

Novel Mitochondrial Metabolism in Excavate Protists

Amber Carol Lynch

This thesis is submitted in partial fulfilment of the requirements for the
degree of Doctor of Philosophy

Lancaster University

January 2016

Abstract

The Excavata super-group contains a massively diverse range of protists that inhabit a plethora of different trophic niches. Many excavates are free-living and little-studied, but there are also many examples of parasitism evolving multiple times within the group. *Trypanosoma brucei* is a euglenozoan parasite that is responsible for causing Human African trypanosomiasis (HAT) and Nagana in livestock. It has a complex heteroxenous life cycle that involves a variety of complex morphological and biochemical changes to the cell as it switches from a mammalian to an insect host. Within the mid-gut of its tsetse fly vector, *T. brucei* replicates as a procyclic trypomastigote; these cells are readily cultured and can be subjected to genetic manipulation in the laboratory. Despite the complete genome sequence being available, there are still many novel proteins within *T. brucei* that have yet to be assigned a function within a cell. *T. brucei* appears to have an extensive array of histidine phosphatases, many of which are uncharacterised (Brown, 2011). The histidine phosphatases are an ancient, ubiquitous enzyme superfamily that catalyse the dephosphorylation of a substrate, and function is critically dependent upon an active site histidine. Given that adaptation to parasitism is classically associated with metabolic stream-lining, it is intriguing that *T. brucei* has retained so many histidine phosphatases.

Two of these histidine phosphatases appear to have arisen by gene duplication, although one appears to have degenerated and subsequently has lost the essential catalytically-active histidine. Through phylogenetic analyses and contemporary molecular characterisation techniques, I report the progress made on attempts to assign function to this paralogous gene pair, both of which localise to the mitochondria in procyclic *T. brucei*. My efforts to unsuccessfully delete genes encoding these histidine phosphatases from the genome of procyclic *T. brucei* suggests they are essential.

Another excavate for which genome data is freely available is the free-living heterolobosean *Naelgeria gruberi*. The sequencing and annotation of the *N. gruberi* genome (Fritz-Laylin et al., 2010) revealed a surprising variety of anaerobic metabolic pathways potentially available to the cell, when encountering anoxic conditions. Using *T. brucei* as a host for heterologous expression, I tested the mitochondrial candidature of several of these *N. gruberi* novel proteins associated with anaerobic metabolism. Since the Heterolobosea are sister to the Euglenozoa, evidence for the expression and localisation of *N. gruberi* proteins in *T. brucei* is a good indication of where these proteins may localise within *N. gruberi* cells. These analyses showing mitochondrial localisation of enzymes associated with anaerobic metabolism point to a currently highly unusual repertoire of metabolic flexibility within the mitochondria of a eukaryotic heterotroph.

Declaration

I declare that the content presented in this thesis is my own work, except where explicitly stated, and has not been submitted in substantially the same form for the award of a higher degree elsewhere.

Acknowledgements

I would like to kindly thank my supervisors, Dr Michael Ginger and Dr Paul McKean for all the help, guidance, support and opportunities they have provided throughout my time at Lancaster. It has been a pleasure to be a member of your labs, and I couldn't have wished for better supervisors. I would like to give a special thanks to Dr Jane Andre and Dr Jon Moran for their invaluable help in the lab and their ongoing friendships.

Thank you to Dr Fiona Benson, Dr Mick Urbaniak, Dr Derek Gatherer, Dr Jackie Parry, Janice Drinkall and Dr Steve Roberts for their help and advice, and their kind generosity with reagents and equipment.

A huge thank you to some of the greatest lab friends anyone could ever wish for over the years! Gillian, Xin, Harriet, Kurimun, Laura, Hollian, you are all fabulous and are responsible for some of my best times at Lancaster... And a special thanks to Louise Kerry, my trypanosome friend.

This project was partially funded by the Faculty of Health and Medicine, and this financial support was invaluable to me. However, the remaining funding was provided by my spectacular grandparents, Carol and Jerry Lynch. I wouldn't be where I am today without you, you know how much you've both done for me. Thank you to you both, I love you so much. And finally thank you to my partner Dave Moran, who has been nothing but loving, patient, caring and supportive throughout these challenging years!

Table of Contents

1.	General Introduction	1
1.1	The diversity and origin of eukaryotic life	1
1.2	The radiation of eukaryotic lineages	9
1.3	Diversity of the Excavata	19
1.4	Mitochondrial diversity	25
1.5	Life cycle biology and metabolism of <i>Trypanosoma brucei</i>	34
1.6	Ecology of <i>Naegleria</i>	43
1.7	Overview of thesis results	46
2.	Materials and Methods	47
2.1	Buffers and solutions	47
2.2	Antibiotic and drug stock solutions	50
2.3	Antibodies	51
2.4	Oligonucleotide primers	52
2.5	Polymerase chain reaction	53
2.6	Agarose gel electrophoresis	53
2.7	DNA ligation	54
2.8	Transformations of <i>E. coli</i>	54
2.9	Extraction and purification of plasmid DNA	55
2.10	Ethanol precipitation of plasmid DNA	55
2.11	Restriction digests of DNA	55
2.12	Gel extraction	56
2.13	Induction of recombinant protein expression	56
2.14	Determining protein solubility	57
2.15	Protein sample preparation for IMAC purification under denaturing conditions	57
2.16	IMAC protein purification under denaturing conditions	57
2.17	Protein sample preparation for benchtop purification	58
2.18	Benchtop protein purification under denaturing conditions	58
2.19	Polyclonal antibody production	58
2.20	SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)	59
2.21	Silver staining	59

2.22	Southern blot analyses	60
2.23	Western blot analyses	61
2.24	<i>Naegleria gruberi</i> (NEG-M) cell culture	62
2.25	<i>Naegleria gruberi</i> (NEG-M) protein samples	62
2.26	<i>Trypanosoma brucei</i> cell culture	62
2.27	Transfection of DNA into trypanosomes	63
2.28	Trypanosome protein samples	63
2.29	Isolation of <i>T. brucei</i> genomic DNA	64
2.30	Preparation of slides for fluorescence microscopy	64
2.31	MitoTracker® probing	65
2.32	Gene knockout in trypanosomes	65
2.33	Co-immunoprecipitation	67
2.34	Culturing of <i>T. brucei</i> in SILAC media	67
2.35	Sample preparation for mass spectrometry	
	2.35.1 BioID sample preparation	69
	2.35.2 Co-immunoprecipitation sample preparation	71
2.36	Liquid chromatography–mass spectrometry	73
2.37	Analysis of mass spectrometry data	73
2.38	Analysis of protist FeFe-hydrogenases	73
2.39	Analysis of microarray data	74
2.40	Assembling an Access database for the analysis of TbHP40 and TbHP30 evolution	75
2.41	Protein targeting predictions	76
2.42	Homology modelling	76
3.	Bioinformatics analysis of novel, paralogous histidine phosphatases	79
3.1	An overview of the histidine phosphatase superfamily	79
3.2	TbHP40 and TbHP30 are mitochondrial proteins	82
3.3	Phylogenomic distribution of TbHP40 and TbHP30 orthologues	83
3.4	Homology modelling	86
4.	Evidence for the essentiality of TbHP40 and TbHP30	106
4.1	The generation of a TbHP40/TbHP30 ^{+/-} heterozygous cell line	106
4.2	Generation of a TbHP40/TbHP30 ^{-/-} - attempt 1	108
4.3	Generation of a TbHP40/TbHP30 ^{-/-} - attempt 2	109

4.4	Generation of a <i>TbHP40/TbHP30</i> ^{-/-} - attempt 3	112
4.5	Progress with the generation of a conditional <i>TbHP40</i> ^{-/-} cell line	114
5.	Identifying interacting partners for the novel histidine phosphatases <i>TbHP40</i> and <i>TbHP30</i>	126
5.1	A proximity-dependent biotinylation approach to identifying candidate interacting partners for <i>TbHP40</i>	128
5.2	Identifying potential interacting partners of <i>TbHP40</i> using a co-immunoprecipitation approach	133
5.3	Summary of attempts to identify interacting partners of <i>TbHP40</i> and <i>TbHP30</i>	138
6.	Heterologous expression in <i>Trypanosoma brucei</i> of candidate mitochondrial proteins from <i>Naegleria gruberi</i>	157
6.1	Bioinformatic analysis of a range of eukaryotic FeFe-hydrogenases	158
6.2	Antibody-based detection of <i>N. gruberi</i> FeFe-hydrogenase	160
6.3	Heterologous expression of <i>N. gruberi</i> enzymes in <i>T. brucei</i>	162
6.4	Summary of characterisation attempts of <i>N. gruberi</i> novel mitochondrial proteins	166
7.	General Discussion	183
8.	Appendices	191
8.1	Appendix I – abbreviations from Figure 1.12	191
8.2	Appendix II – restriction maps for plasmids used in this study	191
9.	References	196

Abbreviations

a.a.	Amino acids
ABC	Ammonium bicarbonate
AdoMetDC	S-adenosylmethionine decarboxylase
AMP	Adenosine monophosphate
ASCT	Acetate:succinate CoA-transferase
BioID	Proximity-dependent biotinylation identification
BLAST	Basic Local Alignment Search Tool
blast	Blasticidin
blastp	Protein-protein basic local alignment search tool
<i>Bleo</i>	Bleomycin
bp	Base pairs
BPP	Benchtop purification protocol
Co-IP	Co-immunoprecipitation
DAPI	4', 6-diamidino-2-phenylindole
DHS	Deoxyhypusine synthase
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EB	Elution buffer
ELISA	Enzyme-linked immunosorbent assay
euBACs	Universal eukaryotic genes of bacterial origin
gDNA	Genomic deoxyribose nucleic acid
GFP	Green fluorescence protein
HA	Human influenza hemagglutinin
HAT	Human African trypanosomiasis
HK	Hexokinase
Hmd	Methylenetetrahydromethanopterin
HRP	Horseradish peroxidase
hyg	Hygromycin
IAA	Iodoacetamide
IG	In-gel
IMAC	Immobilised metal-ion affinity chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside

JGI	Joint Genome Institute
kDNA	Kinetoplast deoxyribose nucleic acid
KO	Knockout
LACA	Last animal common ancestor
LBA	Long branch attraction
LCA	Last common ancestor
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LECA	Last eukaryotic common ancestor
LS	Long slender
MROs	Mitochondrial-related organelles
MS	Mass spectrometry
mtDNA	Mitochondrial deoxyribose nucleic acid
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NCBI	National Centre for Biotechnology Information
NIT	Nitrate-nitrite sensing
ORF	Open reading frame
OXPHOS	Oxidative phosphorylation
PAM	Primary amoebic meningoencephalitis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PDH	pyruvate dehydrogenase
PEP	phosphoenolpyruvate
Pfam	Protein family
PFK	Phosphofructokinase
PFO	Pyruvate:ferredoxin oxidoreductase
PGK	Phosphoglycerate kinase
phleo	Phleomycin
PMSF	Phenylmethylsulfonyl fluoride
PPDK	Pyruvate phosphate dikinase
puro	Puromycin
PYK	Pyruvate kinase
RMC	Relaxed molecular clock
S/N	Supernatant
SAR	Stramenopiles Alveolata Rhizaria

SARP	Stramenopiles Alveolata Rhizaria Plantae
SB	Southern blot
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SILAC	Stable isotope labelling by amino acids in cell culture
SMART	Simple Modular Architecture Research Tool
SOC	Super optimal broth
SS	Short stumpy
SSU RNA	Small sub-unit ribonucleic acid
Sts-1	Suppressor of T-cell receptor signalling 1
Tax ID	Taxonomy identification number
TFA	Trifluoroacetic acid
TIGAR	Tumour protein 53-induced glycolysis and apoptosis regulator
UIR	Upstream intergenic region
UTR	Untranslated region
VSG	Variant surface glycoprotein
ZMG	Zimmerman's post-fusion medium

Chapter 1

General Introduction

The research presented in my thesis focuses upon experiments to study novel aspects of mitochondrial biology in excavate protists. To provide context for the results presented in chapters 3-6, my introductory chapter covers a variety of topics:

- An overview of eukaryotic evolution
- An introduction to the Excavata
- A discussion of mitochondrial diversity
- Lifecycles and relevant essential biology of *Naegleria gruberi* and *Trypanosoma brucei*

1.1 The diversity and origin of eukaryotic life

Life has been present on Earth for at least the past 3.5 billion years (Schopf, 2006). Microfossil, molecular biomarker and isotopic evidence dictate that the earliest life-forms on the planet belong to the prokaryotic lineages; reviewed by Schopf (2006) (see Figure 1.1). The evolution of eukaryotes is believed to have occurred later in Earth's long history.

Microfossils also provide evidence for the existence of early eukaryotic life, and to some extent, provide a record for the evolution of eukaryotic cells. Some of the earliest eukaryotes are thought to be ~1.8 billion years old (Lamb et al., 2009), which dates back to the mid-Paleoproterozoic era (See Figure 1.1). Microfossil examples of these early eukaryotes have been found across the planet, including northern China (Lamb et al., 2009), West Africa (El Albani et al., 2014), northern Australia (Javaux et al., 2003), Siberia (Nagovitsin, 2009) and Canada (Cohen et al., 2011). These microfossils are necessary for inferring the lives and complexity of some of the earliest eukaryotes. All eukaryotic life found on the planet today evolved from the elusive Last Ekaryotic Common Ancessor (LECA) (*i.e.* the root of the eukaryotic tree of life). All descendants from LECA are referred to as crown-group eukaryotes (Butterfield, 2015; Eme et al., 2014). All extinct eukaryotic lineages that existed before LECA (after a eukaryotic ancestor emerged from within the Archaea) are referred to as stem-group eukaryotes (Eme et al., 2014).

Within the fossil record, both prokaryotic and eukaryotic cells can yield microfossils; however the likely soft bodies and small sizes of organisms yielding these microfossils make this fossil record difficult to study. Some microfossils can be identified as cyanobacteria, whereas others can be assigned eukaryotic taxonomic classifications based on modern protistan lineages. However, the majority of microfossils that are characterised end up classified as 'acritarchs' (Butterfield, 2005). Using the original language of Evitt (1963), acritarchs were defined as:

"small microfossils of unknown and probably varied biological affinities consisting of a central cavity enclosed by a wall of single or multiple layers and of chiefly organic composition; symmetry, shape, structure and ornamentation varied; central cavity closed or communicating with the exterior by varied means" (Evitt, 1963).

Eukaryotic microfossils are initially identified by morphologies that are classically only associated with eukaryotic cells. Size is usually the first indication of whether an organism is prokaryotic or eukaryotic, since eukaryotic cells are usually substantially larger (Butterfield, 2015). However, size alone cannot determine biological classification. Modern cyanobacteria can measure up to 60 μm in diameter, and certain filamentous cyanobacteria, such as *Microcystis*, can secrete extracellular sheaths (Kessel and Eloff, 1975). This can result in microfossils that are deceptively large enough to be interpreted as being potentially eukaryotic (Butterfield, 2015). On the contrary, extant pico-eukaryotes can measure $<1\text{-}3\ \mu\text{m}$ in diameter (Baldauf, 2003), and so could easily be mistaken for prokaryotic cells, based on their size alone (van Ooijen et al., 2012).

Further characterisation of eukaryotic microfossils relies upon the identification of mineralised cell ultrastructures from early protists that were resistant to decay and accumulated in sedimentary rock layers (Knoll, 2014). Evidence of complex cell wall morphologies, excystment and vesicle structures are indicative of the organism possessing a cytoskeleton, which is a cytological feature not present in prokaryotes. For instance, microfossils found in the Yukon and dated at ~800 million years old (mid-Neoproterozoic era), exhibit a scale structure on the fossil surface (Cohen et al., 2011) (Figure 1.2 A). This was taken as evidence of the organism containing Golgi-like structures that would have been involved in the synthesis of scales, as is the case in some modern-day protists, such as Haptophyta algae (e.g. *Pleurochrystis*) (Melkonian et al., 1991). The suggested presence of Golgi apparatus indicates a morphological complexity not seen in prokaryotes, and a complex cytoskeleton and vesicular trafficking system would also be required in order to aid

in the translocation of scales from Golgi to the cell surface. Chuar rocks from the Grand Canyon (USA) yielded vase-shaped microfossils that were ~742 million years old (see Figure 1.1). These microfossils were so well preserved and morphologically diverse that they could be taxonomically defined as testate amoebae (Amoebozoa or Rhizaria) (Porter et al., 2003) (see Figure 1.2 B). Eukaryotes, but not the vast majority of prokaryotes, produce C₂₇₋₂₉ sterols that degrade to give stable steranes and sterenes which, unlike proteins and nucleic acids, preserve well in sedimentary rocks (Brocks and Summons, 2003; Knoll, 2014). These can be useful biomarker molecules and through structural and isotopic analysis, can be very informative with regard to understanding ancient cellular diversity, the trophic associations of ancient organisms, and the environmental conditions in which these organisms lived (Brocks and Summons, 2003). For instance, the temperature of the surrounding environment can be predicted based on the preserved biomarkers. Organisms adapted to warm temperatures produce different, distinctive biomarkers compared to organisms adapted to cooler temperatures, since water temperatures have an effect on the levels of dissolved CO₂, and therefore an impact on the carbon assimilation by photosynthetic organisms (such as green algae) (Brocks and Summons, 2003).

Some of the earliest eukaryotic microfossils known currently were found in the Changzhougou Formation, North China and are estimated to be ~1.8 billion years old (Lamb et al., 2009). Thirteen distinct eukaryotic morphologies were identified in this study, and provided a glimpse into the possible diversification of eukaryotes by the Paleoproterozoic era (Lamb et al., 2009) (see Figure 1.1). Other microfossils from the the Biacaoping and Beidajian Formations, North China, and the Deonar Formation in India, are thought to be ~1.6 billion years old, and thus from the late Paleoproterozoic era. Some of the microfossils isolated from both regions were identified as eukaryotic 'acanthomorphic acritarchs' (Butterfield, 2015), including specimens from the genus *Tappania*. Other *Tappania* microfossils have been identified in northern Australia (Javaux et al., 2003), north-western USA (Butterfield, 2015) and central Siberia (Nagovitsin, 2009). These fossils are thought to be from the early-middle Mesoproterozoic era (~1.5-1.2 billion years ago). This indicates that members of the *Tappania* genus were widespread and persistent in the environment for at least 200-300 million years (Butterfield, 2015). *Tappania* is considered to have been a large (30-60 µm) multicellular eukaryote with a sophisticated cytoskeleton and a complex, highly variable morphology (Butterfield, 2005). It was ocean-dwelling with various filamentous, branched tubular extensions emanating from the cell surface (Butterfield, 2005, 2015). Although no direct modern counterparts to *Tappania* survives, there have been tenuous suggestions of

assigning it to the fungal kingdom (Butterfield, 2005), although there is still not enough convincing evidence to draw firmer conclusions on this taxonomic placement (Butterfield, 2015). Another ancient 'acanthomorphic acritarch' isolated from samples from the Biacaoping and Beidajian Formations, named *Shuiyousphaeridium*, was at 100-300 μm , much larger in size than *Tappania* (Butterfield, 2015). It was covered with fine protrusions that emanated from a cell wall, had an inner vesicle wall made up of a mosaic of scale-like plates, and another vesicle referred to as a 'reproductive cyst'- the proposed reproductive cyst is still seen in extant dinoflagellates (Butterfield, 2015). Again, there is no evidence for any remaining *Shuiyousphaeridium* sub-clades existing today. However, the variation in the structures observed between *Tappania* and *Shuiyousphaeridium* suggests eukaryotic life had diversified considerably with respect to both morphological and developmental complexity by the beginning of the Mesoproterozoic era. Both organisms would have contained a nucleus, a complex cytoskeleton and signalling network, functional mitochondria (or perhaps mitochondrial-related organelles) capable of energy generation and a cell wall that would likely have been similar to structures found in today's plants and fungi. Other notable fossilised eukaryotes that can be confidently assigned to extant eukaryotic groups include the Mesoproterozoic red alga, *Bangiomorpha*, from northern Canada (Butterfield, 2000), and the Neoproterozoic green algae *Palaeastrum* and *Proterocladus*, both found in Svalbard (Butterfield and Knoll, 1984).

Fossils can provide approximate dates on when these ancient eukaryotes lived and subsequently became extinct. Yet, fossil evidence alone provides minimal or no data on the divergences which took place as crown eukaryotic lineages evolved. However, quantitative estimates of species divergence can be made using molecular clocks. The original molecular clock hypothesis was based on the study of differences in haemoglobin sequences in humans (Zuckerandl and Pauling, 1965). The concept was devised that the evolutionary rate of sequence divergence between orthologous proteins from different species occurs at a relatively constant rate (Zuckerandl and Pauling, 1965). This would therefore result in molecular clock-derived estimates providing a powerful molecular phylogenetics tool for investigating at which points in history lineages diverged (Hug and Roger, 2007; Roger and Hug, 2006). However, due to variation in mutation rates within lineages, it has proved difficult to generate a molecular clock model that produces viable time estimates for speciation events generally. Moreover, results from molecular clock estimates often contradict the palaeontological evidence provided by fossils, reviewed by Eme et al., (2014); Hug and Roger, (2007); Roger and Hug, (2006). To account for the variations in evolutionary

rate, a variety of 'relaxed molecular clock' (RMC) methods were subsequently developed. RMCs combine molecular phylogeny estimates with fossil records and molecular biomarkers, to provide calibration points for the clock estimates (Eme et al., 2014). Only fossils that have been assigned definitive crown-group lineage classifications should be used as molecular clock time constraints. Often, it has proved more difficult to associate the sparse, patchy Proterozoic fossil specimens with extant crown-group eukaryotes than the much more extensive Phanerozoic specimens (~0.6 billion years old, appearing after the radiation of the crown-group lineages - see Figure 1.1) (Eme et al., 2014).

Despite the refinements made to RMC approaches and the original molecular clock hypothesis, RMC estimates are still riddled with errors and uncertainties (reviewed by Eme et al., (2014); Hug and Roger, (2007); Roger and Hug, (2006)). Such errors may be caused by the differences in methods used in generating RMC estimates; in the selection of fossils and their assignments at the nodes of trees generated; consensus views on eukaryotic phylogeny at the time of any study also influences the accuracy of RMC estimates; the paucity in parts of the eukaryotic fossil record; and the difficulties surrounding fossilisation events in the first place (for example, many of the earliest unicellular post-LECA eukaryotes likely lacked hard structures suitable for fossilisation) (Eme et al., 2014). The dates assigned to the fossils themselves are also often open to question. There can be palaeontological uncertainties in the dates assigned to the rocks containing the fossils, and there can also be a time gap between the actual divergence date of individual eukaryotes and when those organisms were fossilised. Finally, genetic divergences invariably occur before detectable morphological changes in an organism. Thus, date estimates of a fossil based on palaeontological studies alone are often underestimated.

