, putative
Tb927.5.570	hypothetical protein, conserved
Tb927.5.580	prefoldin subunit, putative
Tb927.5.760	cell cycle sequence binding phosphoprotein (RBP33), putative
Tb927.5.830	hypothetical protein, conserved
Tb927.6.1180	hypothetical protein, conserved
Tb927.6.3040	small GTP-binding protein Rab28, putative
Tb927.6.3150	Hydin
Tb927.6.3330	hypothetical protein, conserved
Tb927.6.3580	ATP-dependent DEAD/H DNA helicase recQ, putative
Tb927.6.4100	hypothetical protein, conserved
Tb927.6.4300	glyceraldehyde 3-phosphate dehydrogenase, glycosomal (GAPDH)
Tb927.6.4670	MORN repeat-containing protein 1 (MORN1)
Tb927.6.4710	calmodulin, putative
Tb927.6.4780	DNA ligase I, putative
Tb927.6.5090	hypothetical protein, conserved
Tb927.6.860	hypothetical protein
Tb927.7.1050	40S ribosomal protein S16, putative
Tb927.7.1860	hypothetical protein, conserved
Tb927.7.190	thimet oligopeptidase, putative (THOP1)
Tb927.7.1920	paraflagellar rod component, putative (PFC5)
Tb927.7.2390	hypothetical protein, conserved
Tb927.7.3160	Cytoplasmic dynein 1 heavy chain (DYNC1H1), putative
Tb927.7.3330	hypothetical protein, conserved
Tb927.7.3590	hypothetical protein, conserved
Tb927.7.3990	mitochondrial DNA polymerase I protein C (POL1C)
Tb927.7.4810	HD domain containing protein, putative
Tb927.7.4870	hypothetical protein, conserved
Tb927.7.4910	hypothetical protein, conserved
Tb927.7.5020	60S ribosomal protein L19, putative
Tb927.7.5480	dihydrofolate reductase-thymidylate synthase (DHFR-TS)
Tb927.7.5660	hypothetical protein, conserved
Tb927.7.570	prefoldin, putative
Tb927.7.5920	mercaptopyruvate sulfurtransferase, putative (MST)
Tb927.7.6230	ADP-ribosylation factor, putative (ARF3)
Tb927.7.630	hypothetical protein, conserved
Tb927.7.920	dynein heavy chain, putative
Tb927.8.1550	paraflagellar rod component, putative (PFC3)

Tb927.8.1630	MSP-B, putative
Tb927.8.3050	Chitobiase/beta-hexosaminidase C-terminal domain containing protein, putative
Tb927.8.3060	cytosolic leucyl aminopeptidase, putative
Tb927.8.3250	dynein heavy chain, putative
Tb927.8.3410	Inositol hexakisphosphate, putative
Tb927.8.3790	paraflagellar rod component, putative (PFC2)
Tb927.8.4230	hypothetical protein, conserved
Tb927.8.4880	DNA polymerase alpha catalytic subunit
Tb927.8.4950	kinesin, putative
Tb927.8.5400	hypothetical protein, conserved
Tb927.8.550	peptide methionine sulfoxide reductase, putative
Tb927.8.5600	transaldolase, putative
Tb927.8.6060	2-amino-3-ketobutyrate coenzyme A ligase, putative
Tb927.8.6070	Trypanosome basal body component protein (TBBC)
Tb927.8.610	hypothetical protein, conserved
Tb927.8.6240	STOP axonemal protein
Tb927.8.6420	beta-ketoacyl-ACP reductase 2 (KAR2)
Tb927.8.6580	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (SDH1)
Tb927.8.6660	paraflagellar rod component, putative (PFC1)
Tb927.8.6980	hypothetical protein, conserved
Tb927.8.7950	Flagellar Member 4 (FLAM4)
Tb927.8.7970	hypothetical protein
Tb927.8.810	radial spoke protein RSP9, putative
Tb927.9.10010	chaperone protein DNAj, putative
Tb927.9.10440	DNA polymerase epsilon catalytic subunit, putative
Tb927.9.11230	calmodulin-like protein, putative
Tb927.9.11850	structural maintenance of chromosome 1, putative (SMC1)
Tb927.9.12390	hypothetical protein, conserved
Tb927.9.13440	Flagellar Member 5 (FLAM5)
Tb927.9.1510	D-ala D-ala ligase C-terminus/SET domain containing protein, putative
Tb927.9.15450	Domain of unknown function (DUF4205), putative
Tb927.9.1700	btb/poz domain containing protein, putative
Tb927.9.1750	Fibronectin type III domain containing protein, putative
Tb927.9.2310	hypothetical protein, conserved
Tb927.9.2940	hypothetical protein, conserved
Tb927.9.4420	hypothetical protein, conserved
Tb927.9.5190	proliferative cell nuclear antigen (PCNA), putative
Tb927.9.5590	DNA topoisomerase ii (TOP2)
Tb927.9.6100	TFIIF-stimulated CTD phosphatase, putative
Tb927.9.6130	calmodulin, putative
Tb927.9.6760	hypothetical protein, conserved
Tb927.9.7180	adenosine monophosphate deaminase, putative

Tb927.9.7550	adenylosuccinate lyase, putative (ADSL)
Tb927.9.8760	hypothetical protein, conserved
Tb927.9.8990	Ankyrin repeats (3 copies), putative
Tb927.9.9230	hypothetical protein, conserved
Tb927.9.9730	hypothetical protein, conserved
Tb927.9.9810	hypothetical protein, conserved

Supplementary Table 8: Proteins classified as cell-cycle regulated that are not identified in transcriptomic data.

Gene ID	Product Description
Tb927.2.1820	protein kinase, putative
Tb927.2.3720	ubiquitin-conjugating enzyme, putative
Tb927.2.4900	hypothetical protein, conserved
Tb927.4.3290	Trypanosomal VSG domain containing protein, putative
Tb927.4.4400	hypothetical protein, conserved
Tb927.4.4950	hypothetical protein, conserved
Tb927.6.300	receptor-type adenylate cyclase GRESAG 4, putative
Tb927.6.820	pumilio RNA binding protein, putative (PUF4)
Tb927.6.3360	hypothetical protein, conserved
Tb927.7.1270	hypothetical protein, conserved
Tb927.7.1630	hypothetical protein, conserved
Tb927.8.910	hypothetical protein, conserved
Tb927.8.1370	hypothetical protein, conserved
Tb927.8.4770	small GTP-binding protein Rab18 (TbRAB18)
Tb927.8.6910	cyclophilin, putative
Tb927.9.1850	60S ribosomal protein L35, putative
Tb927.9.2680	hypothetical protein, unlikely
Tb927.9.12630	glycerol kinase, glycosomal (glk1)
Tb927.9.13920	kinetoplastid membrane protein KMP-11 (KMP-11)
Tb927.10.2010	hexokinase (HK1)
Tb927.10.7300	hypothetical protein, conserved
Tb927.10.11840	WD domain, G-beta repeat, putative
Tb927.10.12800	Zinc finger CCCH domain-containing protein 38 (ZC3H38)
Tb927.10.13650	ARF-like 2-binding protein, putative
Tb11.02.5105b	hypothetical protein, conserved
Tb11.1810b	retrotransposon hot spot protein (RHS, pseudogene), putative
Tb11.v5.0452	GTPase activating protein, putative
Tb11.v5.0553	ubiquitin hydrolase, putative
Tb11.v5.0609	DNA-directed RNA polymerase III largest subunit, putative
Tb11.v5.0971	mitochondrial DNA polymerase I protein A, putative
Tb927.11.20300	variant surface glycoprotein (VSG), putative

Tb927.11.2350	hypothetical protein, conserved
Tb927.11.3620	nucleobase/nucleoside transporter 8.1 (NT8.1)
Tb927.11.7670	Cornifin (SPRR) family, putative
Tb927.11.9620.1	hypothetical protein, conserved
Tb927.11.11750	membrane-bound acid phosphatase, putative
Tb927.11.12070	hypothetical protein, conserved
Tb927.11.12820	ribonucleoside-diphosphate reductase small chain (RNR2)
Tb927.11.13690.1	radial spoke protein RSP11, putative
Tb927.11.16790	mitogen-activated protein kinase (ECK1)
Tb11.1650	variant surface glycoprotein (VSG, pseudogene), putative

Supplementary Table 9: Transcripts classified as cell-cycle regulated that are not identified in proteomic data.