Post-LECA eukaryotes of the Paleoproterozoic period had evolved complex cytoskeletons and signalling networks that would have given them distinct selective advantages over prokaryotes (Fritz-Laylin et al., 2010b). For example, presence of elaborate internal cytoskeletons allows eukaryotic cells to grow much larger than prokaryotes. However, despite selective advantages, it apparently took eukaryotic organisms nearly half a billion years to fully exploit the significant attributes that they had over the prokaryotes, which continued to very much dominate the biosphere up until the late Mesoproterozoic period (Butterfield, 2007) (see Figure 1.1). The radiation of eukaryotic life into the crown lineages recognised today (see Figure 1.4) may have been delayed by environmental constraints such as reduced oxygen abundance (initially atmospheric oxygen, subsequently oceanic oxygen

abundance (Canfield, 1998)) and/or insufficient nitrogen fixation (Butterfield, 2015). Biogeochemical studies suggest that there was a major oxidation event on the planet's surface in the mid-Paleoproterozoic era (~2 billion years ago), although anoxic waters still persisted in the deep oceans until the Neoproterozoic era (Canfield, 1998). Since early eukaryotic life was known to exist in the early-Mesoproterozoic oceans (Knoll et al., 2006), this could provide some explanation as to why there was a delay before the rapid diversification of eukaryotes. Another difficulty post-LECA eukaryotes may have encountered was breaking through into biological niches where the predominant inhabitants were well-established prokaryotic colonies. Most RMC estimates have agreed that radiation of the major eukaryotic crown-groups began during the late Mesoproterozoic/early Neoproterozoic period, after the post-LECA lag period (Eme et al., 2014; Knoll, 2014). Most RMC estimates also agree that the diversification of the eukaryotes happened rapidly over a relatively short period of time (Eme et al., 2014).

Figure 1.1

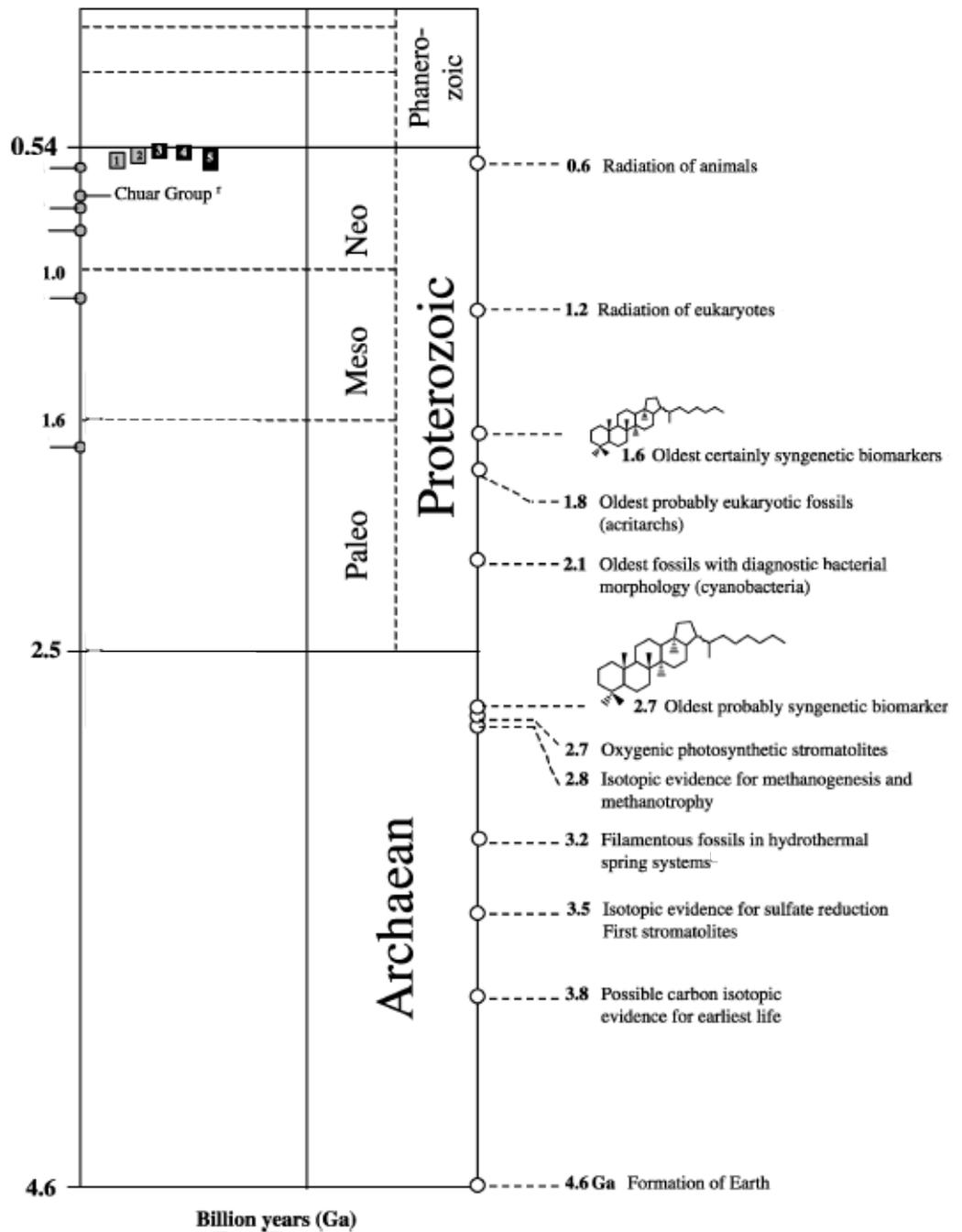


Figure 1.1. A geological timeline highlighting important biological events in Earth’s history. The first eukaryotic life on Earth was thought to have appeared during the mid-Paleoproterozoic period, ~1.8 billion years ago. Towards the end of the Mesoproterozoic era, a major radiation event of eukaryotes was thought to have occurred. Radiation of extant animal lineages is thought to have occurred ~0.6 billion years ago. Modified from Brocks and Summons, (2003).

Figure 1.2

A.

B.

C.

D.

Figure 1.2. A selection of eukaryotic microfossils from the late Paleoproterozoic era and the mid-Neoproterozoic era.

A-B. *Characodictyon skolopium* and *Trigonocyrrillium horodyskii* respectively; microfossils that show the diversity and complexity of the cytoskeletons in likely eukaryotes from the mid-Neoproterozoic era. Reproduced from Butterfield, (2015).

C-D. *Tappania plana* and *Shuiyousphaeridium macroreticulatum* respectively, some of the oldest acanthomorphic acritarch microfossils discovered to date, originating from the late Paleoproterozoic era. Reproduced from Porter (2003) and Knoll (2006).

1.2 The radiation of eukaryotic lineages

Original molecular phylogenetic predictions of the three domains of life (Eukaryota, Bacteria and Archaea) were based on small subunit RNA (SSU RNA) sequence analyses (Sogin and Gunderson, 1987; Woese et al., 1990). This was due to SSU RNAs being one of the most evolutionary conserved molecules across all forms of life on Earth, and thought not to undergo lateral gene transfer (Sogin and Gunderson, 1987). SSU RNAs are also easy to amplify by PCR and contain both highly conserved and hypervariable regions (Burki, 2014). SSU RNA-based phylogenies recovered several apparent ancient ‘deeply diverging’ protist branches (the diplomonads, microsporidians, parabasalids – eukaryotic lineages that were originally termed ‘amitochondriate’ and were thought to have diverged before the eukaryotic mitochondrial acquisition event) reviewed by Burki, (2014). From the SSU RNA phylogenies, most remaining eukaryotic lineages then emerged, clustering around a highly branched, poorly resolved ‘crown structure’ (including animals, plants, fungi, alveolates and stramenopiles). This ‘crown structure’ was thought to be due to these lineages having diverged almost simultaneously from one another. However, the structures of SSU RNA phylogenetic trees are now known to be prone to a significant error, known as long branch attraction (LBA) (Eme et al., 2014; Philippe and Germot, 2000; Roger et al., 1999). Whilst sequence similarity of genes can infer the distance of relationships between organisms, there are only a select number of sites within a gene that can undergo mutation without deleterious effects on the organism, and therefore only a limited number of outcomes (*i.e.* changes in nucleotides and/or protein coding sequence) (Baldauf, 2003). Changes in sequence code accumulate but in distantly-related species where that evolution occurs at a faster rate, similar extensive mutation can occur independently, leading to inappropriate clustering (attraction) of long branches (Baldauf, 2003). Protists originally deemed to be ‘amitochondriate’ were also shown to contain degenerate forms of mitochondria or ‘mitochondrial-related organelles’ (MROs) such as hydrogenosomes or mitosomes (see Section 1.4), and any attraction to the base of the eukaryotic tree was not due to these lineages diverging before the rest of eukaryotes acquired mitochondria, but because of artificial long branched attractions to distant outgroups (reviewed by Burki, (2014); Embley and Martin, (2006)).

As phylogenetic methods developed due to the availability of multigene datasets and from sampling of increasingly diverse ranges of eukaryotes and as the appreciation of LBA improved, phylogenomic analyses led to suggestion of eukaryotic ‘super-groups’. These super-groups define the bulk of known eukaryotic biodiversity and further resolve the

phylogeny of the eukaryotes, particularly the vast diversity of protist lineages (Baldauf, 2003; Baldauf et al., 2000; Hampl et al., 2009; Simpson and Roger, 2004) (see Figure 1.3). The unrooted tree (a tree that makes no inference on the ancestral root, *i.e.* the origin of LECA) shown in Figure 1.3 highlights six super-groups: Opisthokonta, Amoebozoa, Plantae, Chromalveolata, Rhizaria and Excavata (Simpson and Roger, 2004). However, several groups of free-living heterotrophs (the examples shown in Figure 1.3 are the Apusomonads, Colloidiomonads and Centrohelid Heliozoa) lay outside of the defined super-groups, and therefore may be representatives of novel additional groups (Simpson and Roger, 2004). The Opisthokonta + Amoebozoa were sometimes referred to as 'unikonts' and everything else was referred to as the 'bikonts' (Baldauf, 2003; Cavalier-Smith, 2002; Hampl et al., 2009; Minge et al., 2009). The order with which eukaryotic groups diverged, including the identity of the earliest branching eukaryotic lineage, is still debated.

Further progression in methods since the turn of the 21st century led to the development of phylogenomics. Phylogenetics tended to be based on the comparison of one or several genes from various taxa, in order to infer their evolutionary relationships; whereas phylogenomics uses much larger multigene datasets for sequence alignments, in order to analyse complete or large segments of genomes. The acquisition of such large datasets has in part become possible due to next-generation sequencing, which is substantially faster and cheaper than Sanger sequencing methods used to sequence the eukaryotes and prokaryotes chosen first for complex genomic analyses. This has also resulted in a far wider diversity of taxa being sampled, rather than just organisms with medical or financial relevance to humans. A representative species from each major lineage was fully sequenced by 2012 (Burki, 2014). Phylogenomic studies generally continue to support the existence of super-groups, although there have been some re-categorising of the tree presented by Simpson and Roger, (2004) (Figure 1.3). The current view, based on multigene analyses, consists of five clades: Opisthokonta, Amoebozoa, SAR (Stramenopiles/Alveolata/Rhizaria), Excavata and Archaeplastida (often referred to as Plantae) (Adl et al., 2012) (Figure 1.4). Many of these can be further categorised into mega-groups – Diaphoretickes or SARP (Stramenopiles/Alveolata/Rhizaria/Plantae), Amorphea (Opisthokonta/Amoebozoa – formally the unikonts) and Excavata (Adl et al., 2012) (Table 1.1). The Amorphea and Diaphoretickes together form the so-called 'Neozoa' (Cavalier-Smith, 2010), which proposed to be sister group to the Excavata (Cavalier-Smith, 2010; He et al., 2014).

Most multigene phylogeny studies can assign most of the known eukaryotes to one of these three mega-groups (He et al., 2014). However, phylogenomic analyses do not come without

their own errors and limitations. For example, in 2009 Hampl et al. published an often cited study that analysed the possible monophyly of the Excavata and its relationships to other eukaryotic super-groups. These authors produced phylogenetic trees using two different methods, based on multigene datasets containing 143 genes from 48 taxa, with representatives from all of the six major super-groups. No outlier prokaryotic groups were incorporated into the dataset, due to fears of LBA. Their first analyses recovered three clades of excavates; Discoba, Malawimonads and Metamonada (Adl et al., 2012), but did not support the monophyly of the Excavata. The *Malawimonas* taxa branched separately to the rest of the Excavata and grouped with the unikonts with moderate statistical support (Figure 1.5 A) (Hampl et al., 2009). It was noted that many of the metamonad taxa had very long branches, for example *Spironucleus*, *Giardia* and *Trichomonas*, whereas the *Malawimonas* branches were very short in comparison. Hampl et al. hypothesised that these long branches could be responsible for the non-monophyly of the tree because of LBA. To test this theory, various approaches were used to manipulate the outcome of the tree. First, long-branching taxa were removed. To produce the tree seen in Figure 1.5 B, a total of 27 taxa were removed, including *Giardia*, *Leishmania*, *Trypanosoma*, *Tetrahymena*, *Blastocystis*, *Euglena* and *Reclinomonas* (Hampl et al., 2009). The reasoning for this approach was that the robustness of a group can still be tested, even with just one representative from that group present in the analysis. Upon the removal of *Spironucleus* and *Giardia*, the Excavata became monophyletic, and with the removal of *Trichomonas*, the statistical support for the monophyly greatly increased. Removal of the 27 taxa also supported the unikont/bikont divide. Another approach used to infer monophyly was to remove long-branches. By doing this, fast-evolving sequences that had accumulated the largest numbers of changes were deleted from alignments, rather than eliminating specific taxa. The tree seen in Figure 1.5 C was produced after 1750 of the longest-branched gene sequences were removed and this provided limited support for monophyly of the Excavata (Hampl et al., 2009). Analysis of the complete dataset available for the study did not yield monophyly of the Excavata. To infer monophyly, various sequences or indeed whole taxa had to be removed from the dataset. Whilst the approaches employed by Hampl et al., (2009) eventually resulted in a statistically supported monophyly of Excavata, it did not support the monophyly of many of the other super-groups such as the Chromalveolata and Rhizaria (now referred to as SAR, see Figure 1.4) and Archaeplastida. The positions of the representative taxa from these three groups moved around dramatically depending upon the version of the dataset analysed (Figure 1.5 A-C).

More recently, another dataset presented by He et al., (2014) was also used to analyse deep rooted phylogenies of the super-groups defined fully by Adl et al., (2012), and to also explore a probable position of the eukaryotic root. The dataset designed by He et al., (2014) consisted of 76 taxa containing 37 nuclear-encoded proteins of close bacterial origin (euBacs) that were present in LECA and had not undergone early gene duplications or lineage-specific losses. Prokaryotic outgroups of the closest bacterial relatives to the eukaryotes, which contained all 37 euBacs, were also included. Phylogenetic analysis produced a highly resolved tree which supported the monophyly of all the eukaryotic and prokaryotic clades represented, with high statistical significance (Figure 1.6 A) (He et al., 2014). The organisation of the three mega-groups, Amorphea, SA[R]P (Rhizaria was not represented), and Discoba (the parental taxon being the Excavata), were also supported in this tree. The Excavata were identified as the sister lineage to the Amorphea and SARP, meaning that the root of the tree lay between the Excavata and the neozoans (Figure 1.6 B). To eliminate the potential for any artefacts within this rooted tree (Figure 1.6 A), such as LBA, the euBac dataset was tested by removing any fast-evolving, long branching sequences and distant out-group sequences. The removal of the fast-evolving sequences produced an identical phylogenetic tree and neither control affected the positioning of the root. Despite the range of eukaryotes represented in this dataset, there were still a number of potentially important taxa that were absent due to inadequate sequence availability or poor taxon sampling. The diplomonads, microsporidians, parabasalids (*i.e.* excavate protists that possess degenerate forms of mitochondria) are also not represented; since these protists lack aerobic mitochondria, they therefore lack many, if not all of the 37 euBacs required for the dataset. The study from He et al., (2014) was important for highlighting the early divergence of the Excavata before the radiation of the lineages that resulted in the plants, animals and fungi, and also that the Excavata are crucial in the understanding of early eukaryotic evolution (He et al., 2014). Two of the most well-studied discobids are the *Trypanosoma* and *Leishmania* parasites, however, the only free-living discobid with a fully sequenced genome available remains *Naegleria gruberi* (Fritz-Laylin et al., 2010b).

The sampling of free-living organisms is vital for the study of the last common ancestor of all eukaryotes (Fritz-Laylin et al., 2010b; He et al., 2014). Medically-important parasites and multicellular plants and animals are often the first choices for gene sequencing. However, it is thought that parasite genomes are often stream-lined as a consequence of adaptation to parasitic niches (Berriman et al., 2005; Morrison et al., 2007), and are therefore subject to gene losses and high sequence divergence. This may or may not be an accurate

generalisation (Jackson et al., 2015; Pombert et al., 2014). However, parasite genomes are therefore not necessarily the most informative genomes to infer ancestral eukaryotic characteristics since many features may not be retained (Fritz-Laylin et al., 2010b). Genome analysis of free-living *Naegleria gruberi* suggested that LECA contained at least four thousand genes which are conserved and present in extant eukaryotes (Fritz-Laylin et al., 2011; Fritz-Laylin et al., 2010b). This analysis further highlights that LECA was far from 'primitive'. It is thought to have contained many features commonly associated with an extant 'classical' eukaryotic cell. LECA contained a complex signalling network, cytoskeleton and flagellum, a nucleus, Golgi apparatus, endoplasmic reticulum, mitochondria, underwent mitosis and meiosis and was heterotrophic (Butterfield, 2007; Fritz-Laylin et al., 2010b; Simpson and Roger, 2004).

In summary for Sections 1.1 and 1.2, when attempting to gain insight into the evolution of the eukaryotes, microfossil evidence is a good, if not patchy, first resource to turn to. Whilst fossil analyses come with their own plethora of limitations and difficulties, such as dating and taxonomic categorisation, they provide a direct account of early eukaryotic life that existed on the planet from more than a billion years ago. Fossils are also useful for providing age constraints and calibration points in molecular clock analyses, which are used for estimating the age of divergence events of eukaryotic lineages. However, the differences in methods used in constraining fossil calibrations for relaxed clock estimates can result in wildly different age estimates on divergence points (in terms of hundreds of millions of years) (Eme et al., 2014; Hug and Roger, 2007). To reduce the bias on age estimates, multiple fossil calibrations can be used in clock analyses from carefully chosen fossil sources that are accepted by the paleontological community to be well taxonomically defined and dated (Eme et al., 2014; Hug and Roger, 2007). By improving the fossil record, the accuracy of the clock estimates will also be improved. However, increasing the number of taxa in the dataset used to generate clock estimates does not necessarily have a significant impact on the confidence in the age estimates (Hug and Roger, 2007).

Conversely, to improve the resolution of modern phylogenies, wider taxonomic sampling is required (Fritz-Laylin et al., 2010b; He et al., 2014; Hug and Roger, 2007). However, as shown by Hampl et al., (2009), a multigene dataset that includes a wide diversity of taxa can result in a wide variety of phylogenetic trees being produced, dependent upon how the gene-set is handled.

Figure 1.3

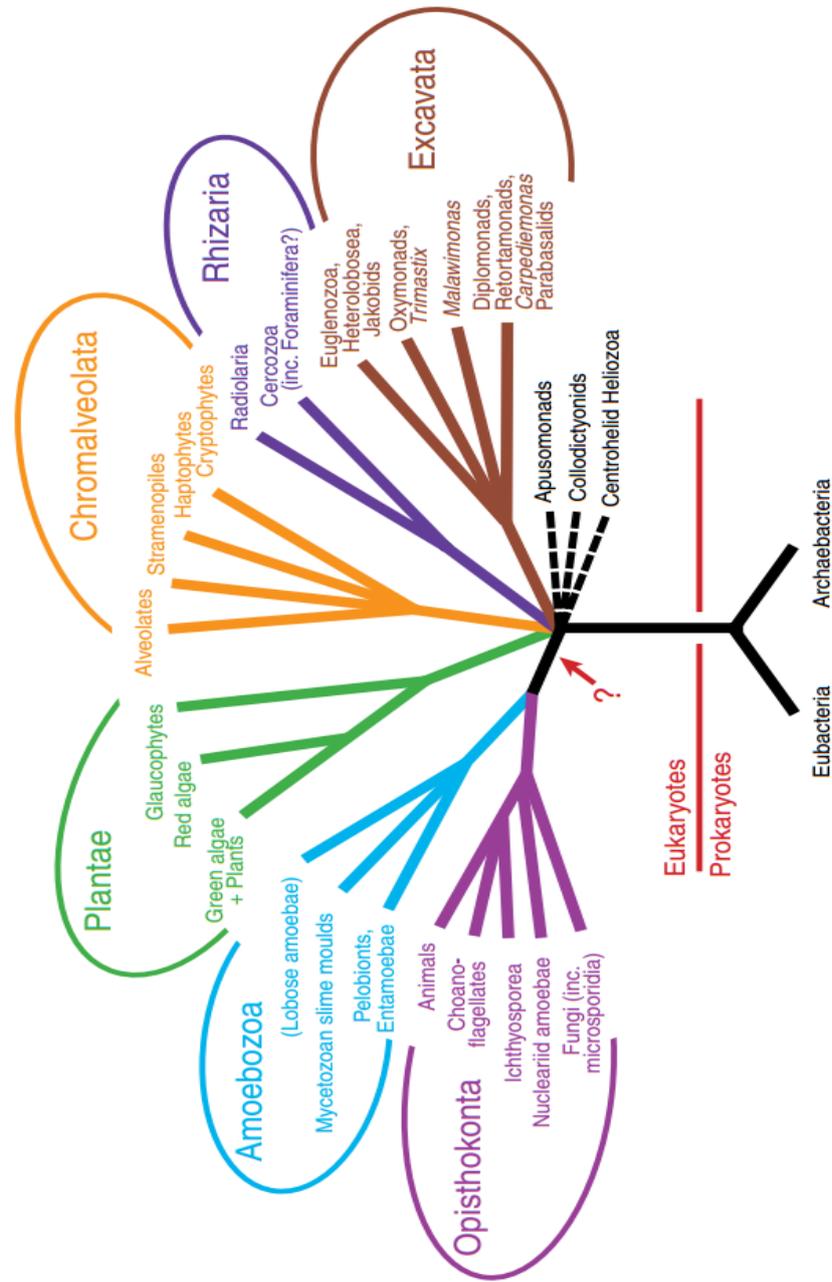


Figure 1.3. A diagram showing the assumed phylogenies of the eukaryotic tree, circa 2004. Reproduced from Simpson and Roger, (2004).

This view describes six major super-groups within the eukaryotes, along with the uncertain placement of the Apusomonads, Collodictyonoids and Centrohelid Heliozoa (shown in black). All eukaryotes are shown to stem from within the Eubacteria and Archaeobacteria. The position of the root of the tree is uncertain (LECA), but the red arrow and question mark indicates a possible location at the base of the branch leading to Opisthokonta and Amoebozoa.