Gene ID	Product Description
Tb927.1.1350	kinesin heavy chain, putative
Tb927.1.1360	hypothetical protein, unlikely
Tb927.1.1740	hypothetical protein
Tb927.1.1920	conserved protein, unknown function
Tb927.1.2290	hypothetical protein, unlikely
Tb927.1.2310	hypothetical protein
Tb927.1.2320	hypothetical protein, conserved
Tb927.1.2370	beta tubulin
Tb927.1.2510	histone H3, putative
Tb927.1.2560	hypothetical protein, unlikely
Tb927.1.2620	hypothetical protein, unlikely
Tb927.1.3150	hypothetical protein, conserved
Tb927.1.3440	hypothetical protein, unlikely
Tb927.1.3560	WW domain containing protein, putative
Tb927.1.4190	hypothetical protein, unlikely
Tb927.1.4500	SNARE associated Golgi protein, putative
Tb927.1.4510	hypothetical protein, unlikely
Tb927.1.4800	hypothetical protein, conserved
Tb927.10.10210	procyclin-associated gene 4 (PAG4) protein (PAG4)
Tb927.10.10220	procyclin-associated gene 2 (PAG2) protein (PAG2)
Tb927.10.10240	procyclin-associated gene 1 (PAG1) protein (PAG1)
Tb927.10.10250	EP2 procyclin (EP2)
Tb927.10.10550	histone H2B, putative
Tb927.10.11710	hypothetical protein
Tb927.10.12670	hypothetical protein, conserved
Tb927.10.12950	hypothetical protein, conserved
Tb927.10.1330	hypothetical protein, conserved

Tb927.10.13380	hypothetical protein, conserved
Tb927.10.1350	hypothetical protein, conserved
Tb927.10.14900	hypothetical protein, conserved
Tb927.10.15250	paraflagellar rod component, putative (PFC15)
Tb927.10.16070	eukaryotic translation initiation factor 4E, putative
Tb927.10.1940	serine/threonine-protein kinase, putative
Tb927.10.3130	myotubularin-associated protein, putative
Tb927.10.4600	predicted SET domain protein
Tb927.10.4730	ribonuclease HII, putative
Tb927.10.5360	40S ribosomal protein S10, putative
Tb927.10.550	hypothetical protein, conserved
Tb927.10.5650	hypothetical protein, conserved
Tb927.10.5960	engulfment and cell motility domain 2, putative
Tb927.10.6500	amino acid transporter, putative
Tb927.10.6520	hypothetical protein, conserved
Tb927.10.7000	hypothetical protein, conserved
Tb927.10.7890	hypothetical protein, conserved
Tb927.10.8450	glucose transporter 1E (THT1E)
Tb927.10.9060	hypothetical protein, conserved
Tb927.10.9730	hypothetical protein, conserved
Tb927.11.10310	hypothetical protein, unlikely
Tb927.11.10610	hypothetical protein, conserved
Tb927.11.10800	Domain of unknown function (DUF4496), putative
Tb927.11.1100	cysteine peptidase, Clan CA, family C2, putative
Tb927.11.11540	unspecified product
Tb927.11.11740	membrane-bound acid phosphatase, putative
Tb927.11.11920	molybdenum cofactor biosynthesis protein, putative
Tb927.11.12980	hypothetical protein, conserved
Tb927.11.13030	calmodulin
Tb927.11.14370	hypothetical protein, conserved
Tb927.11.1470	65 kDa invariant surface glycoprotein-like protein
Tb927.11.14800	hypothetical protein, conserved
Tb927.11.14840	chromosomal passenger protein (CTC2)
Tb927.11.14860	hypothetical protein, conserved
Tb927.11.14930	hypothetical protein, conserved
Tb927.11.15090	hypothetical protein, conserved
Tb927.11.16140	DNA replication licensing factor MCM7, putative
Tb927.11.2310	hypothetical protein, conserved
Tb927.11.2700	hypothetical protein, conserved
Tb927.11.2710	cell division cycle 45 (CDC45), putative
Tb927.11.3000	60S ribosomal protein L37, putative
Tb927.11.340	hypothetical protein, conserved

Tb927.11.4590	hypothetical protein, conserved
Tb927.11.4600	unspecified product
Tb927.11.5010	TFIIF-stimulated CTD phosphatase, putative
Tb927.11.5340	Cell division protein kinase, putative
Tb927.11.5670	Mnd1 family, putative
Tb927.11.5980	conserved protein
Tb927.11.6180	60S ribosomal protein L28, putative
Tb927.11.6550	PUF nine target 1 (PNT1)
Tb927.11.780	hypothetical protein, conserved
Tb927.11.7800	hypothetical protein, conserved
Tb927.11.8750	hypothetical protein, conserved
Tb927.11.890	hypothetical protein, conserved
Tb927.11.9030	cation transporter, putative
Tb927.11.9330	helicase-like protein
Tb927.2.4480	hypothetical protein, conserved
Tb927.2.5040	MORN repeat, putative
Tb927.3.1420	hypothetical protein, conserved
Tb927.3.1860	ATP-grasp domain containing protein, putative
Tb927.3.2760	hypothetical protein, conserved
Tb927.3.2790	leucine-rich repeat protein (LRRP), putative
Tb927.3.3310	60S ribosomal protein L13, putative
Tb927.3.3920	serine/threonine-protein kinase, putative
Tb927.3.4310	Paraflagellar rod protein 1-3 (PFR1-3)
Tb927.3.4670	cdc2-related kinase 10, putative (CRK10)
Tb927.3.5200	hypothetical protein, conserved
Tb927.3.5260	hypothetical protein, conserved
Tb927.3.590	adenosine transporter, putative
Tb927.3.870	Histone RNA hairpin-binding protein RNA-binding domain containing protein, putative
Tb927.4.1290	hypothetical protein, conserved
Tb927.4.2270	Histone-like transcription factor (CBF/NF-Y) and archaeal histone, putative
Tb927.4.2490	hypothetical protein, conserved
Tb927.4.2780	hypothetical protein, conserved
Tb927.4.3440	hypothetical protein, conserved
Tb927.4.4080	C-5 sterol desaturase, putative
Tb927.4.4130	hypothetical protein, conserved
Tb927.4.4180	hypothetical protein, conserved
Tb927.4.4480	hypothetical protein, conserved
Tb927.4.4900	hypothetical protein
Tb927.4.5110	kinetoplastid kinetochore protein 8 (kkt8)
Tb927.4.5330	hypothetical protein
Tb927.5.1550	mitochondrial carrier protein (MCP23)
Tb927.5.2390	hypothetical protein, conserved

Tb927.5.2970	suppressive immunomodulating factor (TSIF)
Tb927.5.3460	hypothetical protein, conserved
Tb927.5.3680	GPI-GlcNAc transferase complex, PIG-H component, putative
Tb927.5.4240	histone H4, putative
Tb927.5.4530	fatty acid elongase, putative
Tb927.5.470	monocarboxylate transporter-like protein
Tb927.5.850	hypothetical protein, conserved
Tb927.6.1220	hypothetical protein, conserved
Tb927.6.1710	hypothetical protein, conserved
Tb927.6.3180	hypothetical protein, conserved
Tb927.6.3430	protein kinase, putative
Tb927.6.3880	hypothetical protein, conserved
Tb927.6.900	hypothetical protein, conserved
Tb927.7.1590	mitotic cyclin (CYC8)
Tb927.7.1750	60S ribosomal protein L7, putative
Tb927.7.2110	kinetoplastid kinetochore protein 11 (kkt11)
Tb927.7.2580	zinc finger protein family member, putative (ZC3H19)
Tb927.7.2650	hypothetical protein, conserved
Tb927.7.2870	histone H2A, putative
Tb927.7.3970	hypothetical protein, conserved
Tb927.7.4580	Cornifin (SPRR) family, putative
Tb927.7.4860	kinetoplastid kinetochore protein 13 (kkt13)
Tb927.7.490	hypothetical protein, conserved
Tb927.7.5040	kinesin, putative (KIN-B)
Tb927.7.5220	protein kinase, putative
Tb927.7.5370	hypothetical protein, conserved
Tb927.7.5650	kinesin, putative
Tb927.7.5740	conserved protein
Tb927.7.6050	receptor-type adenylate cyclase GRESAG 4, putative
Tb927.7.6060	receptor-type adenylate cyclase GRESAG 4, putative
Tb927.7.6070	receptor-type adenylate cyclase GRESAG 4, putative
Tb927.7.610	mitochondrial DNA ligase homolog, LIG k-alpha
Tb927.7.6310	polo-like protein kinase (PLK)
Tb927.7.6590	hypothetical protein, conserved
Tb927.7.6630	hypothetical protein, conserved
Tb927.7.6730	hypothetical protein, conserved
Tb927.7.6840	Centrosomal spindle body, CEP44, putative
Tb927.7.7150	hypothetical protein, conserved
Tb927.7.720	conserved protein
Tb927.7.7270	hypothetical protein, conserved
Tb927.7.7320	hypothetical protein, conserved
Tb927.7.750	hypothetical protein, conserved

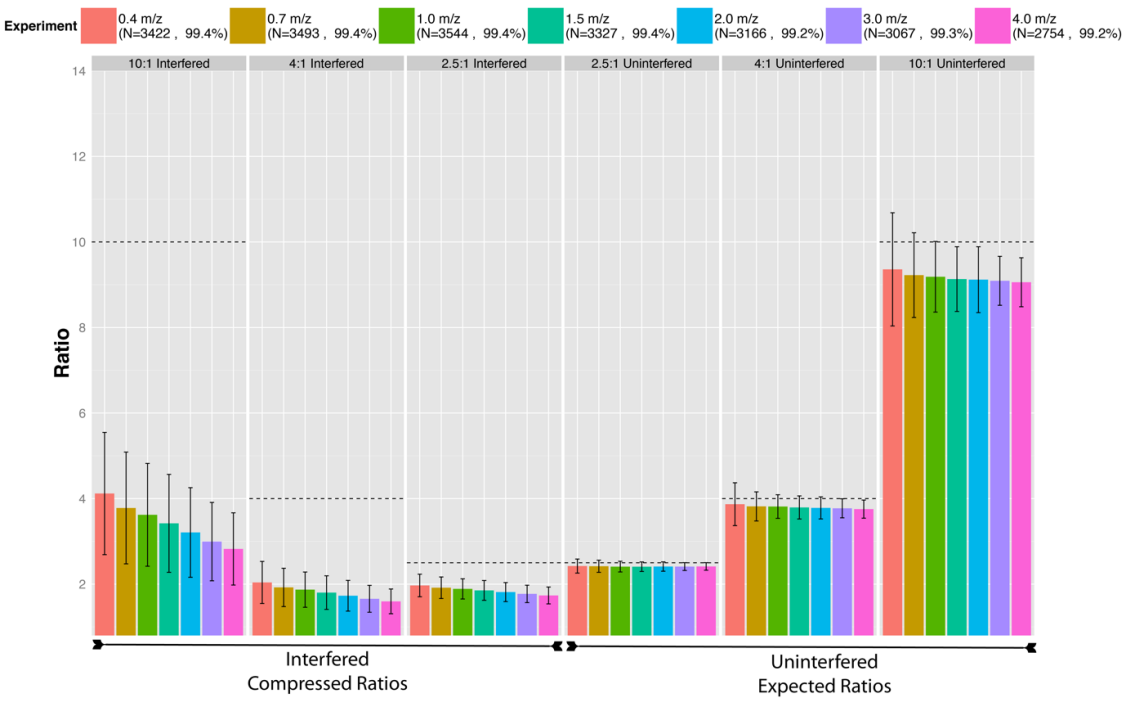
Tb927.8.1150	kinetoplastid kinetochore protein 9 (kkt9)
Tb927.8.1920	histone-lysine n-methyltransferase (DOT1A)
Tb927.8.3850	PSP1 C-terminal conserved region, putative
Tb927.8.3990	hypothetical protein, conserved
Tb927.8.4060	flagellum-adhesion glycoprotein, putative
Tb927.8.4350	hypothetical protein, conserved
Tb927.8.4730	amino acid transporter, putative
Tb927.8.490	hypothetical protein
Tb927.8.4900	hypothetical protein, conserved
Tb927.8.4990	Paraflagellar rod protein 2-3 (PFR2-3)
Tb927.8.5570	transporter, putative
Tb927.8.6120	hypothetical protein, conserved
Tb927.8.6300	hypothetical protein, conserved
Tb927.8.7070	hypothetical protein, conserved
Tb927.9.10480	hypothetical protein, unlikely
Tb927.9.10540	hypothetical protein, unlikely
Tb927.9.11690	hypothetical protein, conserved
Tb927.9.12310	hypothetical protein, conserved
Tb927.9.12590	glycerol kinase, glycosomal (glk1)
Tb927.9.14300	hypothetical protein, conserved
Tb927.9.14400	leucine-rich repeat protein (LRRP), putative
Tb927.9.15520	BARP protein (BARP)
Tb927.9.15580	BARP protein (BARP)
Tb927.9.15600	BARP protein (BARP)
Tb927.9.1580	hypothetical protein, unlikely
Tb927.9.1670	protein kinase, putative
Tb927.9.4550	hypothetical protein, conserved
Tb927.9.4580	hypothetical protein, unlikely
Tb927.9.5920	hypothetical protein, unlikely
Tb927.9.6090	PTP1-interacting protein, 39 kDa
Tb927.9.6140	hypothetical protein, unlikely
Tb927.9.8620	hypothetical protein, conserved