Figure 1.4

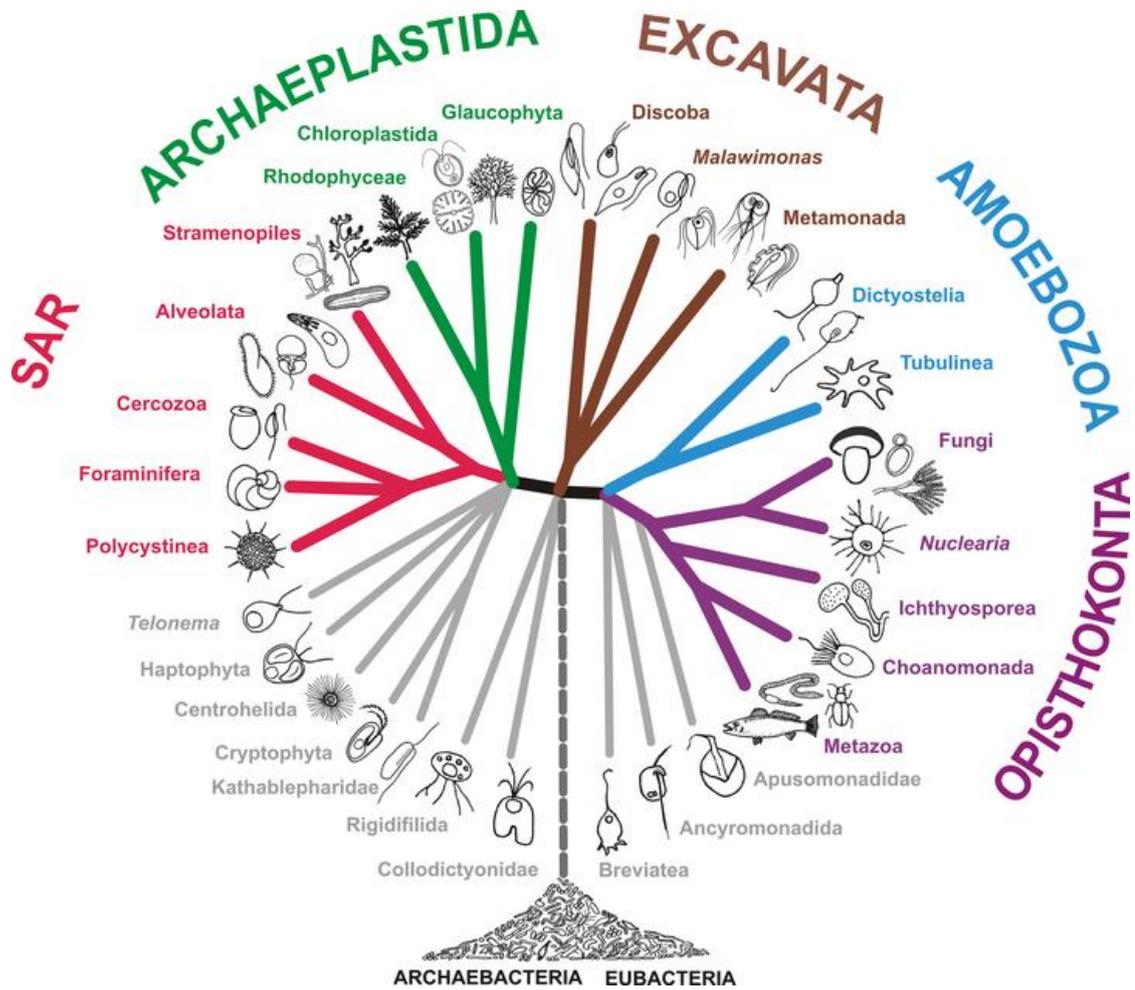


Figure 1.4. An updated view of the eukaryotic tree of life, circa 2012. Compared to Figure 1.3, the number of super-groups has been condensed down to five, and there is still uncertain placement of a variety of protists (lineages shown in grey). The root of the eukaryotic tree is still thought to stem from the Archaeobacteria and the Eubacteria. No inference to the origin of LECA is made in this diagram. Reproduced from Adl et al., (2012).

Table 1.1

Table 1.1. A table showing an alternative representation of the classification of eukaryotes. The super-groups shown in Figure 1.4 are here organised into the two mega-groups: Amorphea and Diaphoretickes. Reproduced from Adl et al., (2012).

	Super-groups	Examples		
Eukaryota	Amorphea	Amoebozoa	Tubulinea Mycetozoa	
		Opisthokonta	Fungi	Choanomonada
			Metazoa	Apusomonada
				Breviata
			Excavata	Metamonada Malawimonas Discoba
	Diaphoretickes		Cryptophyceae	
			Centrohelida	
			Telonemia	
			Haptophyta	
		Sar	Cercozoa	Foraminifera "Radiolaria"
				Alveolata
				Stramenopiles
	Archaeplastida	Glaucophyta Rhodophyceae Chloroplastida		
Incertae sedis Eukaryota		Incertae sedis, and table 3		

Figure 1.5

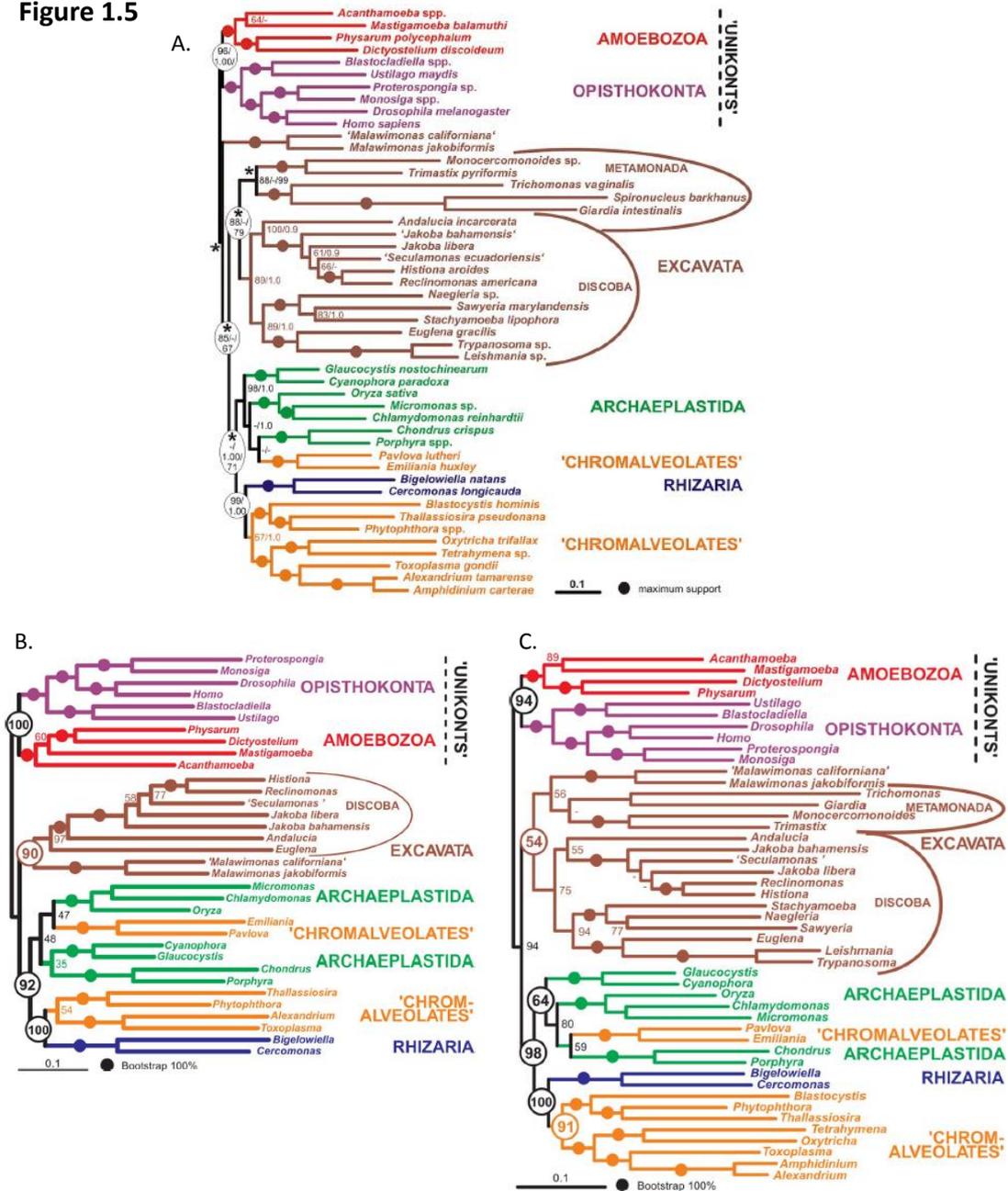


Figure 1.5. Phylogenetic tree analyses by Hampl et al., (2009) for the support of the monophyly of the Excavata super-group.

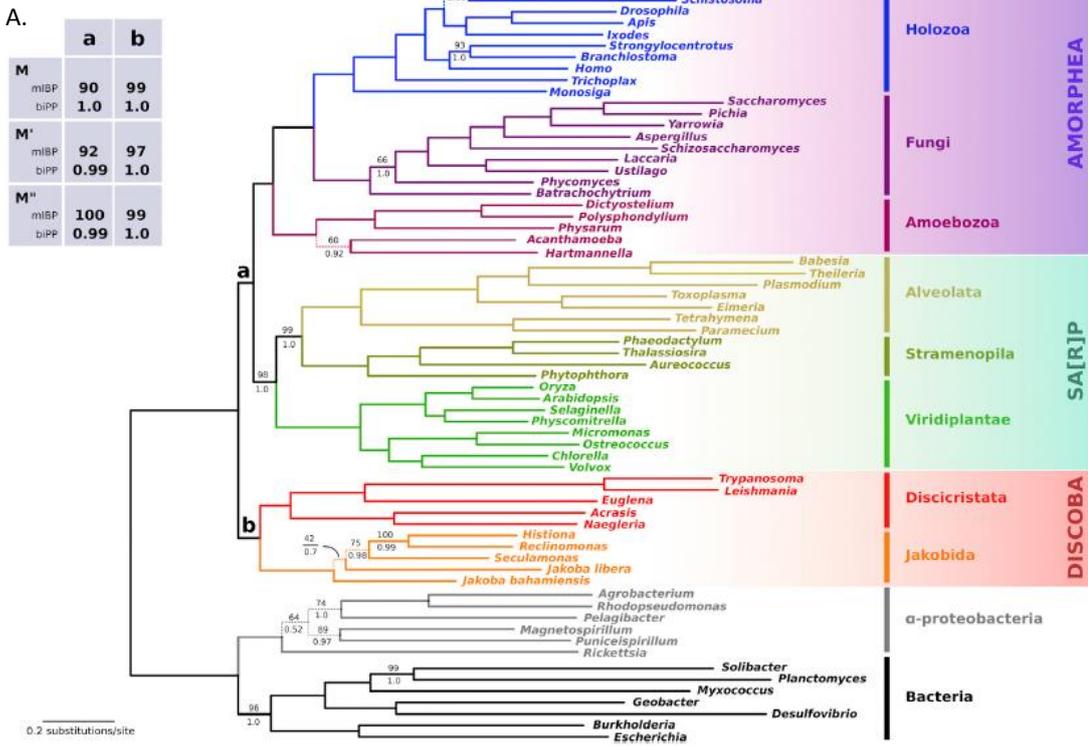
A. The phylogenetic tree produced using the main dataset. Prokaryotic sequences were not included over fears of LBA. This did not support the monophyly of the Excavata, which here consists of three separate branching clades.

B. The phylogenetic tree produced after the removal of long branch taxa. This tree supports the monophyly of the Excavata after the removal of 14 taxa.

C. The phylogenetic tree produced after the removal of the longest branching sequences in the dataset. This tree again supports the monophyly of the Excavata, after 1750 sequences were removed.

Reproduced from Hampl et al., (2009).

Figure 1.6



B.

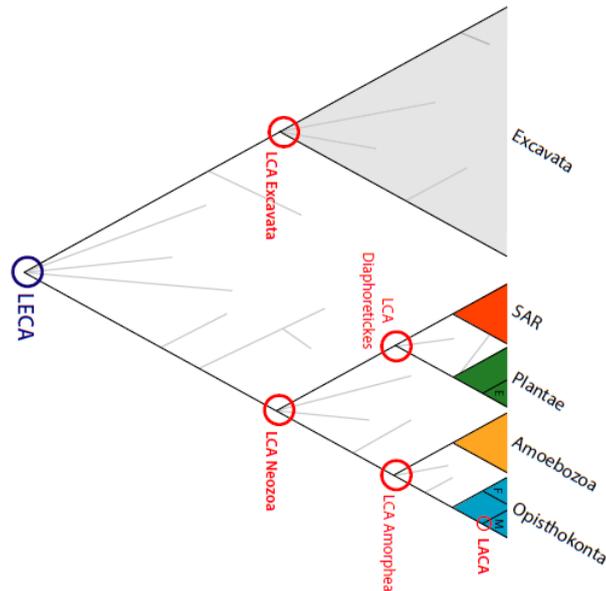


Figure 1.6. Phylogenetic support for the super-groups described by Adl et al., (2012), and insight into LECA.

A. A rooted phylogenetic tree generated using the euBac dataset by He et al., (2014). This shows strong monophyletic support for the described super-groups and a root falling between the neozoa and the excavates. Statistical support for the nodes denoted 'a' and 'b' from maximum-likelihood bootstrapping and Bayesian inference posterior probabilities is shown in the grey box. A red asterisk indicates the root position. Reproduced from He et al., (2014)

B. A cartoon schematic showing the divergence of the crown-group eukaryotes from LECA. LECA, last eukaryotic common ancestor; LCA, last common ancestor; LACA, last animal common ancestor; E, Embryophyta; F, Fungi; M, Metazoa. Reproduced from Butterfield, (2015).

1.3 Diversity of the Excavata

The protists on which my research focuses, *Naegleria gruberi* and *Trypanosoma brucei*, belong to the super-group Excavata. This group was proposed based a combination of characteristic morphological features and phylogenetic analyses (Adl et al., 2012; Cavalier-Smith, 2002; Hampl et al., 2009; He et al., 2014; Simpson, 2003), and encompasses a wide, ecologically diverse collection of free-living, symbiotic and parasitic protists.

In its classic form, the Excavata are comprised of three lineages: i) Metamonada, ii) Discoba and iii) Malawimonas (Figure 1.4) (Adl et al., 2012; Hampl et al., 2009). The excavates were originally defined according to their morphology, as unicellular flagellates with an anterior feeding groove (Patterson, 1999). 'Typical' excavates (in terms of morphology), which includes the Jakobids, Retortamonads, *Carpodimonas*, *Malawimonas* and *Trimastix* (see Table 1.2) all exhibit this characteristic feeding groove and highly similar cytoskeletons. Up to seven distinct cytoskeletal features are shared between various excavates, but none of these features have been observed to date in other eukaryotes (Simpson, 2003) (and listed in Table 1.3). For instance, diplomads and heteroloboseans have feeding grooves, but only some of the other cytoskeletal features exclusive to the excavate group: I fibres, a split R1, and in diplomads possibly singlet roots, too (Simpson, 2003). In contrast, oxymonads lack a feeding groove, but do contain I, B and C fibres, as well as singlet roots. Curiously, the cytoskeletons of the euglenozoans exhibit none of the cytoskeletal features attributed to the excavates and do not have feeding grooves (Simpson, 2003). However, phylogenetic analyses place the group as sister to the heteroloboseans and thus within the Excavata.

To look at each excavate lineage in turn, Metamonada consists of the Preaxostyla, Fornicata and Parabasalia, known members of which are all anaerobic or microaerophilic and have degenerate forms of mitochondria (Adl et al., 2012). Preaxostyla is made up of the oxymonads, which are symbionts found in the hindguts of termites and cockroaches that lack 'classical' mitochondria and *Trimastix*, a free-living anaerobe/microaerophile that have hydrogenosomes rather than classical aerobic mitochondria (Adl et al., 2012; Dacks et al., 2001; Hampl et al., 2008). The Fornicata consists of the diplomads, retortamonads and *Carpodimonas*. All members of these clades possess degenerate forms of mitochondria and inhabit microaerophilic/anaerobic environments (Kolisko et al., 2010; Tovar et al., 2003). Important human and animal parasites are found within the Fornicata, including *Giardia* and *Spironucleus*. Possibly the most well-known parabasalid is *Trichomonas vaginalis*, which causes a sexually transmitted disease in humans. All parabasalids are anaerobic, containing

hydrogenosomes rather than classical mitochondria and like the oxymonads, many are cockroach and termite commensals (Adl et al., 2012; Carpenter and Keeling, 2007). Thus, the last common ancestor of the metamonads was possibly aerobic, and the mitochondrial-related organelles found in many of these anaerobic eukaryotes are due to the degeneration of functional aerobic mitochondria (Cavalier-Smith, 2013; Embley, 2006; Embley and Martin, 2006; Jerlstrom-Hultqvist et al., 2013).

The Discoba consists of: i) Jakobida and ii) Discicristata. The Jakobida is made up of three further clades, *Jakoba*, *Andalucia* and the Histonidae, all of which are small, free-living bacteriovores that have two flagella. *Jakoba* and *Andalucia* are free-swimming, whereas *Reclinomonas* and *Histiona* (Histonidae) are sessile and exist within a lorica. The lorica is a protective structure produced by the organism, made up of a network of fibrils (Flavin and Nerad, 1993). The Histonidae have tubular mitochondrial cristae, whereas in *Jakoba* they are flattened and *Andalucia* lack cristae entirely (Lara et al., 2006; Simpson and Patterson, 2001).

The Discicristata is a much larger group that is comprised of the Heterolobosea and Euglenozoa, most members of which have flattened, often discoidal mitochondrial cristae (Adl et al., 2012). The Heterolobosea is made up primarily of free-living amoebae that move using eruptive short pseudopodia (Adl et al., 2012). Some, like *Psalteriomonas*, *Tetramitus*, *Naegleria* and the closely related *Willaertia*, have a non-feeding, temporary flagellate phase that is not essential for completion of the life cycle. The transition to a flagellate form appears useful for rapid relocation of the cell, before reverting back into the predatory amoebal form (Panek et al., 2012; Robinson et al., 1989). Several heteroloboseans exist only as flagellates, such as *Percolomonas* (Fenchel and Patterson, 1986), whereas others, including *Sawyeria* and *Vahlkampfia*, have no known flagellate stage (O'Kelly et al., 2003). Also found within the Heterolobosea are a number of sorocarpic amoebae, such as *Acrasis*, that aggregate to form multicellular fruiting bodies as part of their complex lifecycle. Using a combination of ultrastructural and phylogenetic analyses, these amoebae have been shown to be *bona fide* members of the Excavata, separate from other sorocarpic amoebae such as *Dictyostelium*, which belongs to the Amoebozoa (Brown et al., 2012). Heteroloboseans are known to inhabit a diverse range of environments, for example, extreme hypersaline, acidic and hot conditions (Baumgartner et al., 2009; De Jonckheere, 2004; Park and Simpson, 2015; Park et al., 2009). Some heteroloboseans live in anaerobic or microoxic conditions, and these heteroloboseans also lack classical mitochondria, and instead contain hydrogenosomes. Of these organisms, *Sawyeria* can be cultured in the presence of trace amount of oxygen,

whereas *Psalteriomonas* and *Monopylocystis* are strictly anaerobic (Barbera et al., 2010; O'Kelly et al., 2003). Surprisingly, the published genome of *Naegleria gruberi* (Fritz-Laylin et al., 2010b), which had always been considered to be an obligate aerobe, provided the first evidence that *Naegleria* could adapt to survive in low-oxygen conditions encountered in soils and muddy sediments (Fritz-Laylin et al., 2010b). *Naegleria* appears to contain genes that encode for an Fe-Fe hydrogenase and all of the associated maturases for anaerobic hydrogen production, and various other enzymes classically associated with anaerobic metabolism (for further detail, see Section 1.4) (Fritz-Laylin et al., 2010b; Ginger et al., 2010). Since the metabolic flexibility in *Naegleria* is so diverse, other heteroloboseans that are typically thought of as being aerobic may also be capable of switching anaerobic metabolic pathways when encountering hypoxic conditions (Panek et al., 2012).

The second clade that makes up the Discicristata, the Euglenozoa, is comprised of the Euglenida, Diplonemea, Symbiontida and Kinetoplastea. Kinetoplastids are the best-studied group within the Euglenozoa (Marande et al., 2005). One of the defining morphological characteristics is the kinetoplast, a network of concatenated mitochondrial DNA structures made up of maxicircles and minicircles (for further detail, see Section 1.4) (reviewed by Lukes et al., (2005); Lukeš et al., (2002)). Kinetoplastids also possess at least one flagellum, with a flagellar pocket at the base that is connected to the kinetoplast via a basal body (Field and Carrington, 2009). The kinetoplastids are subdivided into the Prokinetoplastina and Metakinetoplastina (Adl et al., 2012). The Prokinetoplastina consists of two genera: *Ichthyobodo*, which is an ectoparasite of wild and commercially farmed fish (Callahan et al., 2002) and Perkinsiella-like organisms, that are amoebal endosymbionts (Moreira et al., 2004). The Metakinetoplastina is further divided into the Neobodonida, Parabodonida, Eubodonida and the exclusively parasitic Trypanosomatida (Adl et al., 2012; Moreira et al., 2004; Simpson et al., 2006a). The clades that make up the kinetoplastids are extremely ecologically diverse, ranging from ubiquitous, free-living heterotrophs to obligate parasites of vertebrates, invertebrates and plants (Simpson et al., 2006b). The most well-known and studied kinetoplastids belong to the Trypanosomatid family, which contains medically important species including *Trypanosoma brucei*, the causative agent of African sleeping sickness in humans and the wasting disease nagana in mammals such as cows, horses and camels; *Trypanosoma cruzi*, which is prevalent in South America and is responsible for Chagas disease; and *Leishmania* sp. which causes leishmaniasis, an umbrella term describing a range of cutaneous, muco-cutaneous and visceral pathologies in humans. The closest recognised free-living relative to the trypanosomatids is, so far, the only characterised

member of the Eubodonida sub-group – *Bodo saltans*. It is a heterotrophic bacteriovore that is widely distributed across freshwater and marine environments (Jackson et al., 2008). The recent sequencing and annotation of the *Bodo saltans* genome (Jackson et al., 2015) revealed how the trypanosomatids evolved to become obligate parasites. The parasitic trypanosomatids have stream-lined genomes, with substantial gene losses relating to macromolecular digestion and ion transport; however, in contrast several gene families relating to nutrient scavenging and host invasion have expanded within the parasites, illustrating how unexpected innovations as well as stream-lining can contribute to genome composition during adaptation to parasitism (Jackson et al., 2015).

The Euglenozoa is made up of free-living, aquatic flagellates that are morphologically defined by a complex pellicle, which consists of proteinaceous strips located under the plasma membrane that run from the anterior to the posterior of the cell surface, resembling a superficial cytoskeleton (Leander, 2004; Leander et al., 2007). Euglenids occupy a diverse range of trophic niches because of the wide range of modes of nutrition available to them. Over 1000 species have been described so far, and many of them are phagotrophs that inhabit marine and freshwater sediments. Some, such as *Petalomonas* and *Ploeotia*, consume small prey such as bacteria; others such as *Peranema* and *Dinema* prey upon larger eukaryotic cells including diatoms and green algae. Other euglenids are photoautotrophic or osmotrophic. Photoautotrophic euglenids, such as *Euglena gracilis*, are chimeric cells that obtained their chloroplasts by a secondary endosymbiosis event with green algal cells, which may have once been a common prey item (Leander, 2004).

Finally, the Diplonemea and Symbiontida represent two, possibly much smaller groups within the Discicristata, although the idea that diplomemids are wide-spread in their distribution and very diverse in number has been discussed recently (Lukes et al., 2015). Until recently, only two known genera, *Rhynchopus* and *Diplonema*, made up the Diplonemea. Many are free-living phagotrophs although some parasitic species have been isolated from the blood and gills of marine life (von der Heyden et al., 2004). A recent phylogenetic analysis suggested that *Hemistasia phaeocyticola* may belong within the Diplonemea, despite having a kinetoplast (Yabuki and Tame, 2015). The Symbiontida is the most recently recognised clade within the Discicristata (Yubuki et al., 2009), although may be derived from phagotrophic euglenids even though they lack pellicle structures (Adl et al., 2012). Members, such as *Calkinsia* and *Postgaardi*, live in low-oxygen conditions and are covered in epibiotic bacteria (Yubuki et al., 2009; Yubuki et al., 2013). *Calkinsia* possesses modified mitochondria that structurally resemble hydrogenosomes (Yubuki et al., 2009).

Table 1.2

Table 1.2. A description of the further categorisation of the Excavata super-group. Based on Adl et al., (2012).

A. Shown are the super-phyla, phyla and clade that members of the Excavata are divided into, and an example species for each clade.

B. Light microscopy image of *Giardia duodenalis*. Reproduced from Hahn et al., (2013)

C. Light microscopy image of *Jakoba incarcerata*. Reproduced from Simpson and Patterson (2001)

D. Light microscopy image of *Sawyeria marylandensis*. Reproduced from Panek et al., (2011)

E. Light microscopy image of *Bodo saltans*. Reproduced from Lee et al., (2005)

A.

Super-group	Super-phylum	Phylum	Clade	Example	
Excavata		Preaxostyla	<i>Oxymonadida</i>	<i>Oxymonas</i>	
			<i>Trimastix</i>	<i>Trimastix</i>	
		Fornicata	Diplomonadida	<i>Giardia</i>	
				<i>Retortamonadida</i>	<i>Retortamonas</i>
			Carpediemonas-like	<i>Carpediemonas</i>	
				Trichomonadea	<i>Trichomonas</i>
					Cristamonadea
		Trichonymphea	<i>Trichonympha</i>		
			Spirotrichonymphea	<i>Holomastigotes</i>	
			Jakobida	<i>Jakoba</i>	<i>Jakoba</i>
		Histionidae		<i>Reclinomonas</i>	
				<i>Andalucia</i>	<i>Andalucia</i>
		Discobata		Discicristata	Heterolobosea
			Euglenozoa		<i>Trypanosoma</i>
		<i>Malawimonas</i>	<i>Malawimonas</i>		

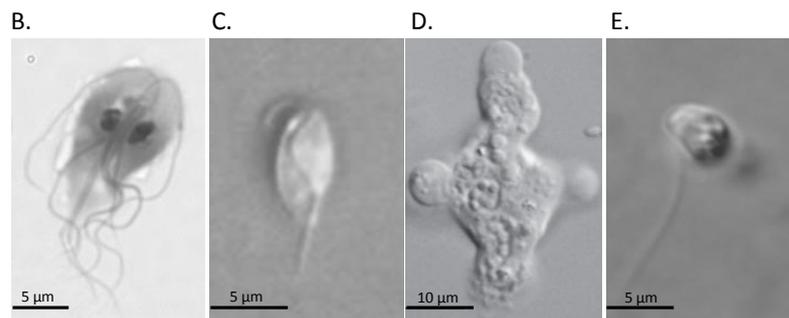


Table 1.3

Table 1.3. 'Typical' common cytoskeletal features shared amongst many excavate protists. All excavates share at least one or more of these features, apart from the Euglenozoa, which completely lack any of the characteristics described above. However, phylogenetic analyses place the Euglenozoa firmly within the Excavata.

Distinctive Excavate Feature	Description
Feeding groove	Distinctive groove used for suspension-feeding
I fibre	Non-microtubular fibre associated with the microtubular root R1
B fibre	Non-microtubular fibre associated with the microtubular root R1
C fibre	Non-microtubular fibre associated with the microtubular root R2
Split R1	Microtubular root R1 splits shortly after origin into inner and outer sections
Singlet root	Single microtubule that originates between R1 and the basal body, then runs along the floor of the groove
Flagellar vanes	Prominent vanes that run along the length of flagellum 1
Composite fibre	Dense non-microtubule fibres that originate in association with outer section of split R1 and run down the groove

1.4 Mitochondrial diversity

Classic textbook metabolism in aerobic chemotrophic eukaryotic cells is based upon respiration in mitochondria, utilising carbohydrates, fatty acids and amino acids as carbon sources for the generation of ATP by oxidative phosphorylation, and using oxygen as the terminal electron acceptor for respiration. A classic mitochondrion is comprised of four essential structures – the outer membrane, inner membrane, intermembrane space and the matrix. Within the matrix, the inner membrane is folded into invaginations known as cristae, which greatly increases the surface area of the inner membrane. Within mitochondria essential enzymes of energy metabolism are pyruvate dehydrogenase, the components of the Krebs cycle, electron transport chain and ATP synthase. A variety of other essential pathways and processes also occur in textbook mitochondria (Mootha et al., 2003).

Mitochondria are semi-autonomous organelles and contain their own genome or mitochondrial DNA (mtDNA), which is also located within the matrix. Mitochondrial genomes vary greatly in size, complexity and structure. In general, Metazoan mtDNA maps as circular molecules, similar in structure to bacterial circular genomes, and typically encode 22 tRNA molecules, 2 rRNA molecules and 13 polypeptides (Taanman, 1999). All 13 polypeptides are involved in the formation of proteins that are required for the electron transport chain, and so are critical for mitochondrial function. Metazoan mtDNA generally contains very little non-coding sequence and also generally contains no introns (Burger et al., 2003). An exception are Cnidaria (Beagley et al., 1998), which are aquatic animals such as sea anemones and corals. For instance, the sea anemone *Metridium senile* has only two tRNA molecules that are coded for by mtDNA, yet it was the first metazoan organism to be found that contains self-splicing group I introns (ribosomes that catalyse their own splicing from precursor mRNA) (Beagley et al., 1998).

Most of the diversity in the shape, size and complexity of mitochondrial genomes can be found amongst the plants and protists. This is unsurprising in that protists account for the majority of eukaryotic biodiversity and even though the exact root of the eukaryotic evolutionary tree still remains elusive, the last common ancestor of all eukaryotic life was almost certainly a primitive unicellular protist with a protomitochondrion, acquired through a crucial symbiotic event involving a α -proteobacterium (Embley and Martin, 2006; Gray, 2012). Because animal, plant and fungal lineages do not extend as far back in evolutionary history as many protists, it is necessary to collect a wide spread of both parasitic but more importantly free-living protist mitochondrial genomes, in order to compare animal mtDNA

with ancestral protist mtDNA. By doing these comparisons, it is possible to look at (complex) patterns of gene loss and gene transfer to the nucleus during eukaryotic evolution (Gray et al., 1998; Martin, 2003; Timmis et al., 2004). It should even be possible to infer with high confidence the architecture and composition of mtDNA in LECA.

Often mtDNA is circular in structure, like the metazoan mtDNA described above, but a variety of unicellular protists including the ciliates *Tetrahymena pyriformis* (Suyama and Miura, 1968) and *Paramecium aurelia* (Goddard and Cummings, 1975), the chlorophycean green algae *Chlamydomonas reinhardtii* (Vahrenholz et al., 1993), and apicomplexans such as *Theilera* sp. (Kairo et al., 1994) and *Babesia* sp. (Brayton et al., 2007; Hikosaka et al., 2010) possess linear double-stranded mtDNA. Linear mtDNA is not exclusive to protist lineages and has also been found in the metazoan phylum Cnidaria (Beagley et al., 1998), yeasts (Fukuhara et al., 1993) and some fungi (Martin, 1995). However, one of the most unusual mitochondrial genome structures can be found in the kinetoplastids. Within the mitochondrial matrix of the kinetoplastids two different types of interlocked (or catenated) circular structures known as maxicircles and minicircles are found, which forms the kinetoplastid DNA (kDNA) (Lukeš et al., 2002; Morris et al., 2001). Maxicircles range in size from ~20 kb to ~40 kb depending upon species, and code for rRNA subunits and electron transport components and some proteins of still unknown function (Lukeš et al., 2002). A few dozen copies of the maxicircle are present per cell. In contrast, minicircles are usually ~0.5 kb to ~2.5kb in size with several thousand minicircles per cell. Minicircles encode guide RNAs that are needed for editing cryptic pre-mRNAs into translatable mRNAs (Lukeš et al., 2002).

Size and gene content of mitochondrial genomes can vary greatly between eukaryotes. The parasitic alveolate *Plasmodium falciparum* has a small mitochondrial genome that is only 5,967 bp (Gardner et al., 2002) in size, coding for only 3 proteins (Feagin, 1992; Feagin et al., 1991); whereas the mitochondrial genome of the free-living jakobid *Reclinomonas americana* is 69,034 bp, and codes for 97 proteins which is one of the largest gene sets identified so far in mtDNA (Lang et al., 1997). More recently, other jakobids have been identified to have even larger mitochondrial genomes, encoding for even more proteins (Burger et al., 2013). The mitochondrial genome of *Andalucia godoyi* codes for exactly 100 proteins (Burger et al., 2013). The mitochondrial genomes of the jakobid protists have many bacterial-like features. Many genes are retained within the jakobid mitochondrial genome that in other organisms have been transferred to the nucleus, for instance the mtDNA of

Reclinomonas americana still codes for many of the electron transport components and contains Shine-Delgarno motifs, which are classically associated with prokaryotes as the binding site for prokaryotic mRNA (Burger et al., 2013; Lang et al., 1997). Because of these unusual features, the mitochondrial genome of *Reclinomonas americana* and other jakobids have been the most informative in terms of revealing what the protomitochondrial genome of the last common ancestor may have contained (Burger et al., 2013; Gray et al., 1998; Lang et al., 1997). In comparison with the jakobids, the mitochondrial genome of the free-living heterolobosean *Naegleria gruberi* is smaller at 49,843 bp, and encodes for only 43 proteins (Rudinger et al., 2011). This indicates that the jakobids diverged earlier in evolutionary history than the heteroloboseans, and their bacterial-like mitochondria have been retained. Putting size in the context of nucleotides per haploid genome, the flowering land plant *Arabidopsis thaliana* has a mitochondrial genome that is 366,924 bp in size, but it only codes for 57 proteins (Unsold et al., 1997). Approximately 80% of the *Arabidopsis* mtDNA is non-coding sequence.

Whilst many eukaryotes are obligate aerobes, with ATP production revolving around the classic central metabolic processes that are compartmentalised in the mitochondria, there are also eukaryotes that were thought to lack mitochondria and were thus originally termed 'amitochondriate'. These eukaryotes grow within anoxic and/or microoxic environments, but careful molecular, ultrastructural and fluorescence-based microscopy approaches have shown that these eukaryotes contain degenerate forms of mitochondria. Perhaps the most well-known of these degenerate mitochondria are the hydrogenosomes. Hydrogenosomes were discovered in the 1970s (Lindmark and Müller, 1973), but their energy-generating capacity was initially poorly understood, and consequently their relationship to mitochondria took time to be established. The most reduced form of mitochondria are the mitosomes; discovered at the very end of the 20th century (Tovar et al., 1999) mitosomes are both highly reduced in terms of metabolic complexity and very small in size. Thus, a broad spectrum of mitochondrial diversity is apparent and collectively degenerate mitochondria are known as 'mitochondrion-related organelles' or MROs. Currently, there is no example known of a eukaryote that lacks mitochondria entirely.

Due to the huge variability in mitochondrial form and function it is difficult to categorise them even though these organelles share a single origin. Muller et al., (2012) suggested a new system for categorising mitochondria and related organelles into one of five classes: Class I - the classical aerobic mitochondria that generate ATP, have an electron transport chain and use oxygen as their terminal electron acceptor (described above). Class II -

anaerobic mitochondria that generate ATP, have an electron transport chain and can utilise an environmental electron acceptor such as nitrite or an endogenous electron acceptor such as fumarate as the terminal electron acceptor. Class III - hydrogen-producing mitochondria which generate ATP, can use protons as the terminal electron acceptor and still have an electron transport chain, albeit simplified in comparison with aerobic mitochondria. Class IV – hydrogenosomes which generate ATP and use protons as the terminal electron acceptor to generate hydrogen without the presence of an electron transport chain. Class V – mitosomes which do not generate ATP, lack electron transport chains and terminal electron acceptors.

Anaerobic mitochondria are very much like classical mitochondria as in they have a proton-pumping electron transport chain, but oxygen is not the terminal electron acceptor and an alternative electron acceptor is required. For instance, certain types of fungi, such as *Fusarium oxysporum*, rely on nitrate respiration under low-oxygen conditions, and they utilise environmentally-derived NO_3^- or NO_2^- to act as the terminal electron acceptor. These denitrifying mitochondria contain enzymes not found in aerobic mitochondria (Morozkina and Kurakov, 2007) such as nitrite reductase (NirK) for the reduction of NO_2^- to NO (Kobayashi et al., 1996; Kobayashi and Shoun, 1995) or nitrate reductase (NaR) for the reduction of NO_3^- to NO_2^- (Kobayashi et al., 1996; Uchimura et al., 2002). NO can be further reduced to N_2O by the nitric reductase enzyme (NoR) (Nakahara and Shoun, 1996). Endogenous terminal electron acceptors such as fumarate can also provide an alternative electron sink to oxygen. Here, glucose is converted to oxaloacetate then reduced further to form malate which is then either oxidised to form acetate or reduced further to form succinate and ultimately propionate, this dismutation occurs to retain a redox balance (Muller et al., 2012). Electron transport chains generally require a quinone to function as part of the electron transport shuttle. Under aerobic conditions, this function is performed by ubiquinone, but the redox potential of ubiquinone is too high to allow efficient electron transfer to fumarate (Van Hellemond et al., 1995). Alternative quinones with lower redox potentials that are capable of donating electrons to fumarate have been found in eukaryotes, such as rholoquinone in parasitic helminths (Van Hellemond et al., 1995) and the excavate *Euglena gracilis* (Tielens et al., 2002).

Perhaps the best characterised organism to contain class III hydrogen-producing mitochondria is *Nyctotherus ovalis*, a ciliate that colonises the hindgut of cockroaches. These mitochondria still retain a mitochondrial genome which encodes for an electron transport chain and some aerobic respiratory enzymes such as pyruvate dehydrogenase are still coded for in the nuclear genome; however they have a functioning hydrogenase, the archetypal

enzyme found in hydrogenosomes, which produces hydrogen. The characterisation of hydrogen production and mitochondrial genome architecture in *N. ovalis* provided a previously missing link that confirmed the mitochondrial ancestry of hydrogenosomes (Boxma et al., 2005b).

The anaerobic excavate *Tritrichomonas foetus* is an obligate bovine parasite that lacks classical mitochondria and like its better known relation *Trichomonas vaginalis*, favours alternative anaerobic metabolic pathways for energy generation (Carlton et al., 2007). *Tritrichomonas foetus* was the first eukaryote described as producing molecular hydrogen via a highly active hydrogenase enzyme localised in 'microbody-like' organelles. These organelles were regarded to be the site of anaerobic catabolism where protons functioned as the terminal electron acceptors and thus named the 'hydrogenosome' (Lindmark and Müller, 1973). Since this seminal discovery, other unicellular eukaryotes, both parasitic and free living, from a wide variety of eukaryotic groups including fungi, excavates and alveolates, have been shown to possess hydrogenosomes (Barbera et al., 2010; Boxma et al., 2007; Carlton et al., 2007; Embley et al., 2003; Makiuchi and Nozaki, 2014; Muller et al., 2012; Yubuki et al., 2009).

Hydrogenosomes are oxygen-sensitive organelles that generate H₂ following the catabolism of typically glucose or malate. Pyruvate is derived from glucose by glycolysis or by the decarboxylation of malate and then oxidised using the enzyme pyruvate:ferredoxin oxidoreductase, rather than pyruvate dehydrogenase, to form acetyl CoA. Subsequently, rather being metabolised through the Krebs cycle, as in classic aerobic mitochondrial metabolism, the acetyl CoA formed is converted to acetate, H₂ and CO₂, thereby generating ATP exclusively by substrate-level phosphorylation. Protons are used as the terminal acceptor for electrons removed during substrate oxidation, in order to produce molecular hydrogen. This is a radically different biochemistry to that seen in classic mitochondria which typically use oxygen as the terminal electron acceptor and drives the generation of ATP by oxidative phosphorylation linked to the electron transport chain. By using protons as the terminal electron acceptor in the hydrogenosome instead of oxygen, an overall requirement for oxidative phosphorylation using either oxygen or other terminal acceptors such as NO₃⁻, NO₂⁻ or fumerate, is eliminated (Muller, 1993).

[FeFe]-hydrogenase is a metalloenzyme that is oxygen-labile and catalyses the reversible reaction $2\text{H}^+ + 2\text{e}^- \rightleftharpoons \text{H}_2$. It is the key characteristic enzyme that defines a class IV hydrogenosome. [FeFe]-hydrogenases are the fastest known catalysts for the formation of

molecular hydrogen, operating at turnover frequencies greater than 100 s^{-1} (Vincent et al., 2007). Electron carriers such as NADH are also recycled during hydrogen generation, therefore serving an important role in the energy generation process during anaerobic fermentation (enzyme 4 in Figure 1.7 A).

In hydrogenosomes, pyruvate:ferredoxin oxidoreductase (PFO), is another organelle-characteristic, oxygen-sensitive enzyme that like pyruvate dehydrogenase (PDH) catalyses the oxidation of pyruvate to acetyl CoA (enzyme 1 in Figure 1.7 A). Whilst PFO is conserved amongst most anaerobic ciliates (Hackstein et al., 1999; Muller, 1993), no genes encoding PFO were found in *Nyctotherus ovalis* which instead contains a PDH (Boxma et al., 2005b); this provides further evidence that hydrogenosomes have mitochondrial ancestry. Both PFO and PDH are absent from the hydrogenosome-containing chytrid fungi *Neocallimastix* sp. and *Pyromyces* sp., and these organisms instead cleave pyruvate using pyruvate-formate lyase enzyme (Akhmanova et al., 1999).

There are three known types of metal hydrogenase enzymes found in nature with their characterisation based upon their metal dependency: [NiFe]-hydrogenases, [FeFe]-hydrogenases and [Fe]-hydrogenases. Unlike the [FeFe]-hydrogenases and [NiFe]-hydrogenases that both catalyse the $2\text{H}^+ + 2\text{e}^- \rightleftharpoons \text{H}_2$ reaction, [Fe]-hydrogenases function as the so called H_2 -forming methylenetetrahydromethanopterin dehydrogenase (Hmd) to catalyse an intermediary step during the formation of methane from CO_2 . [Fe]-hydrogenases differ from the other two types of hydrogenases due to a lack of any Fe-S clusters and metal atoms in the active site (Vignais and Billoud, 2007).

Many of the eukaryotic hydrogenases are [FeFe]-hydrogenases, which typically localise within hydrogenosomes. However genomic and biochemical analyses of some eukaryotes such as the amoebozoan parasite *Entamoeba histolytica* and the excavate *Giardia* have revealed active [FeFe]-hydrogenases, despite the presence in these organisms of mitosomes rather than hydrogenosomes (Horner et al., 2000; Nixon et al., 2003). In *E. histolytica* and *Giardia*, [FeFe]-hydrogenase is a cytosolic enzyme. It was thought that the excavate fish parasite *Spironucleus* also lacked hydrogenosomes but still had functioning [FeFe]-hydrogenase enzymes (Horner et al., 2000; Nixon et al., 2003). However recently, organelles with all the characteristics of hydrogenosomes were characterised in *Spironucleus salmonicida* (Jerlstrom-Hultqvist et al., 2013).

Thus far, the active sites of all characterised [FeFe]-hydrogenases are found within a highly conserved metal binding domain known as the H-cluster which consists of a [4Fe-4S] cluster

connected to the [2Fe-2S] core via a single cysteinate residue, also attached to the [2Fe-2S] coordination centre are carbon monoxide, cyanide and dithiolate ligands (see Figure 1.7 B) (Camara and Rauchfuss, 2012; Mulder et al., 2011).

All eukaryotes expressing an organellar [FeFe]-hydrogenase also require three maturation proteins, HydE, HydF and HydG, in order for the active holo-hydrogenase to bind its H-cluster and thus exhibit catalytic activity. It is thought HydE and HydG are directly involved in the biochemical synthesis of the H-cluster, allowing the synthesis of the associated ligands, whereas HydF is thought to serve a structural purpose and act as a specialised scaffold protein during the assembly of the mature hydrogenase holoenzyme (Mulder et al., 2011).

Up until relatively recently, it was thought to be impossible for multicellular organisms to survive and reproduce in extreme anoxic conditions such as those found in hypersaline ocean basins. Due to the complete absence of oxygen and light, such basins were thought to be exclusively inhabited by chemotrophic prokaryotes and some unicellular eukaryotes. Yet Danovaro et al., (2010) reported the first metabolically active metazoan that completed its entire life cycle in the hypersaline sulphidic environment found in a deep Mediterranean ocean basin. These metazoans belong to the phyla Loricifera and thrive despite the complete lack of oxygen caused by the near-salt saturated brine that is denser than the oxic water above, and the toxic conditions caused by the sulphide produced by sulphate-reducing prokaryotes that also inhabit this environment (van der Wielen et al., 2005). The Loricifera are the first animals to be found that cope without the 'absolute requirement' for oxygen at any point in their life cycle because of a high abundance of hydrogenosomes in the cells, this illustrates how, at least in part, these animals have evolved and adapted to life in an extreme environment (Danovaro et al., 2010).