Appendix: Optimisation of TMT quantitation

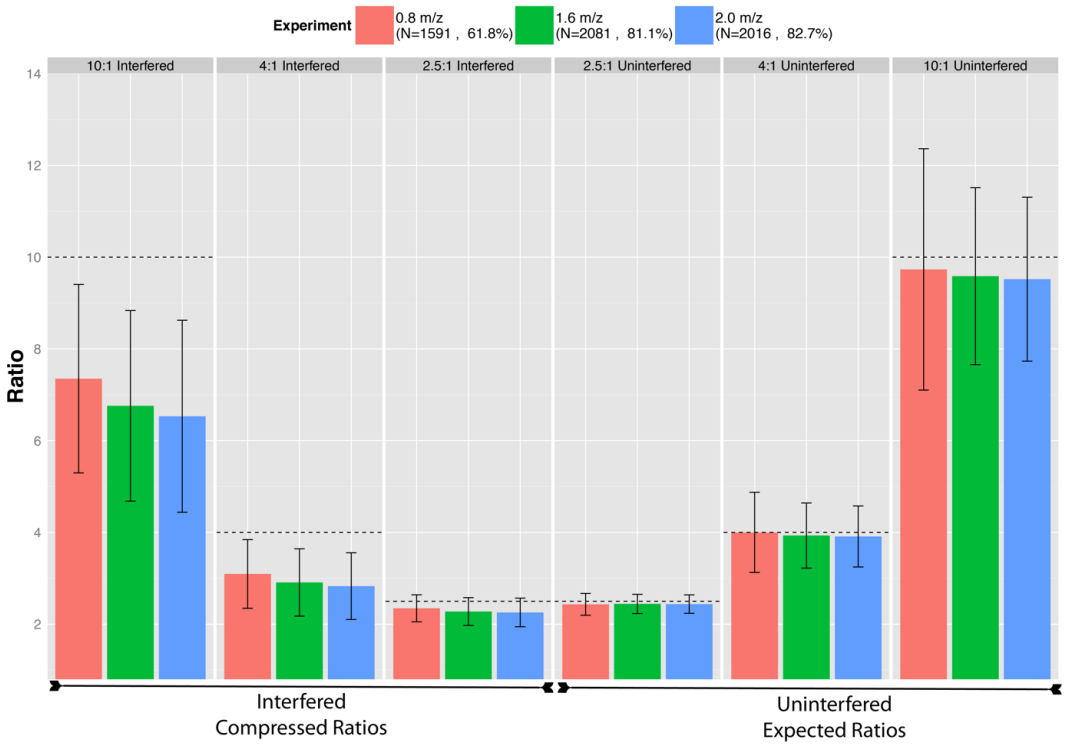
7.1 Extension of data from Chapter 4

The data presented in this Appendix is an extension of the data produced in Chapter 4. A mixture of TMT labelled peptides was produced in the manner described in section 4.3.3, mixing human peptides, labelled in ratios of 10:4:1:1:4:10 with six-plex TMT reagents that generate reporter ions at *m/z* 126, 127, 128, 129, 130 and 131, and trypanosome peptides labelled at a 10:10:10 ratio in the *m/z* 126, 127 and 129 TMT reporter ion channels. The addition of trypanosome TMT labelled peptides mimics the effect of compression when quantifying human peptides in the first three reporter ion channels (126, 127 and 128), while the latter three (129, 130, 131) act as a control, demonstrating the accuracy and precision of quantitation of the expected 10-fold, 4-fold and 2.5-fold ratios of intensity between each TMT reporter.

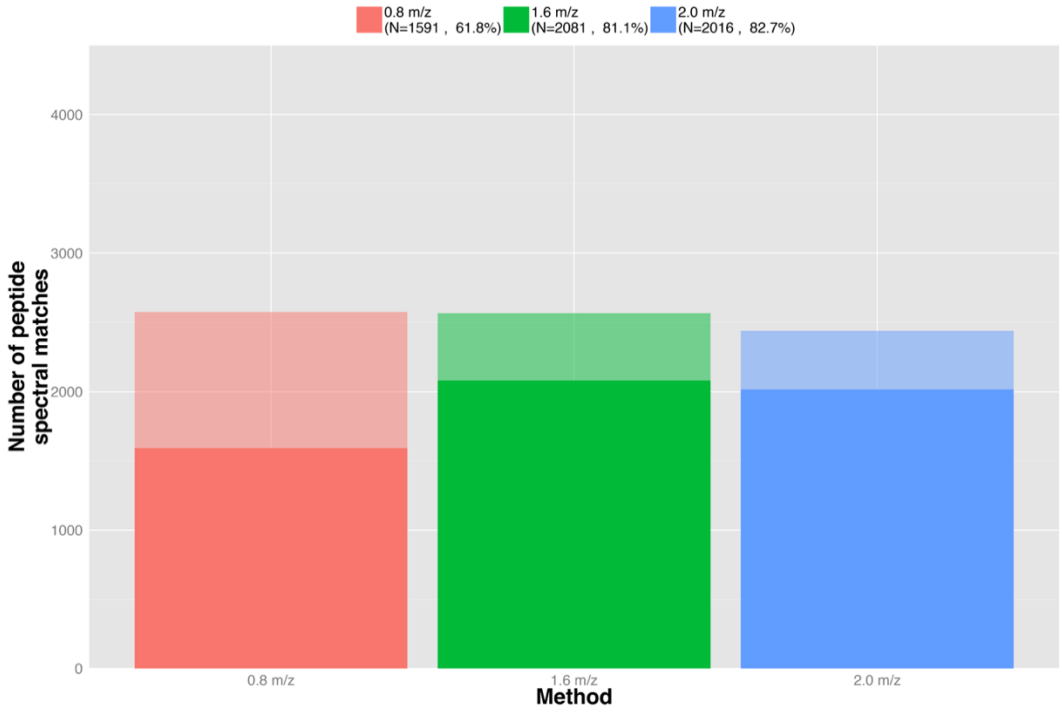
Supplementary Figure 1 shows an extended range of the precursor ion mass isolation widths tested on a QExactive+ mass spectrometer, compared to Figure 4.8. The general trend for more accurate quantitation of the three compressed ratios with narrower isolation widths holds for this extended set of data including 0.4, 0.7, 1.0, 1.5, 2.0, 3.0 and 4.0 *m/z*. In Chapter 4 we also saw an increase in the number of identified psms as we decreased the isolation window. With the extended range of isolation widths tested we see an increase in the number of psms identified from 2,754 to 3,544 as we go from 4.0 *m/z* to 1.0 *m/z*. A further decrease of the isolation window leads to a lower number of psms identified, to 3,493 at 0.7 *m/z* and 3,422 at 0.4 *m/z*.



Supplementary Figure 1: Effect on TMT quantitation of modifying the isolation window on a QExactive+ MS. N shows the number of psms quantified, together with the percentage of psms identified that pass the quality control criteria for quantification.

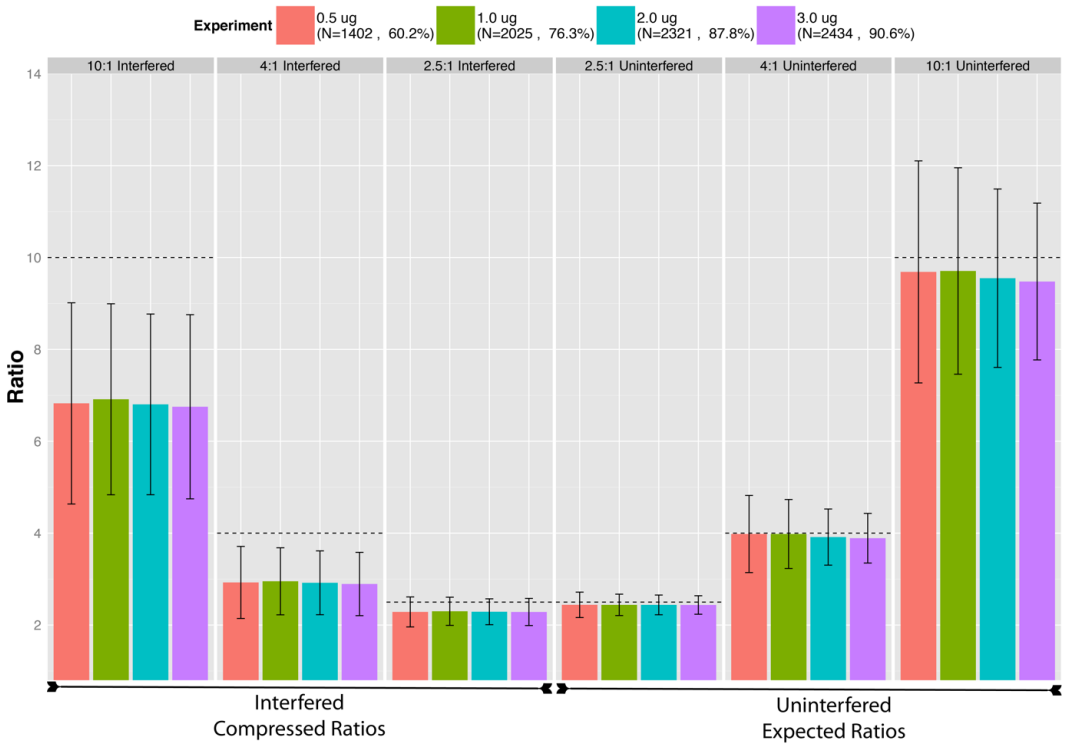


Supplementary Figure 2: Effect on TMT quantitation of modifying the isolation window on a Fusion MS. N shows the number of psms quantified, together with the percentage of psms identified that pass the quality control criteria for quantification.