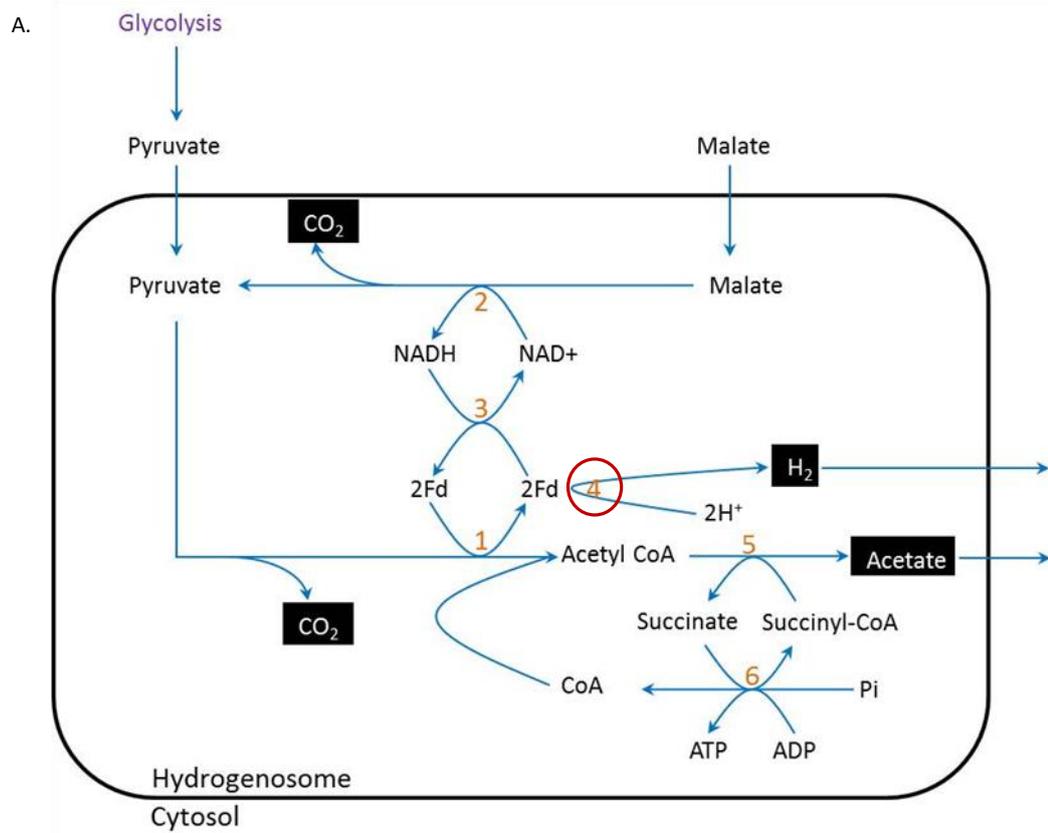
Mitosomes remain highly reduced forms of mitochondria of ill-defined function. These organelles are drastically reduced in both size and biochemical complexity from the original proto-mitochondrion. They were first discovered in *E. histolytica* less than 20 years ago (Tovar et al., 1999). Mitosomes are not a site of ATP generation (Katinka et al., 2001) but the repertoire of enzymes and pathways found in mitosomes is still poorly understood. Mitosomes are bound by a double membrane but they do not contain a genome. Thus proteins active within mitosomes must be synthesised in and imported from the cytosol (Embley and Martin, 2006).

With respect to protein import into mitosomes, some mitochondrial proteins in microsporidian parasites and *Giardia* contain recognisable N-terminal or protein-internal motifs that direct

protein import into mitosomes. This mode of protein targeting also occurs in mitochondrial protein import, but both the complexity of the mitosomal protein import apparatus and the length of N-terminal targeting motifs are reduced in comparison with the mitochondrial protein import apparatus and mitochondrial protein-sorting motifs (Burri et al., 2006; Dolezal et al., 2005; Jedelsky et al., 2011; Regoes et al., 2005). Thus, the cryptic nature of the targeting or sorting motifs present in perhaps many mitosomal proteins adds to the difficulty in understanding precisely the role(s) that mitosomes play in different eukaryotes.

The only biochemical pathway confirmed in mitosomes thus far is used for Fe-S cluster biosynthesis (Jedelsky et al., 2011; Martincova et al., 2015; Tovar et al., 2003). It appears that mitochondria, hydrogenosomes and mitosomes all possess Fe-S assembly pathways, and Fe-S clusters are critical co-factors for a variety of essential proteins found inside and outside of mitochondria and MROs (Lill, 2009). Like the protein targeting and translocation system, the Fe-S cluster pathway also appears to be reduced in mitosomes compared to that found in the mitochondria. For instance, frataxin, a protein associated with Fe-S cluster assembly, is absent in the *Giardia* mitosome, although it has been found in the mitosome of the Microsporidian *Encephalitozoon cuniculi* (Goldberg et al., 2008). Because energy metabolism in *Entamoeba* and *Giardia* parasites is restricted to the cytosol, it is not clear how, or indeed if at all, the mitosomes of these parasites contribute to other cellular processes beyond Fe-S cluster assembly.

Figure 1.7



B.

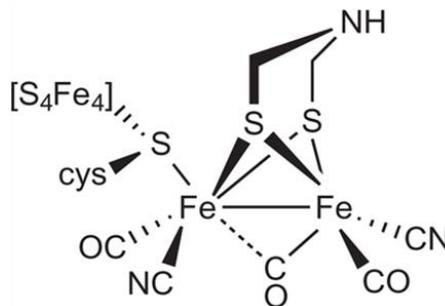


Figure 1.7. Pyruvate metabolism in hydrogenosomes

A. Metabolic pathways seen in hydrogenosomes. End products are shown in white on a black background. 1, Pyruvate:ferredoxin oxidoreductase; 2, malate dehydrogenase; 3, NAD:ferredoxin oxidoreductase; 4, FeFe-hydrogenase (highlighted in a red circle); 5, acetate:succinate CoA-transferase; 6, succinate thiokinase. Modified image from Muller, (1993).

B. Structural model of the active site common to all [FeFe]-hydrogenases. The H-cluster is linked to the FeS-cluster domains with their associated CO, CN⁻ and dithiolate ligands. The H-cluster contains the active site whilst the Fe-S clusters are involved in electron transfer to/from the active site. Modified from Camara and Rauchfuss, (2012).

1.5 Life cycle biology and metabolism of *Trypanosoma brucei*

Trypanosomatids from the genus *Trypanosoma* that are capable of causing human pathology include members of the African trypanosome family and the South American parasite *Trypanosoma cruzi* (Simpson et al., 2006b). African trypanosomes can be the cause of both human and veterinary disease. The infection of susceptible cattle by a variety of African trypanosome species places a very significant economic burden on many sub-Saharan communities and countries (Reid et al., 2012; Sutherland et al., 2015). With respect to human African sleeping sickness, or human African trypanosomiasis, the two *Trypanosoma brucei* sub-species, *T. b. gambiense* and *T. b. rhodesiense*, cause disease. Human African trypanosomiasis (HAT) is endemic in over 25 sub-Saharan African countries putting an estimated 70 million people at different levels of risk for contraction of disease (Simarro et al., 2015; Simarro et al., 2012). 98% of declared sleeping sickness cases are caused by *T. b. gambiense* (Franco et al., 2014) which is endemic in many west and central African countries, particularly the Democratic Republic of Congo (Figure 1.8). The remaining cases of HAT are caused by *T. b. rhodesiense*, which is found in east and southern Africa, but is less adapted to human infection and is primarily a zoonotic disease. Typically, the two species are separated by the Rift Valley, with *T.b. gambiense* being prevalent to the west of the valley, and *T. b. rhodesiense* to the east (Figure 1.8) (Franco et al., 2014). Left untreated, HAT is invariably fatal once clinical symptoms develop. A final *T. brucei* sub-species, *T. b. brucei*, is one of the trypanosome species responsible for African animal trypanosomiasis. It is also the sub-species most amenable to laboratory culture and genetic manipulation and is the sub-species used for the research described in this thesis. The three sub-species of *T. brucei* are morphologically indistinguishable, undergo the same developmental cycles in mammals and the tsetse fly vector, and exhibit the same stage-specific differences in metabolism.

Trypanosomatids have evolved a complex heteroxenous life cycle that requires replication and differentiation in both an insect vector and the vertebrate host. When the tsetse fly takes a blood meal from an infected vertebrate host, trypomastigotes are ingested. In the lumen of the midgut, any long slender forms either transform to short stumpy forms in the anterior of the midgut or die. At the posterior of the midgut, in the endoperitrophic space, short stumpy forms in turn differentiate into procyclics, which are well-adapted for survival in the insect vector (Vickerman, 1985) (Figure 1.9). The length of the cells increases in the procyclic form, the parasites can replicate and the mitochondrion expands throughout the cell, forming an elaborately branched network, which reflects the marked changes in metabolic pathways utilised by the cells (Rotureau et al., 2012; Vickerman, 1985). The

procyclic forms then penetrate the peritrophic membrane to enter the ectoperitrophic space, where they continue to divide and replicate. From here, the parasites migrate to the salivary glands of the tsetse fly, where they progress through a series of stages. The first stage is an adherent epimastigote stage, in which the trypanosome attaches to the epithelial cells of the gland lumen via the flagellum. Epimastigotes are highly proliferative and quickly establish a colony in the salivary glands. Epimastigotes then differentiate into 'premetacyclic' trypomastigotes which are still attached to epithelial cells, can still divide and still have the branched mitochondrion. This stage is characterised by the presence of two flagella. Upon division, one of the cells can give rise to a 'nascent pre-metacyclic' form, which has a variable surface antigen (VSG) coat and can rapidly assume the metacyclic form ready for infection of the mammalian host (Rotureau et al., 2012; Vickerman, 1985). The metacyclic form is the vertebrate-infective, free-swimming stage that cannot proliferate, has a repressed, tubular mitochondrion and possesses a VSG coat.

Tsetse flies infected with trypanosomes display altered behavioural patterns. Flies are forced into taking longer and more frequent blood meals, which increases the likelihood of the parasites infecting the mammalian host (Van Den Abbeele et al., 2010). Once a host is bitten by the tsetse fly, most of the infective metacyclics are ejected from the salivary glands and injected into the host dermis and bloodstream. They are extracellular parasites that must rapidly differentiate into long slender (LS) forms in order to survive and proliferate by binary fission in the bloodstream and lymphatic system of the host. In later stages, LS forms can cross into choroid plexus from where they have access to the cerebrospinal fluid and the rest of the brain (Vickerman, 1985). Once a critical cell density has been reached, LS forms begin to differentiate into short stumpy (SS) forms (Matthews, 1999) which are incapable of proliferation but are viable for differentiation into the procyclic form if ingested by the tsetse fly again. Parasitemia levels fluctuate throughout the course of infection in the mammalian host, dependent upon the host immune system responses and the parasite evasion by expression of various VSG coats (MacGregor et al., 2012).

Bloodstream trypanosomes have had to devise ingenious methods in order to evade the host immune system and the most well-known evasion mechanism is antigenic variation. The surface of *T. brucei* is densely covered by a coat of a single variant surface glycoprotein (Vickerman, 1969) that can be replaced by expressing a different VSG gene, once the immune system raises antibodies against the first (McCulloch, 2004).

A widely held belief is that parasites stream-line their biology as a consequence of adaptation to obligate parasitism. This is most often commented upon with respect to metabolism. *T. brucei* can be characterised by its unusual metabolic compartmentalisation and life cycle remodelling of glycolysis and mitochondrial metabolism. In *T. brucei* and other kinetoplastids, the majority of glycolytic enzymes are exclusively found within peroxisomes (Haanstra et al., 2015; Opperdoes and Borst, 1977). As a consequence of glycolytic enzymes being the most abundant enzymes and the most obvious pathway present in kinetoplastid peroxisomes, these peroxisomes are better known as glycosomes (Haanstra et al., 2015).

As *T. brucei* progresses through its complex life cycle, in addition to dramatic changes in morphological form, there are also significant changes in metabolism that reflect adaptation to different niches in host and vector. In the mammalian bloodstream (a glucose-rich environment) *T. brucei* is pleomorphic in form, with long slender, intermediate, and short stumpy trypomastigote forms present at different points during chronic parasitemia (Matthews et al., 2015). Long slender bloodstream forms utilise glucose as the sole source of carbon for ATP production and rely exclusively on glycolysis for this ATP production (Figure 1.10). The first seven enzymes of glycolysis, which catalyse the catabolism of glucose to 3-phosphoglycerate, occur exclusively in the glycosomes (Opperdoes, 1987; Opperdoes and Borst, 1977; van Hellemond et al., 2005; Vickerman, 1985). The remaining reactions of glycolysis occur in the cytosol, pyruvate is excreted as the waste end product, and net ATP production for the cell is catalysed by pyruvate kinase (PYK). The NADH generated during glycolysis is oxidised via a redox shuttle involving mitochondrial glycerol-3-phosphate dehydrogenase and the alternative oxidase, which uses O₂ as a terminal electron acceptor to produce H₂O (Michels et al., 2000).

Glycolysis in trypanosomes differs from other organisms not only in respect of glycosomal compartmentalisation, but also because there are no negative feedback mechanisms to regulate either hexokinase or phosphofructokinase activities. Glycosomal compartmentalisation either influenced or is influenced by the absence of these regulatory controls (Gualdrón-López et al., 2012; Michels et al., 2000; Tielens and van Hellemond, 2009). The glycosome is essential in trypanosomes as it protects against the accumulation of hexose phosphate intermediates that are produced during the initial steps of glycolysis (glucose 6-phosphate, fructose 6-phosphate and fructose 1,6-bisphosphate) (Haanstra et al., 2014).

The relative simplicity of energy metabolism in long slender bloodstream forms required to provide the energy requirements of the long slender form is also reflected in the structure of the mitochondrion, which is not directly utilized in energy generation. The long slender mitochondrion is greatly repressed and reduced in size and metabolic complexity compared to that in the procyclic form. Thus, in bloodstream *T. brucei*, the mitochondrion is an unbranched linear structure that runs from the anterior to the posterior of the cell and virtually lacks cristae (see Figure 1.9), there is also an absence of cytochromes and a functional Krebs cycle (Michels et al., 2000; Morgan et al., 2004; Opperdoes, 1987; Vickerman, 1985). As *T. brucei* differentiates from the long slender form into the short stumpy form via an intermediate stage, the mitochondrion begins to elongate and swell, forming tubular cristae as it primes its metabolism ready for uptake by a tsetse fly; an environment that is very different from the mammalian bloodstream and requires a more extensive metabolism that requires the processing of amino acids (Vickerman, 1985).

Once in the tsetse midgut, the short stumpy forms differentiate into the procyclic form, which see a dramatic change in the mitochondrion and biochemistry of *T. brucei* since it is an environment that is not abundant in glucose. The mitochondrion becomes more activated with respect to ATP production, resulting in expansion and elongation throughout the cell as it increases in length, forming an extensively branched network with discoid cristae (Vickerman, 1985) (see Figure 1.9). This change in mitochondrial form marks a switch to amino acid metabolism, which now becomes the principal energy source, rather than glucose which is limited in its availability within the insect vector.

As differentiation from the short stumpy form to the procyclic form begins, remodelling of the pathways in the glycosome also occurs. Another isoform of phosphoglycerate kinase (PGK) is expressed, which is found in the cytosol, and expression of glycosomal PGK declines. Expression of pyruvate phosphate dikinase (PPDK) is now seen in glycosomes. This enzyme converts phosphoenolpyruvate (PEP) into pyruvate, synthesising ATP from AMP and PP_i and thus competes at some level with pyruvate kinase for the PEP substrate. Most of the pyruvate generated from PEP is decarboxylated in the mitochondria to form acetyl-CoA by the enzyme pyruvate dehydrogenase (Bringaud et al., 2006). Alternately, PEP can be converted to oxaloacetate by the enzyme phosphoenolpyruvate carboxykinase. Oxaloacetate can be converted into malate for use in succinate production within the glycosome or the mitochondria. Other enzymes including malate dehydrogenase, fumarase and fumarate reductase are also found in the glycosome, and contribute to the formation of

succinate as a metabolic end product; this succinate production has the potential to contribute to glycosomal redox balance (see Figure 1.11).

In procyclic forms, acetyl CoA produced from decarboxylation of pyruvate or amino acid metabolism is not oxidised through the Krebs cycle, which in *T. brucei* is incomplete due to the absence of NAD-dependent isocitrate dehydrogenase. Instead, acetyl CoA is predominantly converted into acetate by the activity of acetate:succinate CoA-transferase (ASCT). ATP is generated when the succinyl-CoA co-product of the ASCT-catalysed reaction is converted back to succinate by succinyl-CoA synthase. Electrons transferred to mitochondrial NADH through a variety of enzyme-catalysed reactions, including reactions catalysed by PDH and the Krebs cycle enzymes which are present, enter the mitochondrial electron transport chain, for cytochrome dependent respiration and oxidative phosphorylation (Bringaud et al., 2006).

Figure 1.8

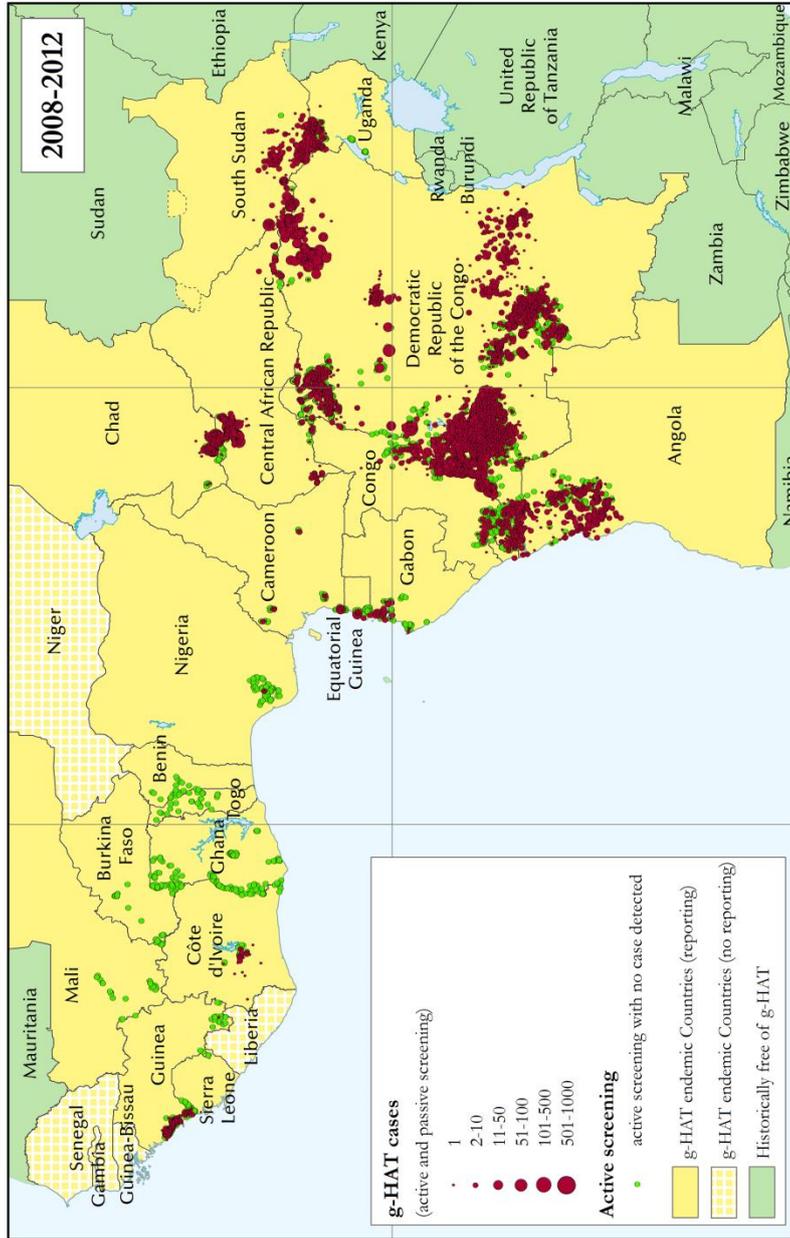


Figure 1.8. A map of sub-Saharan Africa highlighting the distribution of Human African trypanosomiasis caused by *T. b. gambiense* between the years 2008 to 2012. Modified from Simarro et al., (2015).

Figure 1.9

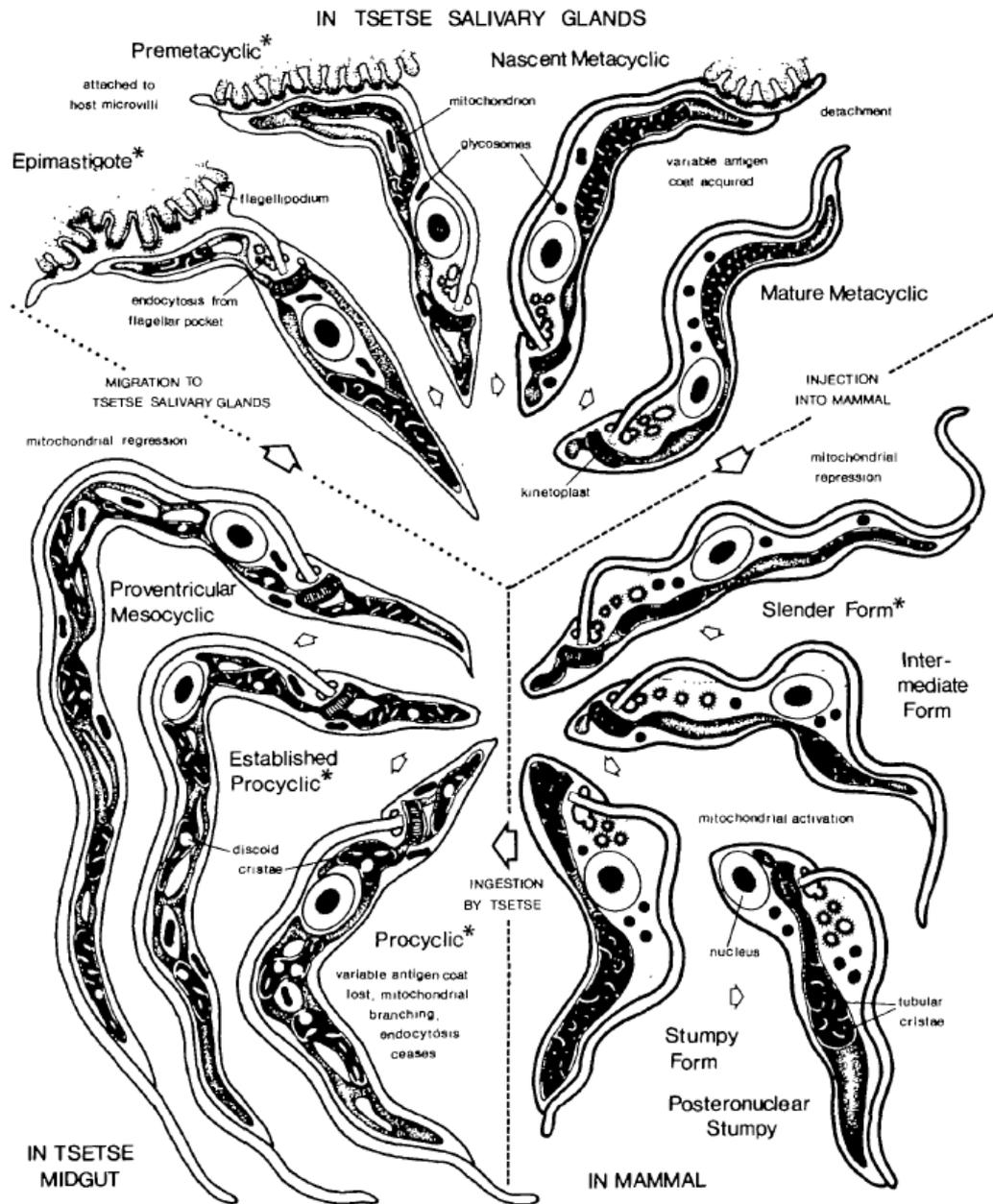


Figure 1.9. The biochemical changes *T. brucei* undergoes as it proceeds through its complex life cycle, including stages in both the mammalian and insect hosts. The cartoons are reproduced from Vickerman, (1985).