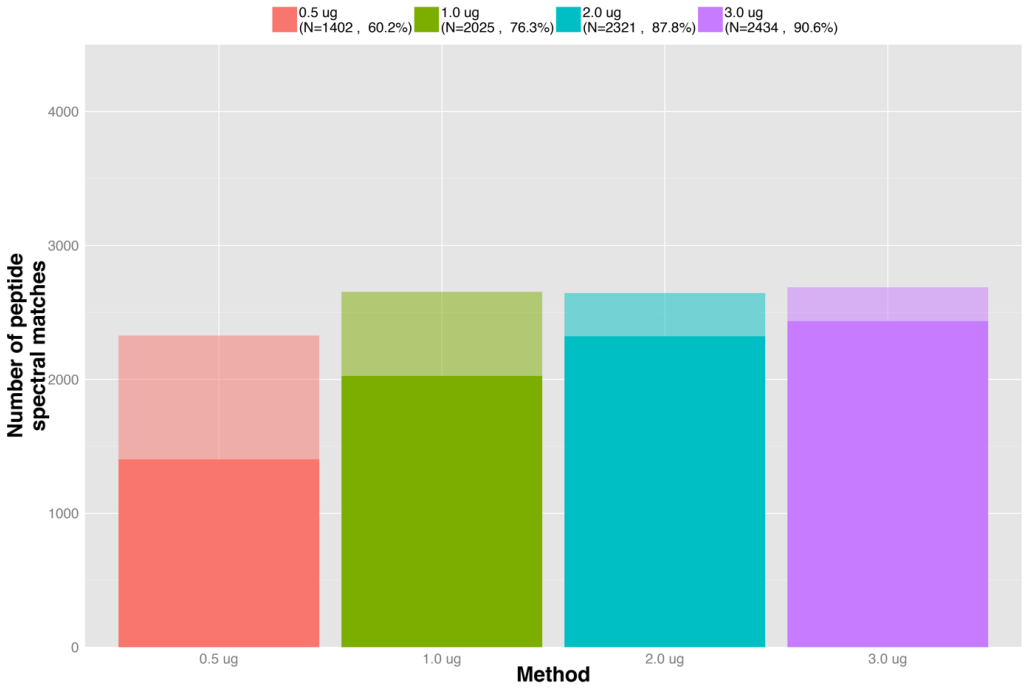


Supplementary Figure 3: Effect on psm identification and *quantification* of isolation window on a Fusion MS. N shows the number of psms quantified, together with the percentage of psms identified that pass the quality control criteria for quantification.

In Supplementary Figures 2 and 3 we tested the effect of modifying the isolation width for precursor ion selection on a Fusion MS. Supplementary Figure 2 shows that a narrower isolation width produces more accurate quantitation of compressed ratios, though more variability in the quantitation of the uncompressed ratios is also observed. Supplementary Figure 3 also shows that although all three isolation widths tested produced similar numbers of psms identified, narrowing the isolation width to 0.8 *m/z* resulted in a large drop in the number of psms both identified and *quantified* to 1,591 (61.8% of those identified), from 2,081 at 1.6 *m/z* (81.1% of those identified).



Supplementary Figure 4: Effect on TMT quantitation of peptide load on a Fusion MS. N shows the number of psms quantified, together with the percentage of psms identified that pass the quality control criteria for quantification.

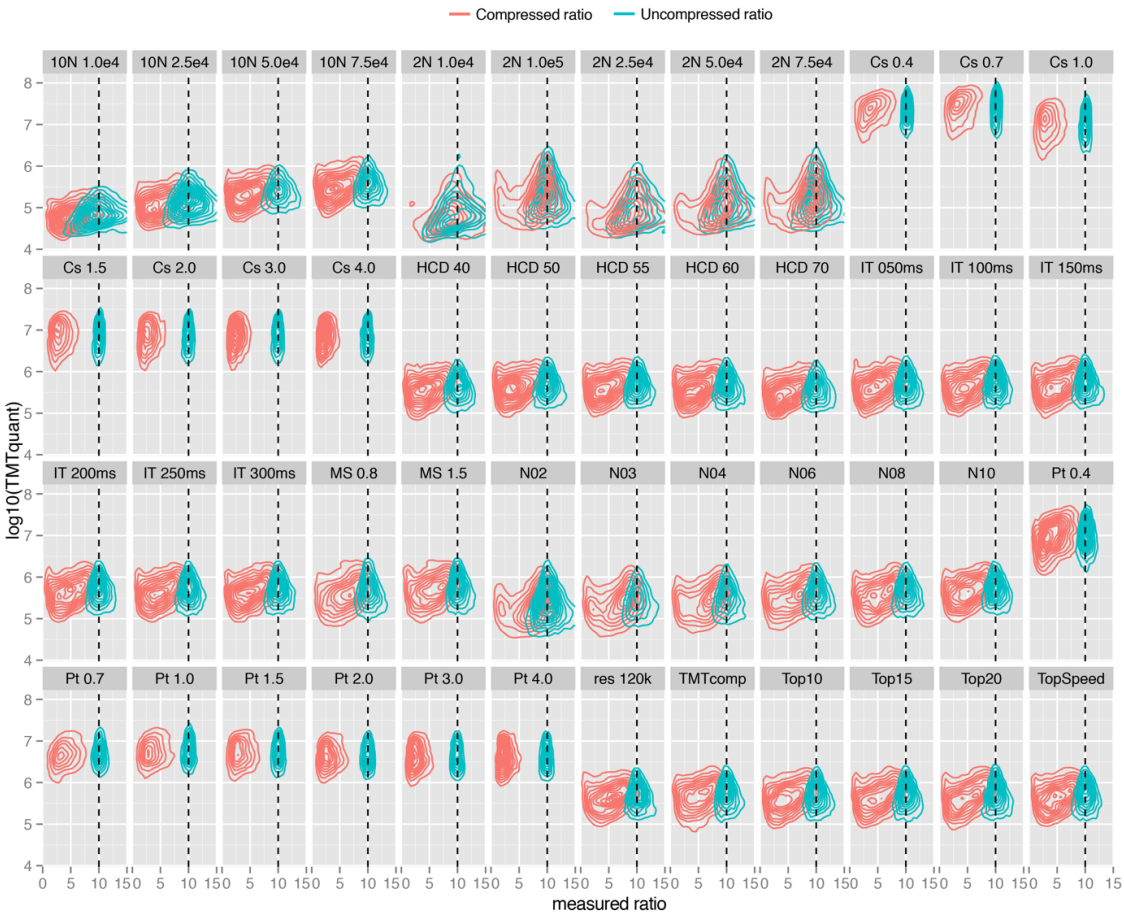


Supplementary Figure 5: Effect on psm identification and *quantification* of peptide load on as Fusion MS. N shows the number of psms quantified, together with the percentage of psms identified that pass the quality control criteria for quantification.

Supplementary Figures 4 and 5 demonstrate the effect on TMT quantitation of modifying the peptide load in a mass spectrometry run. In this set of experiments there appears to be little difference in the quantitation of either the compressed or uncompressed ratios when loading anything from 0.5 to 3.0 µg of peptide, with similar accuracy and precision. However, as shown in Supplementary Figure 5, there are significant differences between the number of psms identified and *quantified* between these samples. At 0.5 µg peptide loads only 60.2% of psms identified pass our quality control for quantitation, described in Chapter 4, producing data for 1,402 psms. This increases to 2,025 psms at 1.0 µg, 2,321 at 2.0 µg, and 2,434 at 3.0 µg. At a peptide load of 3.0 µg we see that over 90% of psms identified are also being quantified.

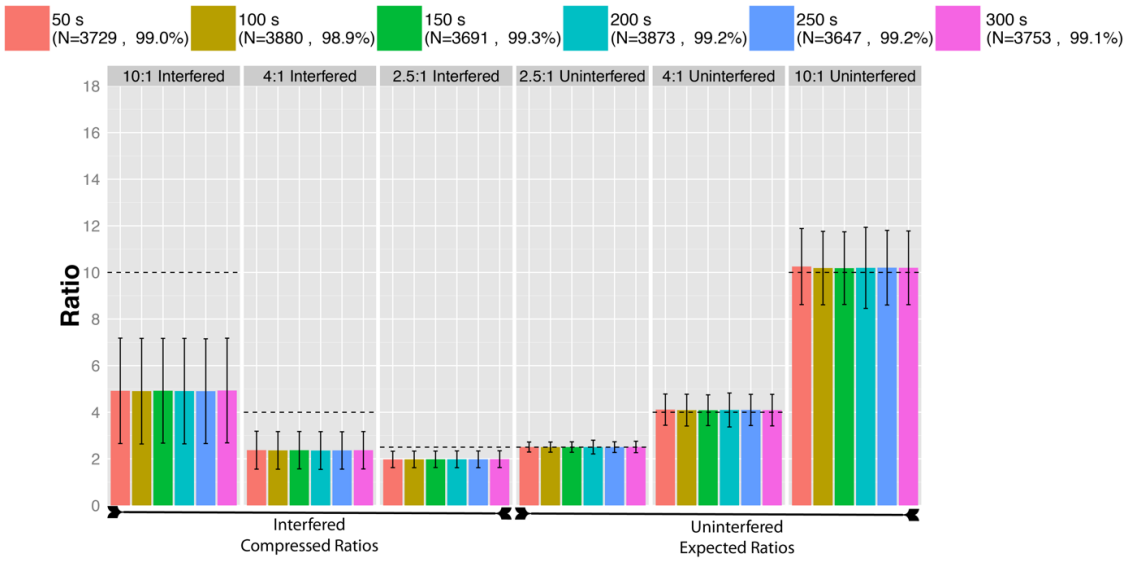
7.2 Preparation of second batch of TMT peptide mixture

The experiments that follow utilise a second preparation of TMT-modified peptides produced by mixing mouse and trypanosome peptides in a similar manner to that described in Section 4.3.3. Instead of using human peptides to assess effects on compression we are now using mouse peptides, mixed in the ratios of 10:4:1:1:4:10 with six-plex TMT reagents that generate reporter ions at *m/z* 126, 127, 128, 129, 130 and 131. Trypanosome peptides were again used to mimic interfering peptides, and labelled in the *m/z* 126, 127 and 128 TMT reporter ion channels. A number of experiments are displayed which are described in the caption of Supplementary Figure 6, a subset of which are discussed in more detail in later figures.



Supplementary Figure 6: Contour plots of compressed (red) and uncompressed (blue) psms with an expected 10-fold ratio with a range of methods tested on QExactive, QExactive+ and Fusion with mouse and trypanosome TMT-peptide mix. Facet titles indicate the experiment performed. 'Pt' indicates experiments performed on a QExactive instrument with numbers indicating isolation widths tested. Similarly, 'Cs' indicates experiments performed on a QExactive+ instrument. All other experiments were performed on a Fusion. 10N or 2N 1.0e4, 2.5e4, 7.5e4 and 1.0e5 shows the effect of modifying the AGC target for the MS3 scan at either 10 SPS-notches or 2 SPS-notches. HCD40, 50, 55, 60 and 70 show the effect of modifying the collision energy used for reporter ion fragmentation. IT 050ms, 100ms, 150ms, 200ms, 250ms and 300ms show the effect of modifying the MS3 injection time. MS 0.8 and 1.5 show the effect of modifying isolation width. N02, 03, 04, 06, 08 and 10 show the effect of modifying the number of SPS-notches used for the quantitative MS3 scan. The res120k experiment shows the effect of acquiring data at a higher Orbitrap resolution. TMTcomp uses methods reported in another publication. Top10, 15, 20 and Speed show the effect of different ion selection methods.

7.3 Differential effects of increasing the MS3 scan maximum injection time

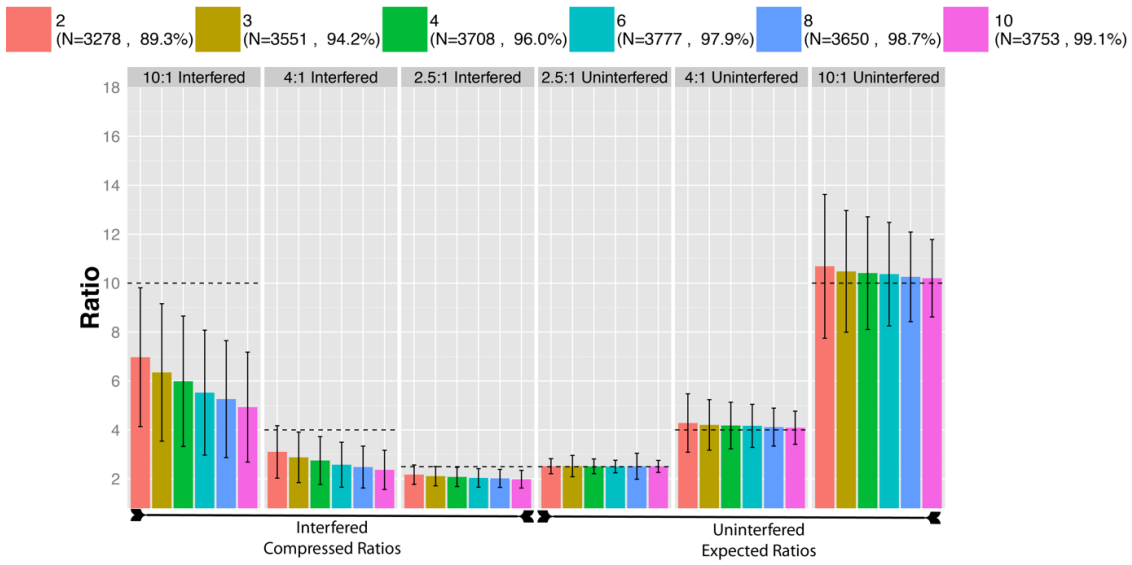


Supplementary Figure 7: Effect on TMT quantitation of modifying the MS3 scan maximum injection time on a Fusion MS with mouse and trypanosome TMT-peptide mix. N shows the number of psms quantified, together with the percentage of psms identified that pass the quality control criteria for quantification.

The maximum injection time for MS3 scans does not appear to modify the accuracy or precision of quantitation of compressed or uncompressed ratios when using the mouse and trypanosome peptide mixture. There is also no effect on the number of psms identified that pass quantification criteria, with all experiments quantifying >99% of the psms identified. This is in contrast to the data presented in Figures 4.10 and 4.11 in Chapter 4, where we observed an increased precision in quantification of the uncompressed ratios, and an increase in the number of peptides passing quantification criteria as we increased the MS3 maximum injection time from 105 ms to 300 ms. We believe this is likely due to inaccuracies in peptide quantitation which meant that, inadvertently, more peptide was loaded in the mouse and trypanosome peptide mix compared to the human and trypanosome peptide mix described in Chapter 4. This can also be observed from the scatterplots of the 10-fold ratios and total TMT reporter ion intensity for each identified psm in Supplementary Figure 6 and Figure 4.9. In the

experiment in Figure 4.9, there are many psms with a total TMT reporter ion intensity below 100,000 units. However, in the experiments where maximum ion injection time was modified, displayed in Supplementary Figure 6 (IT 050ms, IT 100ms, IT 150ms, IT200ms, IT250ms, IT300ms), there are few psms that are detected below a total TMT reporter ion intensity of 100,000. It may therefore be the case that due to a higher load of peptide in these experiments, the maximum injection time was not being met for any of the peptides quantified, even at a maximum injection time of 50 ms.