Figure 1.10

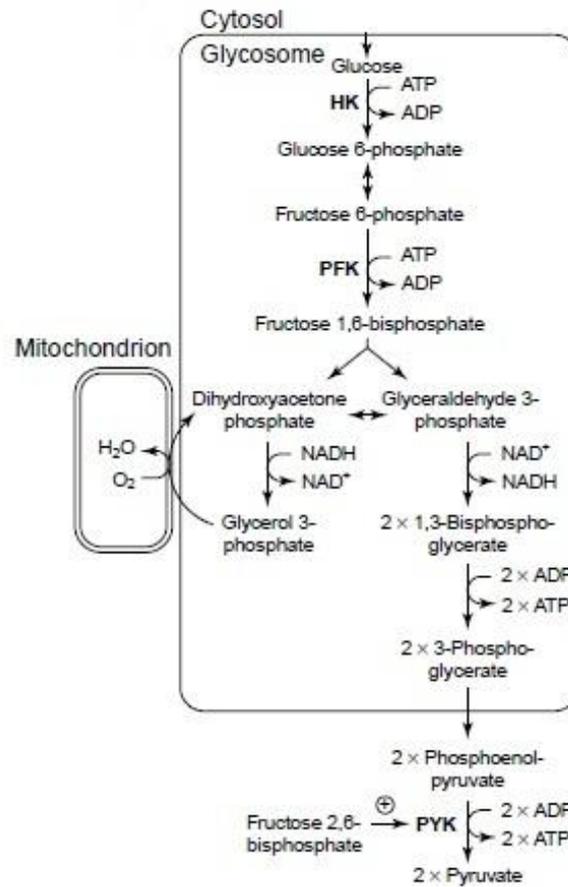


Figure 1.10. Metabolic pathways in the bloodstream form of *T. brucei* showing the conversion of glucose into pyruvate via the glycolytic pathway. Image adapted from the original Michels et al., (2000).

Figure 1.11

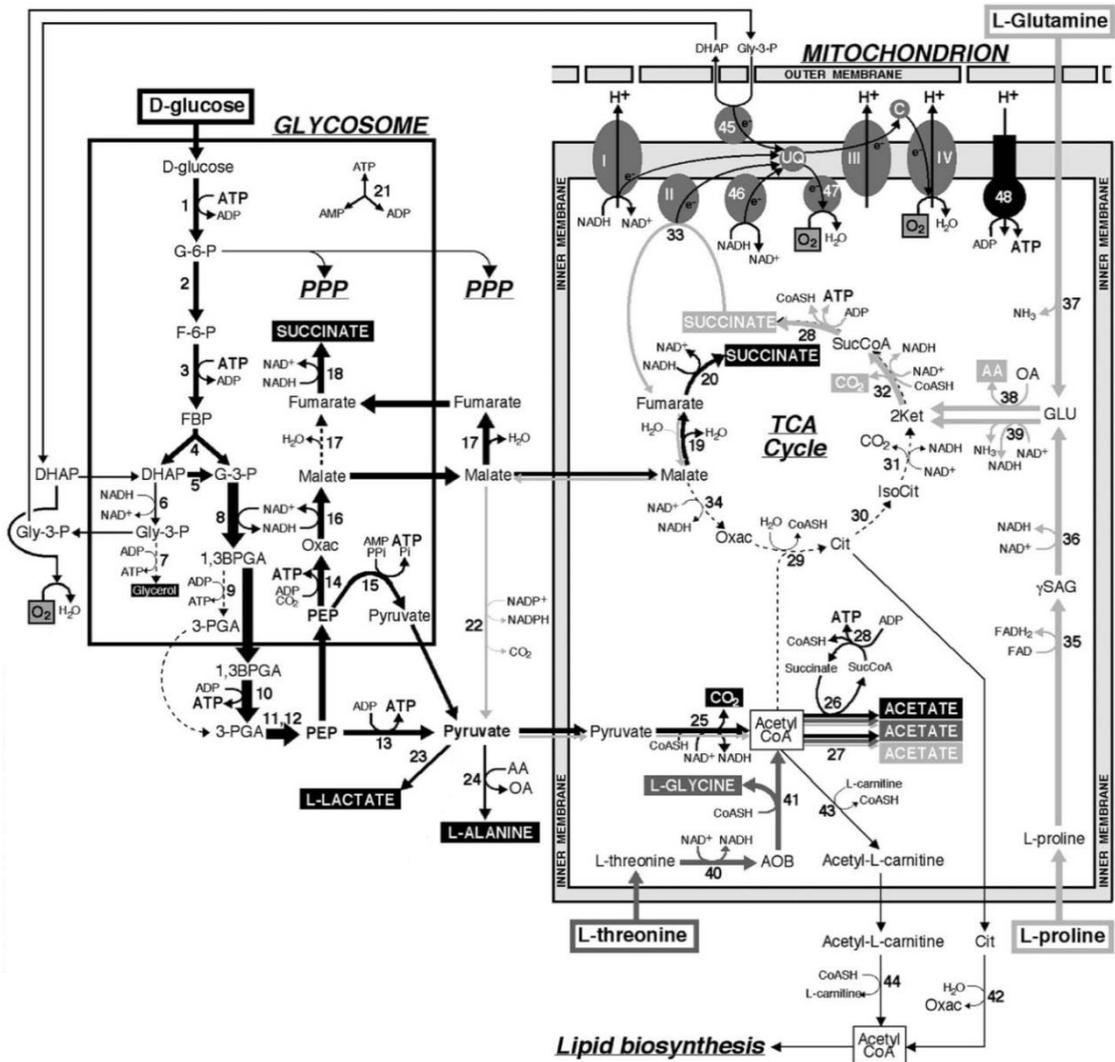


Figure 1.11. A schematic map detailing the metabolic pathways found in the procyclic form of *T. brucei*. End products are shown in white on a black background. Abbreviations are shown in Appendix I. Modified from Bringaud et al., (2006).

1.6 Ecology of Naegleria

The only free-living excavate protist for which a nuclear genome sequence is freely available is *Naegleria gruberi* (Fritz-Laylin et al., 2010b). The heterolobosean group to which *Naegleria* belongs is sister to the Euglenozoa (the group to which the trypanosomatids belong) in many phylogenetic analyses. *Naegleria* species have been isolated from freshwater and soil environments worldwide, including in Polar regions (De Jonckheere, 2004, 2006). *Naegleria* can be found as any of three morphological forms. These protists grow, feed and replicate as amoebae, preying on bacteria within the environment (although they can be cultured in the laboratory using axenic media (Fulton et al., 1984)). In response to various environmental cues, for instance oxygen levels, pH and prey density, the amoebae rapidly differentiate into a biflagellate form (Dingle and Fulton, 1966; Fritz-Laylin et al., 2010a; Fulton, 1977; Fulton and Walsh, 1980). In order for this differentiation process to take place, the actin-based cytoskeleton is replaced by a microtubule-based cytoskeleton which requires *de novo* assembly of the two basal bodies from which flagellum growth is templated (Fritz-Laylin et al., 2010a; Fritz-Laylin and Cande, 2010). *Naegleria* cannot replicate or feed in its flagellate form, but the two flagella are able to locomote the cell to potentially more hospitable environments more quickly than the amoebal form can crawl. Finally, encystment can be stimulated by a wide variety of biotic or abiotic factors such as overcrowding in a population niche, complete absence of food or a significant increase in salt concentration. *Naegleria* cysts are refractory to damage from desiccation, rapid changes in temperature and nutrient withdrawal (De Jonckheere and van de Voorde, 1976; Marciano-Cabral, 1988). Pores are present in the walls of the cysts, and these are sealed by a mucoid plug (Schuster, 1975). The pores allow for rapid escape of the excysting *Naegleria*, once environmental conditions become more favourable for growth and division. Triggers for excystment include the introduction of water or increased concentrations of carbon dioxide, which serves as an indicator for elevated numbers of bacterial prey (Averner and Fulton, 1966). Once excysted, the amoebae leave behind their cyst walls.

Whilst *N. gruberi* is non-pathogenic to humans, *N. fowleri* is a deadly opportunistic pathogen. *N. fowleri* is a moderate thermophile which can replicate in temperatures of up to 45°C (Visvesvara et al., 2007). Within the environment, *N. fowleri* is easily found in warm fresh water including lakes, streams and rivers in at least North America, central Europe, Asia and Australia. It can also be isolated from the runoff emanating from nuclear power stations

(Huizinga and McLaughlin, 1990; Jamerson et al., 2009). Human infection can arise following exposure to fresh, warm water environments where the amoebal load is high.

The amoebal trophozoites enter the body through the nasal passages and are thought to be phagocytosed by the structural cells that line the olfactory epithelium (Visvesvara et al., 2007). The trophozoites then migrate along the olfactory nerves, pass through the small holes in the cribriform plate and penetrate the delicate arachnoid mater to enter the subarachnoid space. From there, the trophozoites can penetrate the pia mater and enter the brain parenchyma (Visvesvara et al., 2007). Once *N. fowleri* reaches the brain it begins to multiply and cause necrosis of the surround tissue, which initiates a strong inflammatory immune response and haemorrhaging. This infection is referred to as primary amoebic meningoencephalitis (PAM) (Cervantes-Sandoval et al., 2008; Marciano-Cabral and Cabral, 2007; Visvesvara et al., 2007).

Fortunately, infection of people by *N. fowleri* is rare, but sadly the fatality rate from infection is greater than 95%; many of the victims of PAM are children (De Jonckheere, 2011). The earliest recorded fatalities from *N. fowleri* infections date back to 1961 (Fowler and Carter, 1965) and as well as infections stemming from bathing in open warm water, *N. fowleri* infections have also occurred as a consequence of bathing in swimming pools when chlorination has not been optimum and poor maintenance of the pool structure allows the amoebae to persist in concrete crevices (Kadlec et al., 1978). Unfortunately, there are no effective treatments for *N. fowleri*-dependent PAM and with death ensuing rapidly following the onset of symptoms (as quickly as within a week of infection), rapid diagnosis of *N. fowleri* as the cause of infection would also be required for any new anti-*Naegleria* medicine to be effective.

Sequencing and annotation of the nuclear genome of *Naegleria gruberi* NEG-M has highlighted the extent of *Naegleria's* potential metabolic flexibility (Fritz-Laylin et al., 2011; Fritz-Laylin et al., 2010b; Ginger et al., 2010). On the one hand, the *Naegleria* genome contains genes encoding enzymes and proteins required for a mitochondrial respiratory chain for aerobic respiration similar to that seen in *T. brucei* and a complete Krebs cycle (Fritz-Laylin et al., 2010b; Opperdoes et al., 2011). The first step of glycolysis in many organisms is catalysed by hexokinase; *Naegleria*, however, lacks this enzyme and instead has glucokinase which has specificity only for glucose but not other carbohydrates. *N. gruberi* also lacks the second glycolytic enzyme, ATP-dependent phosphofructokinase (PFK), and has pyrophosphate- (PP_i-) dependent PFK in its place – PP_i-dependent PFK is commonly

associated with anaerobes and prior to genome sequencing, was the only clue for a potential for anaerobic metabolism in this ubiquitous amoeba.

On the other hand, in addition to the potential for classic aerobic respiration, *N. gruberi* also appears to possess genes for a complex anaerobic metabolism too (Fritz-Laylin et al., 2010b). In the absence of O₂, it is possible that in *Naegleria* NO₃⁻ or NO₂⁻ respiration may take its place as the terminal electron acceptor, as indicated by the identification of homologues to fungal nitrite reductase (NirK) enzymes (Ginger et al., 2010). Denitrification consists of four reduction steps as shown in the equation: NO₃⁻ → NO₂⁻ → NO → N₂O → N₂ to produce gaseous nitrogen from nitrate and/or nitrite ions and such systems are closely associated with a wide range of bacteria and certain denitrifying fungi such as *Fusarium oxysporum* and *Aspergillus oryzae* (see Section 1.4). Curiously, the NirK enzyme in *Naegleria* is encoded for in the nuclear genome, along with four putative adenylate cyclases that have been identified to contain nitrate-nitrite-sensing (NIT) domains (Fritz-Laylin et al., 2010b; Ginger et al., 2010). NIT domains are found in certain bacteria and recognise NO₃⁻ and NO₂⁻ ions for the regulation of cellular processes such as gene expression, enzyme activity and metabolism, dependent upon nitrate and nitrite concentrations (Shu et al., 2003). This perhaps points to a small metabolic network centred upon NO₂⁻ detection and metabolism in *Naegleria*.

Another method of anaerobic respiration may be available to *Naegleria* by using substrate-level phosphorylation of ADP to generate ATP using the enzyme acetyl CoA synthetase (ADP forming) (Fritz-Laylin et al., 2010b; Ginger et al., 2010). Under anaerobic conditions, *i.e.* in the absence of a functional Krebs cycle, the acetyl CoA derived from pyruvate could potentially be converted to acetate in order to generate ATP as shown in the following equation: acetyl CoA + ADP + P_i ⇌ acetate + CoA + ATP, as seen in the anaerobic amoebozoan parasite *Entamoeba histolytica* (Jones and Ingram-Smith, 2014).

When oxygen and nitrate are both absent, the genome sequence of *N. gruberi* suggests a capacity to produce molecular hydrogen by hydrogen fermentation. Thus the nuclear genome revealed the presence of a FeFe-hydrogenase enzyme, along with its three associated post-translational maturases (HydE, HydF and HydG) that contain N-terminal mitochondrial transit peptides (Fritz-Laylin et al., 2010b; Ginger et al., 2010). In a fermentation role, *i.e.* presumably in *Naegleria* anaerobic metabolism, FeFe-hydrogenase catalyses the formation of molecular hydrogen, using protons from reduced electron carriers (such as NADH) that acted as electron acceptors during fermentation (Vignais and Billoud, 2007). Due to the presence of the *Naegleria* maturation system, it suggests that Fe-

hydrogenase activity would be organellar rather than cytosolic (Fritz-Laylin et al., 2010b). Whilst I was carrying out some of the experiments described in this thesis, it was suggested that laboratory-adapted *N. gruberi* is indeed capable of a low rate of H₂ production (Tsaousis et al., 2014), although this work also gave FeFe-hydrogenase localisation that is contradictory to the data I show in Chapter 6.

1.7 Overview of thesis results

In the results chapters that follow, I first present data from experiments to determine the function of a paralogous pair of novel mitochondrial histidine phosphatases in *T. brucei*. Intriguingly, orthologues of these proteins are present in only a few eukaryotes including other excavates with aerobic mitochondria. Finally, in Chapter 6, I will report results of experiments designed to determine the localisation of enzymes from *Naegleria*'s predicted repertoire of anaerobic metabolism.

Chapter 2

Materials and Methods

2.1 Buffers & solutions (listed in alphabetical order)

Agarose Gel Electrophoresis 6x Sample Loading Buffer – 0.15% orange G, 30% glycerol in H₂O

Blocking Buffer A – 1x PBS, 0.05% Tween-20, 5% Dried Skimmed Milk

Blocking Buffer B – 1x PBS, 0.01% BSA

BPP Buffer B – 100 mM NaH₂PO₄, 10 mM Tris·Cl, 8 M Urea, pH 8.0

BPP Buffer C – 100 mM NaH₂PO₄, 10 mM Tris·Cl, 8 M Urea, pH 6.3

BPP Buffer D – 100 mM NaH₂PO₄, 10 mM Tris·Cl, 8 M Urea, pH 5.9

BPP Buffer E – 100 mM NaH₂PO₄, 10 mM Tris·Cl, 8 M Urea, pH 4.5

Column Wash Buffer* – 2 M NaCl, 50 mM Tris·Cl (pH 7.5)

Coomassie Brilliant Blue Stain – 45% Methanol, 10% Glacial Acetic Acid, 45% H₂O, 0.05% Coomassie Brilliant Blue R-250

Coupling Buffer – 0.1 M Na₃C₆H₅O₇, 0.05 M Na₂CO₃ (pH 7.2)

Destain Solution – 45% Methanol, 10% Glacial Acetic Acid, 45% H₂O

Elution Buffer* (EB) – 10 mM Tris-HCl, pH 8.0

Equilibration Buffer* - 150 mM NaCl, 50 mM Tris·Cl (pH 7.4)

Genomic DNA (gDNA) extraction buffer – 50 mM Tris, 100 mM NaCl, 1 mM EDTA

Hemin Stock* (for SDM-79) – 4% Hemin in 0.1 M NaOH

IMAC Binding Buffer* - 8 M Urea, 29 mM Tris·Cl (pH8.0), 500 mM NaCl, 0.02% Triton-X, 25 mM Imidazole, 10% Glycerol

IMAC Denaturing Buffer* - 6 M Guanidine, 20 mM Tris·Cl (pH 8.0), 500 mM NaCl, 0.02% Triton-X, 25 mM Imidazole, 10% Glycerol

IMAC Elution Buffer* - 8 M Urea, 29 mM Tris·Cl (pH8.0), 500 mM NaCl, 0.02% Triton-X, 500 mM Imidazole, 10% Glycerol

IMAC Nickel Sulphate* – 100 mM NiSO₄

Inhibitor Cocktail – Halt Protease Inhibitor single-use Cocktail (ThermoFischer), 200 mM TLCK, 200 mM PMSF, 1 M Benzamidine

Laemmli Buffer – 100 mM Tris-Cl (pH 6.8), 200 mM Dithiothreitol (DTT), 4% SDS, 0.2% Bromophenol Blue, 20% Glycerol

LB Agar - 10 g/l Tryptone, 5 g/l NaCl, 5 g/l Yeast Extract, 1.5% (w/v) Agar

LB Broth (Low Salt) – 10 g/l Tryptone, 5 g/l NaCl, 5 g/l Yeast Extract

Lysis Solution (Fermentas) – GeneJET™ Plasmid Miniprep Kit; GeneJET Plasmid Maxiprep Kit

MS-ABC Solution – 50 mM Ammonium Bicarbonate

MS-Elute Solution – 0.1% Trifluoroacetic acid, 70% Acetonitrile

MS-IAA Solution – 8 M Urea, 1% SDS, 50 mM Iodoacetamide

MS-IG-Extraction Buffer – 1:2 (v/v) 5% Formic Acid : 100% Acetonitrile

MS-IG-IAA Solution – 55 mM Iodoacetamide, 100 mM Ammonium Bicarbonate

MS-Solubilisation Buffer – 8 M Urea, 1% SDS

MS-TFA Solution – 0.1% Trifluoroacetic acid

M7 Axenic Media* - 308 mM L-methionine (1.5 g/l), 29 mM D-glucose (270 g/l), 0.02% Buffer Solution ([133 mM KH₂PO₄ [18.1 g/l]], [176 mM Na₂HPO₄ [25 g/l]]), Yeast Extract (25 g/l), 10% heat-inactivated foetal calf serum. Stored at 4°C

Neutralization Solution (Fermentas) – GeneJET™ Plasmid Miniprep Kit; GeneJET Plasmid Maxiprep Kit

PBS-Tween – 1x PBS, 0.05% Tween-20

Protein Transfer Buffer – 39 mM Glycine, 48 mM Tris, 0.037% SDS, 20% Methanol

Quenching Buffer – 1 M Tris.Cl (pH 7.4)

ReddyMix® PCR Master Mix (Thermo Prime) – 1.25 U ThermoPrime Plus DNA Polymerase, 75 mM Tris-Cl (pH 8.8), 20 mM (NH₄)₂SO₄, 1.5 mM MgCl₂, 0.01% (v/v) Tween-20, 0.2 mM dATP, 0.2 mM dCTP, 0.02 mM dGTP, 0.02 mM dTTP, Precipitant and Red Dye for Electrophoresis

Resuspension Solution (with RNaseA added) (Fermentas) - GenJET™ Plasmid Miniprep Kit; GeneJET™ Plasmid Maxiprep Kit

SB-1 Buffer – 0.25 M HCl

SB-Denaturation Buffer – 0.5 M NaOH, 1.5 M NaCl

SB-Neutralisation Buffer – 1 M Tris HCl (pH 8.0), 1.5 M NaCl

SB-SSC (Standard Saline Citrate) 20x – 3 M NaCl, 300 mM Trisodium Citrate

SB-Hybridization Buffer – 12% Urea, 0.5 M NaCl, 4% Blocking Reagent (from Amersham AlkPhos Direct Labelling and Detection System Kit [RPN3690])

SB-Primary Wash Buffer – 2 M Urea, 0.1% SDS, 0.5 M NaH₂PO₄·H₂O, 150 mM NaCl, 1 mM MgCl₂, 0.2% Blocking Reagent (from Amersham AlkPhos Direct Labelling and Detection System Kit [RPN3690])

SB-Secondary Wash Buffer 20x* – 1 M Tris, 2 M NaCl, pH 10.0

SDM-79* - SDM-79 powder (Gibco) (Brun and Schonenberger, 1979) hydrated in 5 L of H₂O with 2 g/l sodium bicarbonate and filtered into 500 ml aliquots (pH 7.3). Each aliquot supplemented with 7.5 mg/l haemin (10 mg/ml stock in 0.1 M NaOH) and 50 ml heat inactivated foetal calf serum before use. Everything stored at 4°C

SDS-PAGE 8% Tris-glycine Resolving Gel – 8% Acrylamide mix, 1.5 M Tris-Cl (pH 8.8), 8% SDS, 0.04% Ammonium persulphate, 0.1% N, N, N', N',-tetramethylethylenediamine (TEMED)

SDS-PAGE 5% Stacking Gel – 5% Acrylamide, 1 M Tris (pH 6.8), 10% SDS, 10% Ammonium Persulphate, 0.02% TEMED

SDS-PAGE Running Buffer (5x Stock) – 125 mM Tris, 1.25 M Glycine, 0.5% SDS

SDS-PAGE Transfer Buffer – 48 mM Tris, 39 mM Glycine, 0.037% SDS, 20% Methanol

Slide Fixative – 3.7% (w/v) PFA in 1x PBS

S.O.C medium – 2% Tryptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM Glucose

SS-Fixation Buffer – 50% Ethanol, 10% Glacial Acetic acid

SS-Rehydration Buffer – 3% Ethanol, 0.83 M Sodium Acetate, 0.00125% Gluteraldehyde, 0.5mM Sodium Thiosulphate (Na₂S₂O₃·5H₂O)

SS-Silver Solution – 5.89 mM Silver Nitrate, 2.66 mM Formaldehyde

SS-Developer Solution – 236 mM Sodium Carbonate, 1.33 mM Formaldehyde

SS-Stop Solution – 0.04 M EDTA, Disodium Salt, Dihidrate (C₁₀H₁₄N₂NA₂O₈·2H₂O)

TAE Buffer (1x) – 40 mM Tris Acetate, 1 mM EDTA (pH 8.5)

TBE Buffer (0.5x) – 45 mM Tris-borate, 1 mM EDTA

Wash Buffer – 1x PBS, 0.05% Tween-20, 5% Dried Skimmed Milk, 0.5% BSA

Wash Solution (concentrated) (with EtOH added) (Fermentas) – GeneJET™ Plasmid Miniprep Kit; GeneJET Plasmid Maxiprep Kit

ZMG Buffer – 132 mM NaCl, 8 mM KCl, 8 mM Na₂HPO₄, 1.5 mM KH₂PO₄, 0.5 mM Magnesium acetate, 0.09 mM Calcium acetate, 1% Glucose, pH 7.0

* Required Filtration before Use

2.2 Antibiotic and drug stock solutions

Table 2.1. The antibiotics and drugs used and the filtered stock solutions they were made up at.