7.4 Optimising the number of SPS notches

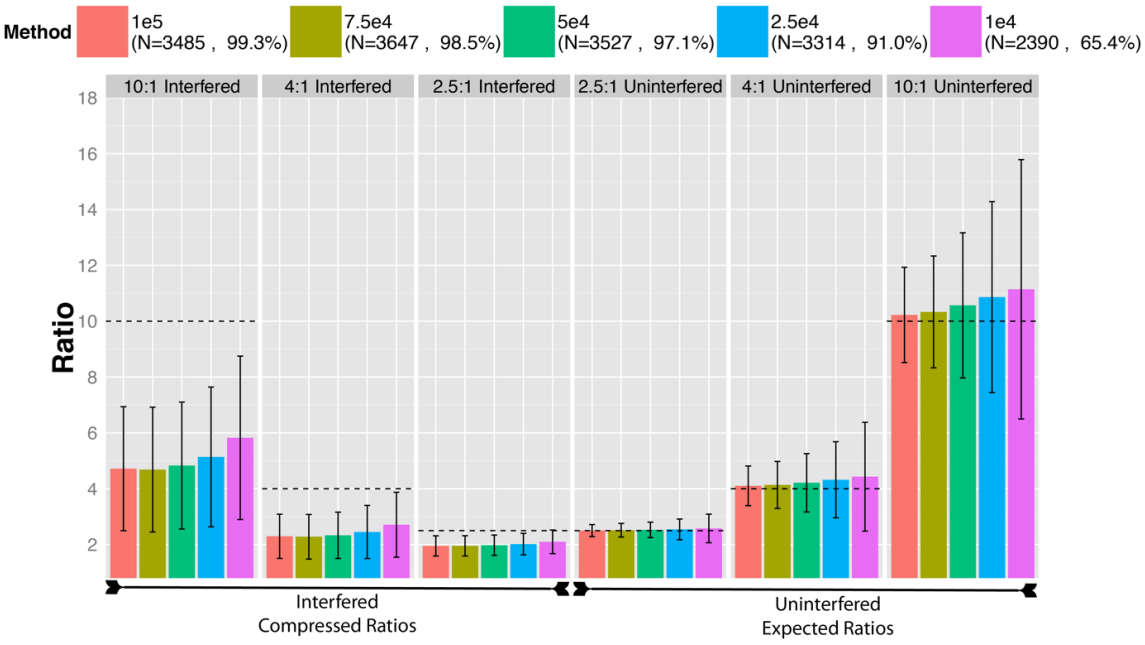


Supplementary Figure 8: Effect on TMT quantitation of decreasing the number of SPS-notches used on a Fusion MS with mouse and trypanosome TMT-peptide mix. N shows the number of psms quantified, together with the percentage of psms identified that pass the quality control criteria for quantification.

The original method published for synchronous precursor selection (SPS) MS3 for isobaric tag quantitation utilised 10 ‘notches’ to select fragment ions for higher energy collision dissociation for fragmentation of reporter ions for the quantitative MS3 scan. This means that 10 fragment ions produced from fragmentation of the precursor ion were re-isolated for this second round of high energy fragmentation. In Supplementary

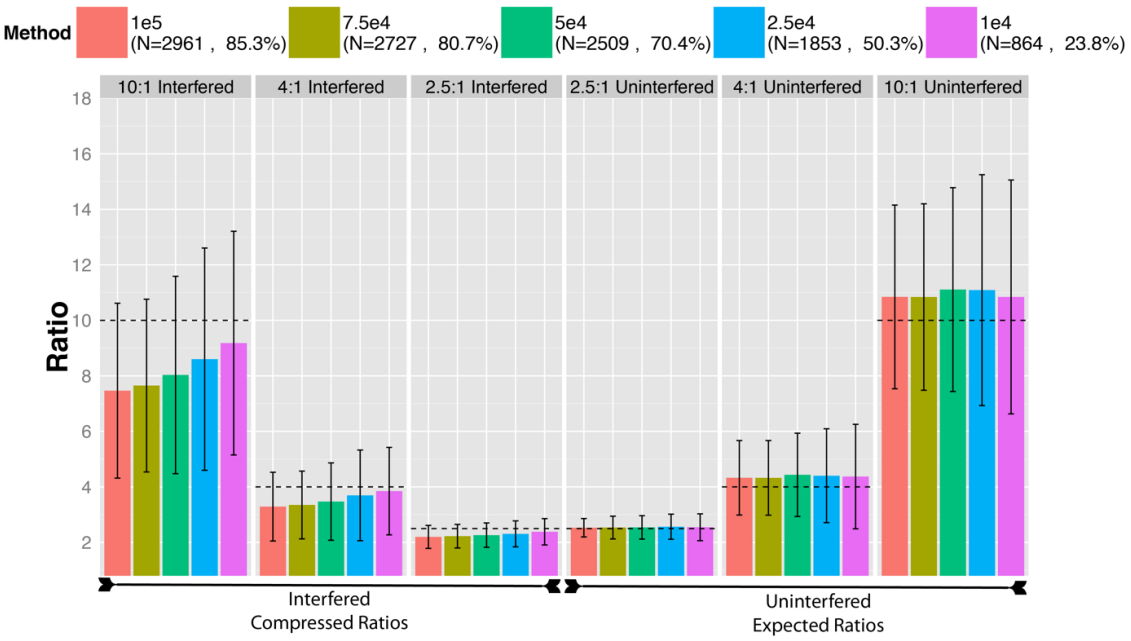
Figure 8 we investigated the effect on quantiation of TMT by reducing the number of fragment ions (or 'notches') that were re-isolated for fragmentation for the quantitative MS3 scan. Reducing the number of notches increases the accuracy of quantitation of compressed ratios. This may be due to reduced effects of co-isolation of contaminant peptide ions. By using a higher number of notches you increase the chance that any one of the notches selected may contain a fragment ion of a contaminating peptide, therefore leading to the effect of compression as shown in Figure 4.7. Reducing the number of notches reduces the likelihood of this occurring, especially as fragment ions from the ion of interest are likely to be higher intensity, and therefore selected first. However, we also see in the uncompressed ratios an increased imprecision in quantitation. This likely occurs due to a decrease in the amount of TMT reporter ion available for quantitation, which leads to increased variability in quantitation too. This can be observed in Supplementary Figure 6 (N02, N03, N04, N06, N08, N10), where the contour plots of data acquired with a higher number of notches shows a higher total TMT reporter ion intensity, though a wider deviation in the compressed and uncompressed ratios.

7.4 Modification of MS3 AGC target at 10 and 2 SPS notches



Supplementary Figure 9: Effect on TMT quantitation of decreasing the automatic gain control target using 10 SPS-notches on a Fusion MS with mouse and trypanosome TMT-peptide mix. N shows the number of psms quantified, together with the percentage of psms identified that pass the quality control criteria for quantification.

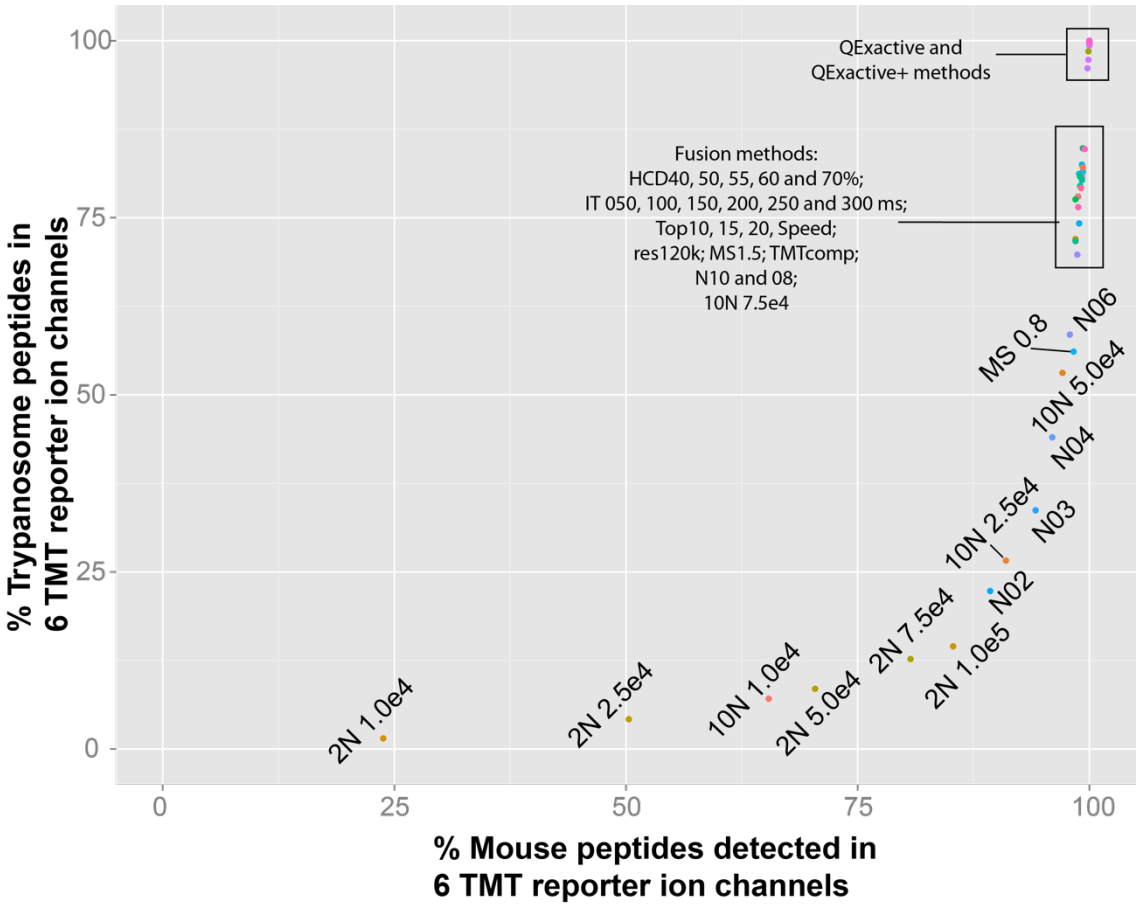
Decreasing the automatic gain control target of ions to be used for the quantitative MS3 scan from $1x10^5$ down to $1x10^4$ shows a trend towards more accurate quantitation of compressed ratios. For low abundance peptides in particular, it may be that a lower AGC target limits the amount of co-isolated contaminant ions that can also be accumulated prior to the quantitative MS3 scan. Unfortunately, although we see an increase in accuracy of compressed ratio quantitation, we also see an increased imprecision in quantitation of the uncompressed ratios as we decrease the AGC target. We also observe the effect of a decreased number of identified ions passing quality control criteria to be utilised for quantitation. At an AGC target of $2.5x10^4$, 3,314 psms are quantified (91% of those identified). This drops to 2,390 psms at an AGC target of $1x10^4$ (65.4% of those identified). This effect is much more marked when the number of SPS-notches used is decreased from 10 to 2.



Supplementary Figure 10: Effect on TMT quantitation of decreasing the automatic gain control target using 2 SPS-notches on a Fusion MS with mouse and trypanosome TMT-peptide mix. N shows the number of psms quantified, together with the percentage of psms identified that pass the quality control criteria for quantification.

At 2 notches we again see an improvement of the compressed ratio accuracy and the variability of the uncompressed ratio is similar across all AGC targets tested. However, we see a large drop in number of psms quantified as we reduce the AGC at 2 notches. Starting from 2,961 psms (85.3% of those identified) at an AGC target of $1x10^5$, we see a drop to 2,509 psms (70.4% of those identified) at an AGC target of $5x10^4$, further dropping to 864 psms (23.8% of those identified) at an AGC target of $1x10^4$. These results again highlight the balance between achieving accurate quantitation of compressed ratios and precise quantitation of the uncompressed ratios.

7.5 Identification of compression through analysis of trypanosome peptides



Supplementary Figure 11: Percentage of mouse and trypanosome peptide spectral matches identified in all 6 TMT reporter ions channels. Nomenclature for experiments plotted are identical to Supplementary Figure 6.

Another method to assess the effects of compression is to determine how many psms identified as trypanosome peptides are quantified in all 6 TMT channels. As these peptides have only been labelled using the 126, 127 and 128 *m/z* TMT reporter ions, they should only be detected in 3 channels. Identification in 6 channels indicates the co-isolation of mouse peptides, which have been labelled with all 6 TMT reporter ions. Therefore, in the ideal scenario in Supplementary Figure 11 we would expect to see experiments plotted in the right hand corner, with 0% of trypanosome peptides and 100% of mouse peptides detected in all 6 TMT reporter ion channels.

We see that for all QExactive and QExactive+ methods tested, where we varied the isolation width, almost all trypanosome and mouse peptides were identified with an intensity from all 6 TMT reporter ions (Supplementary Figure 11). For data acquired on a Fusion MS, using the SPS-MS3 methodology, we see the bulk of experiments identifying a signal for all 6 TMT reporter ions in >65% of trypanosome peptides and almost 100% of mouse peptides. This group includes experiments where the HCD collision energy, the maximum injection time for MS3 scan, or the ion selection method was modified. We also observe experiments where 10 or 8 SPS notches were utilised. A further reduction in trypanosome peptides identified in all 6 TMT reporter ions is detected by: decreasing the number of SPS notches used (N08, N06, N04, N03 and N02); decreasing the isolation width for precursor ion selection to 0.8 *m/z* for the MS3 scan (MS 0.8); and decreasing the AGC target at 10 notches (10N 5e4 and 2.5e4) or 2 notches. However, we also observe that as we decrease the number of trypanosome peptides with reliably identified signal for all TMT reporter ions, we also begin to see a large drop this category of mouse peptides too. Just under 25% of mouse peptides are detected in all 6 TMT reporter channels for instance. This result highlights that there is a fine balance to be struck between producing accurate quantification of measured ratios using TMT (by trying to reduce the effect of compression), and producing precise and reliable quantification of measured ratios (by increasing the amount of material or signal reaching the Orbitrap mass analyser).

In summary, we envision two situations for the quantitation of proteomic samples using isobaric tag based quantitation. The first, where sample material is plentiful and the amount of peptide available for analysis is not limiting, it would be worthwhile to reduce the effects of compression by decreasing the number of SPS-notches utilised and decreasing the isolation width for selecting the precursor ion for the

MS3 quantitative scan, while increasing the amount of sample loaded per run. These parameters would allow for a reduction of the effect of compression with high sample loads. We would not recommend reducing the AGC target below $1x10^5$ as this would consistently reduce the amount of sample analysed on the Orbitrap mass analyser, resulting in inaccurate quantification. In the second situation, where the amount of sample and peptide available is limiting, it would be worthwhile to maximise the number of SPS-notches used, increase the isolation width for the precursor ion selection for the quantitative MS3 scan, and increase the maximum injection time for the MS3 scan. In this case we are looking to increase the amount of sample that reaches the Orbitrap mass analyser to allow for more precise quantitation of the material which is available.