Antibiotic:	Working Concentration:	Use:
Ampicillin (Sigma)	100 µg/ml	Selectable marker for <i>E. coli</i>
Blasticidin (Sigma)	10 µg/ml	Selectable marker for trypanosomes
Chloramphenicol (Sigma)	25 mg/ml (in 100% ethanol)	Selectable marker for <i>E. coli</i>
Doxycycline (Sigma)	1 µg/ml	Inducible expression in trypanosomes
Hygromycin (Sigma)	20 µg/ml	Selectable marker for trypanosomes
Kanamycin	10 mg/ml	Selectable marker for <i>E. coli</i>
Phleomycin (Sigma)	7.5 µg/ml	Selectable marker for trypanosomes
Puromycin (Sigma)	1 µg/ml	Selectable marker for trypanosomes

2.3 Antibodies

Table 2.2. The antibodies used in Western blotting and fluorescence microscopy.

Antibody:	Animal Raised in:	Dilution:	Source:
anti-HA	Mouse	1/100	Abcam
anti-myc	Mouse	1/1000	Abcam
anti-GFP	Mouse	1/100	Roche
anti-mouse IgG conjugated to Rhodamine	Goat	1/200	Chemicon
anti-mouse IgG conjugated to FITC	Goat	1/200	Chemicon
Streptavidin conjugated to TRITC	N/A	1/200	Abcam
anti- 6x His	Mouse	1/5000	Clontech
NgFeHyd	Rabbit	1/250	Covalab
FeHyd ^{Δ274-752}	Chicken	1/1000	Covalab
anti-mouse IgG HRP conjugate	Goat	1/80000	Sigma
anti-rabbit IgG HRP conjugate	Goat	1/30000	Dako
Streptavidin conjugated to HRP	N/A	1/10000	Abcam

2.4 Oligonucleotide primers

All custom oligonucleotides were ordered from Eurofins MWG Operon and resuspended in filtered EB buffer (Fermentas) to 100 μ M prior to long-term storage at -20°C . Oligonucleotides were then diluted to 10 μ M prior to use in PCR.

Table 2.3. Custom primer sequences used in PCR reactions. F, forward primer; R, reverse primer; * denotes primers used in gene knockout experiments in *T. brucei*.

Primer Name	Primer Sequence (5' to 3')
40ORF F*	cgcgatatcatgtcacgatgtggcgcaagctggatatg
40ORF R*	atactcgagccggtccgttcctcgctgcggaaaagg
<i>Tb</i> HP30HA F	cgcgatatcatggcagttgtacgcaggtg
<i>Tb</i> HP30HA R	atactcgagcatgaaaggcctccagtactc
UpIG0445 F*	ctatgagtctaaactgagcg
UpIG0445 R*	ccgtcatgatgcaaggatc
30ORF F*	tatggatcctcaaccaactaccagacagtg
30ORF R*	tatctcgagtcaactgtcaggttaaacc
4030KOA F*	aagctggatatgtcgagtggttactggctctgggtcccattgttgctc
4030KOB R*	ttgattgtgatatgtcacgatgtggcgcaagctggatatgtcgagtgg
4030KOC F*	tcccagtagctgttgttcgatattgtatattttatggcagcaacg
4030KOD R*	tcacatgaaaggcgtccagtactcatggatcccagtagctgttggc
4030KO5 F*	catatgcacatgtacacacg
4030KO6 R*	atctccctgtacacttctgg
<i>NgHydE</i> F	ccgaagcttatgatgaaggattgaaaaatcagc
<i>NgHydE</i> R	atgctcgagttcctgttcggtactttc
<i>NgHydF</i> F	gccaagcttatgaaatcacatcaacaacacttgatcc
<i>NgHydF</i> R	ttgctcgagttgagatggagtgaagg
<i>NgHydG</i> F	gggaagcttatgaagagtttagctagattgg
<i>NgHydG</i> R	ccactcgagatagtaaatatctctgacaccttcagc

Synthetic genes were also ordered as appropriate from Eurofins MWG Operon and resuspended in filtered EB buffer prior to long term storage at -20°C .

2.5 Polymerase chain reaction

Standard (*i.e.* for some cloning, for amplification of probes for Southern blot analysis and for mapping stable integrations in *T. brucei*) PCR reactions were set up using Reddymix® PCR Master Mix (ThermoPrime), 1 µl of 10 µM forward primer, 1 µl of 10 µM reverse primer, 3 µl of gDNA template (~100 pg) and 7.5 µl of H₂O. The standard PCR program used is described in Table 2.3.

For high fidelity PCR reactions, the Expand™ System (Roche) was used and the following set up: 5 µl of buffer with 15 mM MgCl₂ (10x), 1 µl of dNTPs (10 mM), 1 µl of 10 µM forward primer, 1 µl of 10 µM reverse primer, 3 µl of gDNA template (~100 pg), 1 µl Expand™ Taq DNA Polymerase (2.5 U) and 38 µl H₂O. The standard PCR program used is described in Table 2.3.

PCR products were purified using the GeneJET™ PCR Purification Kit (Fermentas) following the purification protocol as described in Thermo Scientific GeneJET PCR Purification Kit #K0701, #K0702.

Table 2.4. Standard PCR program used.

Number of Cycles	Stages	Temperature (°C)	Time
1	Initial Denaturation	94	5 min
25	Denaturation	94	30 s
	Annealing	50	50 s
	Elongation	72	1 min 50 s
1	Final Extension	72	5 min
-	Hold	4	-

2.6 Agarose gel electrophoresis

Gel electrophoresis was used to analyse PCR products and sub-cloning of DNA between plasmids. 0.8% agarose gels were cast using 1x TAE. Samples were loaded with 1x DNA loading buffer and 4 µl GeneRuler™ 1 kB DNA ladder (Fermentas) was used as standard molecular weight marker. Gels were run at 70 V for approximately 45 min and post-stained using Gel Red (at a concentration of 1 µl per 20 ml of 1x TAE). Gels were imaged using the Gel Doc EZ System (Bio-Rad).

2.7 DNA ligation

Ligation reactions for the insertion of purified PCR amplicons into pGEM[®]-T Easy (Promega) (Appendix II Figure B.1) were prepared using an insert:vector ratio of 3:1, along with T4 DNA ligase and 2x Rapid Ligase Buffer at room temperature overnight. The protocol described in the Promega Technical Manual - pGEM[®]-T and pGEM[®]-T Easy Vector Systems followed precisely.

For protein expression, DNA inserts were cloned into pET-28a (Novagen) (Appendix II Figure B.2). Inserts were prepared for cloning by release from T Easy vectors using stated restriction enzymes and ligated with pET-28a vector that had previously been digested with the same enzymes, and then purified by gel extraction. Ligations were again prepared using an insert:vector ratio of 3:1, along with T4 DNA ligase and 2x Rapid Ligase Buffer at room temperature overnight.

2.8 Transformations of *E. coli*

XL-1 Blue Subcloning-Grade competent cells (Stratagene) were used for most routine cloning, with Rosetta™ 2(DE3)pLysS competent cells (Novagen) used for induction of recombinant protein expression.

Cells were thawed on ice and 100 µl incubated with the appropriate recombinant plasmid on ice for 30 mins. The cells were heat-shocked at 42°C for 1 min and recovered on ice for 5 min. For transformations into Rosetta cells only - recovery from the heat-shock was necessary before streaking onto LB agar plates and this was done by incubation in SOC media at 37°C for 1 h. Cells were streaked out onto an LB agar plate supplemented with ampicillin (XL-1 Blue) or kanamycin/chloramphenicol (Rosetta) and incubated at 37°C overnight. Transformants were selected and were used to inoculate 5 ml LB broth supplemented with ampicillin or kanamycin/chloramphenicol. Cultures were incubated overnight in a shaking incubator at 250-300 rpm at 37°C.

2.9 Extraction and purification of plasmid DNA

For minipreps, plasmid DNA was isolated from the *E.coli* cells using the GeneJET™ Plasmid Miniprep Kit (Fermentas) following the protocol laid out in Thermo Scientific GeneJET™ Plasmid Miniprep Kit #K0502. Purified DNA was eluted in EB buffer and stored at -20°C.

For maxipreps, one colony lifted from the plate of a fresh transformation was firstly used to inoculate 5 ml of LB broth (supplemented with ampicillin) and grown overnight in a shaking incubator at 250-300 rpm at 37°C. This 5 ml culture was then used to inoculate 100 ml of LB broth (supplemented with ampicillin) and grown overnight in a shaking incubator at 250-300 rpm at 37°C. Plasmid DNA extracted using the GeneJET™ Plasmid Maxiprep Kit (Fermentas) following the protocol described in Thermo Scientific GeneJET™ Plasmid Maxiprep Kit #K04921, #K0492. The purified DNA was eluted in EB buffer and stored at -20°C.

2.10 Ethanol precipitation of plasmid DNA

Ethanol precipitation of DNA was carried out as follows: 3 M sodium acetate (pH 5.2) at 0.1x volume of the original DNA volume and 100% ethanol at 2.5x volumes the original eluted DNA volume were added sequentially to the DNA, which was then incubated minimally for 20 min at -20°C. Eppendorfs containing precipitated DNA were centrifuged at 6700 x g for 15 min at 4°C and the supernatant discarded. 1 ml of ice cold 75% ethanol was added and centrifuged again at 6700 x g for 5 min at 4°C. The supernatant was removed and the pellet resuspended in 50 µl of sterile EB buffer.

2.11 Restriction digests of DNA

Plasmids and genomic DNA (gDNA) were digested using restriction enzymes for purposes such as screening for successful cloning, linearisation of a plasmid for trypanosome transfection, isolation of DNA for purification and further sub-cloning into other vectors and Southern mapping. For plasmids, 10 U of the desired enzyme(s) (Roche or Fermentas) was used to digest ~1 µg DNA. For gDNA, ~1/6th of the gDNA extracted from 1x10⁸ cells was digested overnight with 40 U of enzyme in a final volume of 60 µl. 20 U of enzyme in an additional 20 µl was added to the overnight digest the following morning, and incubated for a further 2 h.

Agarose gel electrophoresis was used to analyse completed restriction digests (Section 2.6)

2.12 Gel extraction

The DNA bands of interest were cut out of the agarose gel using a scalpel under a UVP Dual-Intensity Transilluminator. DNA was extracted from the agarose gel using the GeneJET™ Gel Extraction Kit (Fermentas) following the QuickProtocol™ GeneJET™ Gel Extraction Kit, #K0691, #K0692. The purified DNA was eluted in 50 µl EB buffer and stored at -20°C.

2.13 Induction of recombinant protein expression

Small scale induction was always initially carried out to assess the dynamics and extent of recombinant protein expression in Rosetta clones, and to determine protein solubility. Colonies were selected from Rosetta LB agar plates and grown overnight in liquid culture (in the presence of chloramphenicol). 100 µl of the stationary phase culture was then used to inoculate each of the two 10 ml aliquots of LB broth containing no antibiotics. Cultures were incubated in a shaking incubator at 250-300 rpm at 37°C until they reached a cell density of OD₆₀₀ 0.4-0.6. Once the cultures reached a sufficient density, a 1 ml sample was taken and centrifuged at 2500 x *g* for 5 min, the supernatant discarded and the pellet stored at -20°C. Recombinant protein expression was induced using 1 mM IPTG. Cultures were incubated at 37°C in the shaking incubator for 3 h, with 1 ml samples taken at hourly intervals. SDS-PAGE and Coomassie Brilliant Blue staining (Section 2.20) was used to analyse recombinant protein expression over the three hour induction period.

Large scale induction was employed to purify sufficient recombinant protein for antibody production. Here 10 ml of overnight culture inoculated from a single fresh colony was used to inoculate 800 ml of LB broth (no antibiotics) and the culture incubated at 37°C in the shaking incubator until a cell density of OD₆₀₀ 0.4 - 0.6 was reached. Protein expression was once again induced with IPTG and the culture incubated for 3 h at 37°C in the shaking incubator. The entire culture (except for 10 ml which was harvested separately for a test of solubility) was centrifuged at 2500 x *g* for 25 min and the supernatant removed. Cell pellets were stored at -20°C before protein purification was carried out.

2.14 Determining protein solubility

Establishing the solubility of the recombinant protein was essential as it dictated what conditions were required in further downstream purification steps. The 10 ml samples taken at the 3 h time-point post induction (Section 2.13) were resuspended in 1 ml of Protein Solubility Buffer A. The suspension was sonicated on ice; using 10 sec bursts at amplitude 6, with 10 sec intervals in between, until no clumps of cells remained. The samples were centrifuged at $6700 \times g$ for 25 min at 4°C then 50 μl of the supernatant added to 50 μl of Laemmli buffer (with 10% final volume of DTT). The supernatant was decanted off into a separate 50 ml falcon tube and saved for possible use later on depending on the protein solubility result, and the remaining pellet resuspended in 1 ml Buffer A. 50 μl of the resuspended pellet was added to Laemmli buffer (with 10% final volume of DTT).

Both the supernatant and pellet samples were analysed using SDS-PAGE and Coomassie Brilliant Blue (Section 2.20).

2.15 Protein sample preparation for IMAC purification under denaturing conditions

For purification under denaturing conditions of insoluble protein, the cell pellet from a large scale induction was resuspended at 4°C in the IMAC denaturing buffer at a volume of 20 ml/1000 ml of culture. 100 mM PMSF protease inhibitor was added (10 μl per 1 ml of sample) and the sample sonicated on ice using 60 sec bursts at amplitude 18, with 60 sec intervals. The sample was centrifuged at $30000 \times g$ for 2 h and the supernatant decanted off immediately. The supernatant was filtered using a syringe and a 0.20 μm syringe filter.

2.16 IMAC protein purification under denaturing conditions

The filtered sample containing the plasmid for antibody production was isolated using immobilised metal-ion affinity chromatography (IMAC). A 5 ml HisTrap HP column (GE Healthcare) was charged with IMAC nickel sulphate then equilibrated with 4 x column volumes of IMAC denaturing buffer. The filtered protein sample (Section 2.15) was then injected onto the column and any proteins not bound by the hexahistidine tag were washed off using 4 x column volumes of IMAC binding buffer. Proteins were then eluted in a series of fractions (10 x column volumes) using an imidazole concentration gradient (25 mM-500 mM) using the IMAC elution buffer, followed by a further 4 x column volume washes with IMAC

elution buffer. The fractions were collected in 15 ml falcon tubes, numbered according to the order in which they were eluted and stored at -80°C. Fractions of interest were analysed using SDS-PAGE and the protein isolated as Coomassie-blue gel slices on large SDS-PAGE gels (Section 2.20).

2.17 Protein sample preparation for benchtop purification

The cell pellet from the large scale induction was resuspended at room temperature in BPP Buffer B at 5 ml per gram wet weight and gently vortexed until completely lysed. The lysate was centrifuged at 10000 x *g* for 30 min and the supernatant transferred into a fresh tube. 50 µl of sample was taken and added to 50 µl of boiling hot Laemmli buffer and stored at -20°C for future analysis by SDS-PAGE.

2.18 Benchtop protein purification under denaturing conditions

The recombinant protein from the induced *E. coli* cleared lysate (Section 2.14) was again isolated using its hexahistidine tag, but this time eluted using an acid shift rather than an imidazole concentration gradient. 0.5 ml of 50% Ni-NTA slurry (Qiagen) was added to 4 ml of cleared lysate and this was mixed gently on a rotating wheel for 60 min at room temperature. The lysate-resin mixture was then transferred into a Gravity Flow Column (Bio-Rad) and the flow-through saved for future analysis. The column was washed twice with BPP buffer C. The recombinant protein was eluted by 4 x column washes with BPP buffer D, followed by 4 x column washes with BPP buffer E. All washes and elutions were collected in Eppendorfs and 50 µl samples taken for each and added to 50 µl of boiling hot Laemmli buffer. These protein samples were stored at -20°C for future analysis by SDS-PAGE.

2.19 Polyclonal antibody production

Purified proteins (Sections 2.13-2.18) were sent to Covalab for polyclonal antibody production using chicken hosts.

Raw sera from both chickens were tested for reactivity against recombinant FeHyd^{A274-752} protein samples using Western blotting (Section 2.23).

2.20 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

8% resolving gels with an additional 5% stacking gel were cast in Bio-Rad plates and protein samples separated by electrophoresis using a Mini-PROTEAN® 3 cell (Bio-Rad). Protein samples had previously been resuspended in Laemmli buffer and stored at -80°C. The protein samples were heated at 100°C for 5 min prior to analysis and loaded alongside 6 µl of PageRuler™ Plus Prestained Protein Ladder (Fermentas). Electrophoresis was carried out in 1x SDS-PAGE running buffer at a constant 200V for approximately 1 h.

Large 8% resolving gels with an additional 5% stacking gel were cast using Bio-Rad plates and samples separated by electrophoresis using a PROTEAN II xi cell (Bio-Rad) and electrophoresis was carried out 1x SDS-PAGE running buffer at a constant 200 V for approximately 4 h.

SDS-PAGE gels intended for further analysis by mass spectrometry were stained using Instant Blue (Expedeon) for 30 min in a petri dish, then destained with H₂O for 2 h in preparation for protein band excision.

SDS-PAGE gels not intended for further analysis by Western blotting or mass spectrometry were stained with Coomassie Brilliant Blue solution for 4 h, and then washed with Destain solution overnight in preparation for imaging using the BioRad ChemiDoc™ XRS⁺ Imaging System.

2.21 Silver staining

The SDS-PAGE gel was fixed in SS-Fixation solution for 30 min in a clean glass dish. The gel was then transferred into SS-Rehydration solution for 30 min and then washed for 3 x 5 min with H₂O. The H₂O was drained off and SS-Silver solution was added and incubated for 20 min. The SS-Silver solution was removed and the SS-Developer solution added for 2-10 min, until the staining of the bands was dark enough. Once sufficient staining had been obtained, the gel was transferred immediately into the SS-Stop solution for 10 min then washed for 3 x 5 min with H₂O. Silver stained gels were transferred into 10% glycerol for 30 min before being imaged using the BioRad ChemiDoc™ XRS⁺ Imaging System.

2.22 Southern blotting

Genomic DNA was harvested from *T. brucei* cell lines and digested with EcoRI. Agarose gel electrophoresis was carried out in the same manner as described in Section 2.6, except 0.7% agarose gels were cast using 0.5x TBE and were subject to electrophoresis using 0.5x TBE Running Buffer at 100 V for approximately 100 min. The gel was then prepared for wet transfer first by immersion into SB-1 buffer for 15 min with gentle rocking (to 'nick' the DNA in preparation for a more efficient transfer). The gel was then immersed in SB-Denaturation buffer for 2 x 15 mins with gentle rocking (which separated the negatively-charged DNA into separate strands in preparation for downstream hybridization). The gel was then rinsed in H₂O and neutralised using the SB-Neutralisation buffer for 2 x 15 mins with gentle rocking. The gDNA in the gel was then transferred by capillary action to a nylon membrane (Amersham Hybond N⁺) overnight at room temperature in 20x SB-SSC, following the Amersham Hybond N⁺ protocol.

Probing and detection hybridization was performed using the Amersham Gene Images AlkPhos Direct Labelling and Detection System (RPN3690). The protocol provided in the product booklet was always followed exactly.

PCR (Section 2.5) was used to generate the DNA needed for the hybridization probe using primers listed in Table 2.3. The DNA hybridization probe was carried out using the Amersham Gene Images AlkPhos Direct Labelling and Detection System (RPN3690) protocol.

A 30 ml aliquot of hybridization buffer was pre-heated to 55°C in the hybridization oven in a rotating hybridization tube. The gDNA on the Hybond N⁺ membrane was cross-linked using a UVC 500 UV cross-linker (set at 70000 µJ/cm²) (Amersham) and then the membrane placed in the hybridization tube with the pre-heated buffer and pre-hybridized for 15 min at 55°C. The labelled DNA hybridization probe was then added to the tube and hybridization was carried out overnight at 55°C with constant rotation. The following day, 1x SB-Primary Wash buffer was made and 50 ml pre-heated to 60°C. The 20x stock of SB-Secondary Wash buffer was diluted to 1x and supplemented with MgCl₂ (to a final concentration of 2mM). The post-hybridization stringency washes were all carried out using the rotating hybridization tube in the hybridization oven. The Hybond N⁺ membrane was washed with pre-heated 1x SB-Primary Wash buffer at 60°C for 2 x 15 min, and then washed for 10 min with 1x SB-Secondary wash buffer at room temperature. The 1x SB-Secondary wash was repeated twice more but for 5 min each time. The excess liquid was drained from the membrane and then incubated at room temperature for 3 min with 1 ml of CDP-Star detection reagent. The substrate was

drained and the membrane wrapped in saran wrap before being exposed to autoradiography film for 1 h.

The films were developed in the dark room under the red safety light. The film was first incubated in developer solution (Kodak) with gentle agitation for 1 to 5 min, until desired exposure of the bands was obtained. The film was then transferred immediately into fixation solution (Kodak) and gently agitated until it turned completely clear. It was then rinsed in H₂O and left to dry before being imaged using the Bio-Rad Chemidoc XRS⁺ Imaging System.

2.23 Western blotting

Proteins separated by SDS-PAGE were transferred onto Amersham Hybond™-P nitrocellulose membranes using a Mini Trans-blot® transfer cell (Bio-Rad) for 60 min at 100 V. The membranes had previously been prepared by dipping in 100% methanol for 10 sec to permeabilise, then followed by a wash in H₂O for 5 min, and finally equilibrated in transfer buffer for 10 min. Once Western transfers were complete, membranes were left to dry on the bench and were stored at 4°C between two sheets of blotting paper wrapped in saran wrap before probing.

For probing with primary antibody (refer to Table 2.3), Western membranes were re-permeabilised in 100% methanol for 10 sec, washed in H₂O for 5 min then rinsed in 1x PBS for 5 min; then incubated with blocking buffer A for 1 h at room temperature. Membranes were then incubated with the primary antibody, diluted in milk blocking buffer for 1 h at room temperature. Excess primary antibody was removed by 3 x 10 min washes in PBS-Tween and then the membranes were incubated with the appropriate secondary antibody (refer to Table 2.3), diluted in milk blocking buffer, for 1 h at room temperature. The blots were again washed for 3 x 10 min in wash buffer then allowed to drain. 2 ml of Immun-star™ WesternC™ (Bio-Rad) chemiluminescent substrate was added to the blots to detect any protein bound to the membranes, and incubated for 5 min at room temperature. The blots were drained, wrapped in saran wrap and the chemiluminescent signal was detected and imaged using the Bio-Rad Chemidoc XRS⁺ Imaging System.

2.24 Naegleria gruberi (NEG-M) cell culture

Naegleria gruberi (NEG-M) trophozoites were cultured in M7 axenic media supplemented with 10% heat-inactivated foetal calf serum and incubated at 20°C.

2.25 Naegleria gruberi (NEG-M) protein samples

For *N. gruberi*, a 50 ml culture (with a density between $10^6 - 10^7$ cells ml⁻¹) was centrifuged at 1000 x g for 10 min at 4°C, the supernatant removed and the pellet resuspended in 10 ml of 1x PBS, centrifuged at 1000 x g for 10 min at 4°C, the PBS wash discarded and the PBS wash repeated. The pellet was then resuspended in 200 µl of boiling hot Laemmli buffer and heated at 100°C for 5 min. Protein samples were stored at -80°C.

2.26 Trypanosoma brucei cell culture

Procyclic form trypanosome strains were cultured in SDM-79 medium (Brun and Schonberger, 1979) supplemented with 10% heat-inactivated foetal calf serum (FCS) and 8 µM hemin and incubated at 28°C. The non-genetically modified 427 procyclics were grown in the presence of no selectable markers; cell lines transfected with plasmid constructs containing genes encoding *TbHP40::myc₃*, *TbHP30::myc₃*, *TbHP30::GAFSINPAM*, *FeHyd^{Δ61-752}:HA:GFP*, *HydE::myc₃*, *HydF::myc₃*, *HydG::myc₃*, *NirK:HA* or *ACS^{Δ101-1100}:HA:GFP*. In all of these plasmid constructs the gene encoding the epitope-tagged protein replaced the luciferase gene in plasmid pDex377 (Appendix II Figure B.3 A) were grown in the presence of hygromycin; the *TbHP40::myc::BirA** construct, in which the BirA* fusion gene replaced the luciferase gene in plasmid pLew100v5b1d-HYG (Appendix II Figure B.4), was also grown in the presence of hygromycin. Cell lines transfected with *TbHP40::myc₃* with a pDex477 vector backbone (Appendix II Figure B.3 B) were grown in the presence of phleomycin; gene knockout cell lines using the blasticidin drug selectable marker were cultured in the presence of blasticidin and cell lines using the puromycin drug selectable marker were cultured in the presence of puromycin. The SmOx-P9 strain (Poon et al., 2012) was cultured in the presence of puromycin to maintain the tetracycline repressor protein and T7 RNA polymerase. The 29-13 strain (Wirtz et al., 1999) was cultured in the presence of hygromycin to maintain the tetracycline repressor protein and T7 RNA polymerase.

Stabilates were made by thoroughly mixing healthy late-log phase cells with glycerol (added to a final concentration of 10% v/v) and freezing slowly to -80°C for 24 h before transfer to liquid N₂ for long-term storage.

2.27 Transfection of DNA into trypanosomes

Plasmids with a pDex377 or pDex477 vector backbone were linearised with NotI for 4 h at 37°C. The DNA was purified by ethanol precipitation (Section 2.10) and resuspended in filter-sterilised EB buffer.

The procyclic form cells were counted using a Neubaur haemocytometer and the volume required calculated giving 3.0×10^7 cells per transfection (*i.e.* per 0.5 ml of ZMG buffer). The cells were centrifuged at 1000 x *g* for 10 min at 4°C. The supernatant was removed by pipetting and the pellet resuspended in 0.5 ml of ZMG buffer before being transferred into an electroporation cuvette (Bio-Rad 0.4 mm size), along with 5-10 µg of linearised, purified plasmid DNA. The cells were electroporated using a BTX Electro Square Porator ECM830 with 3 x 100 µs pulses at 1700 V with 200 ms intervals and then were left to recover in 10 ml of complete SDM-79 overnight. After 24 h, the cultures were selected for transformants by the addition of the appropriate drug(s).

Stabilates of all transformed cell lines were made by adding glycerol to a final concentration of 10% to a logarithmic culture ($0.5-1 \times 10^7$ cells ml⁻¹). 1 ml aliquots were frozen first to -80°C, and then stored in liquid N₂.

2.28 Trypanosome protein samples

Dense, but always healthy cultures of at least 7.0×10^6 cells ml⁻¹ were centrifuged at 1000 x *g* for 10 min at 4°C, the supernatant removed and the pellet resuspended in 10 ml of 1x PBS, centrifuged at 1000 x *g* for 10 min at 4°C, the PBS wash discarded and the PBS wash repeated. The pellet was then resuspended in an appropriate volume of boiling hot Laemmli buffer to get 4×10^5 cell equivalents ml⁻¹ and heated at 100°C for 5 min. Protein samples were stored at -80°C.

2.29 Isolation of *T. brucei* genomic DNA

10 ml cultures of early stationary phase ($1-2 \times 10^7$ cells ml^{-1}) procyclic *T. brucei* were centrifuged at $1000 \times g$ for 10 min at 4°C and the supernatant removed. Cell pellets were resuspended in 500 μl of genomic DNA (gDNA) extraction buffer along with 50 μl of 10% SDS and 10 μl Proteinase K and incubated overnight at 37°C . The following day, 500 μl of the sample was added to 500 μl of phenol:chloroform (2:1) and inverted gently for approximately 2 min to fully mix the aqueous and organic layers. The sample was then centrifuged at $6700 \times g$ in a bench-top microfuge for 2 min and the aqueous layer gently pipetted into a fresh Eppendorf tube. The sample was placed on ice and 1 ml of 100% ethanol was added to precipitate the gDNA. The gDNA was spooled using a glass Pasteur pipette and dipped in 70% ethanol 6 times to wash, and then resuspended in 50 μl EB Buffer. 5 μl of resuspension solution (with RNase added) (Fermentas) was added to degrade co-purifying rRNA and the samples incubated at 37°C for 1 h. Genomic DNA samples were then stored at 4°C .

2.30 Preparation of slides for fluorescence microscopy

1 ml of *T. brucei* culture (typically 5×10^6 cells ml^{-1}) was centrifuged at $6000 \times g$ in a bench-top microfuge for 3 min and the supernatant removed. The pellet was washed in 1x PBS and centrifuged again at $6000 \times g$ for 3 min; the pellet was then resuspended in 100 μl -500 μl of 1x PBS, depending upon the size of the pellet. The sample was pipetted onto a slide coverslip and left to settle for 5-10 min. The PBS was removed from the coverslip and *para*-formaldehyde added (to a final concentration of 2.7%) for 20 min in order to fix the cells. The coverslips were washed in ice cold methanol for 20 min, and could be stored at -20°C for future use.

Coverslips were removed from methanol and washed in 1x PBS for 10 min in order to rehydrate the cells, then incubated for 1 h in 1% BSA blocking buffer in a hydration chamber at room temperature. Any excess blocking buffer was removed and the coverslips incubated with the primary antibody for 1 h in the hydration chamber at room temperature. The coverslips were washed 3 x 5 min with 1x PBS-Tween and then incubated with the secondary antibody in the hydration chamber for 40 min. The coverslips were again washed with 1x PBS-Tween for 3 x 5 min, drained and then 1 drop of Vectashield Mounting Medium with

DAPI added (Vector Labs) before mounting the coverslip onto a slide, which was then sealed with nail varnish. Slides were stored at 4°C.

All of the immunofluorescence imaging was done using the Applied Precision DeltaVision RT deconvolution microscope and processed using the software SoftWoRx Explorer.

2.31 MitoTracker® probing

MitoTracker® labelling of mitochondria was used to confirm mitochondrial localisation of tagged proteins in procyclic *T. brucei*. A stock solution of MitoTracker® Red CMXRos (Invitrogen) was prepared by dissolving lyophilised MitoTracker® in DMSO to a final concentration of 1 mM ready for further use or storage in the dark at -20°C. The 1 mM stock solution of MitoTracker® was diluted with SDM-79 to a working concentration of 500 nM. 9 ml cultures of *T. brucei* expressing the protein of interest were set up at a 2×10^6 cells ml⁻¹, and 1 ml of the working concentration of MitoTracker® added to these cultures, to give a final working MitoTracker® concentration of 50 nM in a 10 ml culture. The cultures were then incubated at 27°C for 4 h to allow the MitoTracker® probe to passively diffuse across the plasma membrane and accumulate in the active mitochondrion of *T. brucei*. After incubation, 1 ml samples were taken and slides were made following the method laid out in Section 2.30.

2.32 Gene knockout in trypanosomes

T. brucei is diploid, so in order to generate a null mutant for any single-copy protein-coding gene, two alleles needed to be disrupted. Attempts to silence the gene expression of *Tb927.11.2920* and *Tb927.11.2910* were done by trying to replace both endogenous alleles using puromycin, phleomycin and blasticidin drug selectable markers with additional complementary homology flanks. The primers used are outlined in Table 2.6 and marked with an asterisk.

The phleomycin and blasticidin S HCl drug selectable cassettes were flanked by a tubulin processing signal at the 5' end, and an actin processing signal at the 3' end, permitting maturation of the mRNA once expressed the trypanosomes, by the addition of a spliced leader cap at the 5' end, and polyadenylation at the 3' end. The puromycin N-acetyltransferase drug resistance marker was a synthetic gene ordered from Eurofins MWG

Operon. The blasticidin and phleomycin drug resistance markers were amplified using PCR with the primer pair 30-40KOa and 30-40KOc, to give the blasticidin or phleomycin marker with short 30 bp homology flanks at the 3' and 5' ends. The PCR products were run on an agarose gel, purified and the products checked again on an agarose gel.

The PCR products from the first round of amplification were then used as templates for the second round of PCR amplification, using primer pair 30-40KOb and 30-40KOd. These primers added an extra 30 bp to the existing homology flanks onto the blasticidin or phleomycin drug selectable marker, resulting in a total of 60 bp complementary homology flanks at both the 5' and 3' end. The PCR products were run on an agarose gel, purified, confirmed on another agarose gel, ethanol precipitated and resuspended in sterile EB buffer.

The blasticidin and phleomycin drug selectable markers with the 60 bp complementary homology flanks were transfected into separate populations of procyclic 427 *T. brucei* (~3 µg of DNA) to produce two heterozygous (+/-) populations. The cell line with the blasticidin drug selectable marker was selected for successful transformants using blasticidin (10 µg ml⁻¹) (1x10⁶ cells total), and transformants from the phleomycin drug resistance marker cell line were selected for using phleomycin (3 µg ml⁻¹) (1x10⁶ cells total). Once the populations had been selected for using the appropriate drug and had recovered, stabilates were made and genomic DNA extracted (Section 2.29). The gDNA was digested with EcoRI and analysed by Southern blotting (Section 2.22) using the upstream intergenic region (UIR) hybridisation probe. The upstream intergenic probe was made using the primers UpIG0445 F (forward primer) and UpIG0445 R (reverse primer) (detailed in Table 2.3), following the protocol laid out in Amersham Gene Images AlkPhos Direct Labelling and Detection System (RPN3690).

The puromycin resistance cassette was transfected into procyclic 427 *T. brucei* to produce a heterozygous (+/-) population and the heterozygous (+/-) population resistant to blasticidin in an attempt to produce a null (-/-) mutant. The heterozygous cell line with the puromycin drug resistance selectable marker was selected for using puromycin (2 µg ml⁻¹) (1x10⁶ cells total). The candidate null (-/-) cells with the blasticidin and puromycin drug resistance selectable markers were selected for using both puromycin and blasticidin, and just puromycin (2x10⁶ cells total). Once the populations had been selected for using the appropriate drug and had recovered, stabilates were made and genomic DNA extracted. The gDNA was digested with EcoRI and analysed by Southern blotting using the upstream intergenic region (UIR) hybridisation probe.

2.33 Co-immunoprecipitation

All co-immunoprecipitation experiments were carried out using the Pierce® Co-Immunoprecipitation (Co-IP) Kit (Thermo Scientific, catalogue number 26149). *T. brucei* cells co-expressing both *TbHP30::GAFSINPAM* and *TbHP40::myc* were cultured in SDM-79 until a total cell density of 1×10^8 cells was obtained. The cells were harvested and lysed in 180 μ l of IP-Lysis/Wash buffer in the presence of an inhibitor cocktail (20 μ l of the Halt Protease Inhibitor single-use cocktail (ThermoFisher, catalogue number 78430), 2 μ l of TLCK, 2 μ l of benzamide and 2 μ l of PMSF), following the 'Lysis of Cell Suspension Cultures' protocol provided with the Pierce® Co-Immunoprecipitation (Co-IP) Kit (Thermo Scientific, 26149).

Prior to cell lysis, a column was prepared containing 10 μ g of anti-myc antibody that had been coupled to the AminoLink® Coupling Resin (Thermo Scientific, catalogue number 20381) following the 'Antibody Immobilisation' protocol in Pierce® Co-Immunoprecipitation (Co-IP) Kit (Thermo Scientific, 26149). The cell lysate was pre-cleared, following the 'Pre-clear lysate using the Control Agarose Resin' protocol in Pierce® Co-Immunoprecipitation (Co-IP) Kit (Thermo Scientific, 26149). The co-immunoprecipitation steps followed the protocol in Pierce® Co-Immunoprecipitation (Co-IP) Kit (Thermo Scientific, 26149), with protein samples prepared using the 5x Lane Marker Sample Buffer (Thermo Scientific) rather than the standard Laemmli buffer. Protein samples collected from the co-immunoprecipitation experiments were analysed by SDS-PAGE (Section 2.20), Silver Staining (Section 2.21) and Western blotting (Section 2.23).

2.34 Culturing of *T. brucei* in SILAC media

SDM-79-SILAC media was made in accordance to the original custom SDM-79 formulation, Cassion Labs, USA (see Table 2.5) and lacked the essential amino acids L-arginine and L-lysine. Isotopically labelled L-arginine and L-lysine were made up according to Table 2.6 in H₂O and filter sterilised before use.

T. brucei cell lines to be SILAC labelled (and which had previously been cultured in normal SDM-79) were diluted to a starting density of 1×10^6 cells ml⁻¹ in SDM-79-SILAC to a final volume of 10 ml and the appropriate labels (light-, medium- or heavy-) added to each flask. Parental cell lines (*T. brucei* 427 wild-type; SmOx-P9) were always cultured using light-labels.

The cells expressing *TbHP40::myc::BirA** or co-expressing both *TbHP30::GAFSINPAM* and *TbHP40::myc* were cultured separately with both medium- and heavy- labels to provide a built-in independent biological repeat for each experiment. After 48 h incubation at 27°C, the cultures were counted using a haemocytometer (Neubaur) and if the cells had undergone at least seven divisions (the minimum number of divisions required to ensure all proteins synthesised within a cell had incorporated the isotopically-labelled amino acids), the cultures were scaled up to a final volume of 40 ml in SDM-79-SILAC (plus appropriate labels) at a starting density of 1.5×10^6 cells ml⁻¹. At this point, *TbHP40::myc::BirA** (in a SmOx-P9 background) expressing cells were also incubated with doxycycline (this was not required for *TbHP30::GAFSINPAM* and *TbHP40::myc* co-expressing-cells, since expression of these constructs is constitutive in a 427 background). After 24 h incubation at 27°C, excess biotin (to a final concentration of 50 µM) was added to all cultures. After a further 24 h incubation at 27°C, cells were harvested. The 40 ml cultures were counted then centrifuged at 1000 x *g* for 10 min at 4°C and the supernatant removed. The cell pellets were washed three times in filtered 1x PBS then lysed at 1×10^9 cells ml⁻¹ in ice cold filtered H₂O in the presence of an inhibitor cocktail (20 µl of the Halt Protease Inhibitor single-use cocktail [ThermoFisher], 2 µl of TLCK, 2 µl of benzamidine and 2 µl of PMSF) and incubated at room temperature for 10 min. Lysed cells were divided into 0.5×10^9 cells ml⁻¹ (equivalent to ~5 mg/ml total protein content) aliquots and stored at -80°C. Cell pellets (containing light-, medium- and heavy-labelled cells) for BioID experiments (see Section 2.35.1) were combined prior to lysis, whereas cell pellets for co-immunoprecipitation experiments (see Section 2.35.2) were kept strictly separate from one another at the point of lysis.

Table 2.5. Details on how to make L-lysine and L-arginine deficient SDM-79-SILAC media. Note that the solution requires filtering before the addition of FCS and antibiotics.

Ingredient	Product
Hemin stock	Sigma H9039
Sodium bicarbonate	Sigma S5761
D-glucose	100x, 9.3 g/50ml
D-glucosamine	100x, 0.5 g/50ml
L-hydroxyproline	100x, 10 mg/50ml
L-proline	100x, 3.08 g/50ml
i-inositol	100x, 6.5 mg/50ml
L-methionine	100x, 0.45 g/50ml
L-serine	100x, 0.36 g/50ml
<i>Adjust to pH7.3 with NaOH, sterile filter</i>	
Heat inactivated foetal calf serum	Sigma F4135
Glutamax 1, 100x	Gibco 35050-038
Penicillin-Streptomycin, 100x	Gibco 15140-122

Table 2.6. Details on the light-, medium- and heavy-labelled amino acids that can be added to the SDM-79-SILAC media for *T. brucei* SILAC labelling experiments. Arg – Arginine; Lys – Lysine.

Label (100 mg/mL)	Molecular Weight (Da)	30% Concentration (in SDM-79-SILAC)
L-Arg.HCl (R0) <i>Light</i>	210.6	0.30 mM
L-Arg.HCl (R6- ¹³ C ₆) <i>Medium</i>	216.6	0.30 mM
L-Arg.HCl (R10- ¹³ C, ¹⁵ N ₄) <i>Heavy</i>	220.6	0.30 mM
L-Lys.HCl (K0) <i>Light</i>	182.6	0.11 mM
L-Lys.HCl (K4- ² H ₄) <i>Medium</i>	223.1	0.11 mM
L-Lys.HCl (K8- ¹³ C, ¹⁵ N ₄) <i>Heavy</i>	227.1	0.12 mM

2.35 Sample preparation for mass spectrometry

2.35.1 BioID Sample Preparation

T. brucei SmOx-P9 cells or cells expressing *TbHP40::myc::BirA** in a SmOx-P9 background that had been cultured in SDM-79-SILAC (using light-, medium- or heavy- labels) were lysed using the protocol described in Section 2.33. The lysed cells were defrosted on ice, diluted with

MS-Solubilisation buffer at a 1:1 ratio and the samples vortexed until they went clear. The samples were centrifuged at $6700 \times g$ for 10 mins at 4°C , and the supernatant transferred to a LoBind Eppendorf (Sigma-Aldrich). Streptavidin-coated magnetic beads (Pierce™) (100 μl per column) were washed 3 times with filtered 1x PBS in LoBind Eppendorfs on a magnetic rack. The washed magnetic beads were then transferred into the sample supernatant and incubated on a rotating wheel at 4°C for 4 h. This allowed biotinylated molecules to adhere to the streptavidin immobilised on the magnetic beads. After incubation, the LoBind Eppendorfs were placed on the magnetic rack and the supernatant discarded. The beads were resuspended in 1 ml of MS-IAA solution, which contained iodoacetamide for the reductive alkylation of any free cysteine residues. The LoBind Eppendorfs were wrapped in tin foil (since iodoacetamide is light-sensitive) and then incubated on a rotating wheel at room temperature for 20 min. The LoBind Eppendorfs were again placed on the magnetic rack and the supernatant removed. The beads were washed with 3 x 1ml MS-Solubilisation buffer followed by 5 x 1ml washes with filtered 1x PBS. The beads were then resuspended in MS-ABC solution in preparation for the tryptic digestion in order to remove the biotinylated peptides from the streptavidin coated beads.

Digestion was carried out following the protocol provided with the Trypsin Singles, Proteomics Grade, Enzyme (Sigma Aldrich, T7200). The tryptic digest samples were sealed in LoBind Eppendorfs with parafilm and incubated at 37°C for 18 h.

After digestion, the LoBind Eppendorfs were placed in the magnetic rack and the supernatant containing the digested peptides was transferred into a fresh LoBind Eppendorf. The magnetic beads were washed with 100 μl MS-ABC solution and the supernatant transferred to the LoBind Eppendorf containing the digested peptides. The magnetic beads were then washed with 100 μl of 500 mM NaCl and the supernatant again transferred into the LoBind Eppendorf containing the digested peptides. A C18 microspin column (Harvard Apparatus) was washed with 100 μl of methanol and centrifuged at $200 \times g$ for 1 min and the flow-through discarded. The microspin column was then washed 3 x 100 μl with MS-Elute solution and centrifuged each time at $200 \times g$ for 1 min and the flow-through discarded. The column was then washed 3 x 100 μl with MS-TFA solution and centrifuged each time at $200 \times g$ for 1 min and the flow-through discarded. Trifluoroacetic acid (TFA) was added to the LoBind Eppendorf containing the digested peptides to get a 0.1% TFA final concentration. The peptides were then transferred to the microspin column for desalting. The microspin column was centrifuged at $200 \times g$ for 1 min and the flow-through discarded. The microspin column was washed 3 x 100 μl with MS-TFA and centrifuged each time at $200 \times g$ for 1 min

and the flow-through discarded. The peptides were eluted from the column into a fresh LoBind Eppendorf tube with 3 x 100 µl MS-Elute solution, which was centrifuged each time at 200 x *g* for 1 min. The eluted peptides were diluted with 500 µl of autoclaved H₂O and stored overnight at -80°C. The frozen samples were then lyophilised in preparation for analysis by mass spectrometry.

2.35.2 Co-Immunoprecipitation Sample Preparation

T. brucei 427 or cells co-expressing both *TbHP30::GAFSINPAM* and *TbHP40::myc* in a 427 background that had been cultured in SDM-79-SILAC (using light-, medium- or heavy- labels) were lysed using the protocol described in Section 2.33. The three separate lysed cell aliquots (labelled with light-, medium- or heavy- lysine and arginine isotopes) were thawed on ice and pre-cleared at 4°C, following the 'Pre-clear lysate using the Control Agarose Resin' protocol in Pierce® Co-Immunoprecipitation (Co-IP) Kit (Thermo Scientific, catalogue number 26149).

Prior to lysis, three separate columns were prepared, each containing 20 µg of anti-myc antibody that had been coupled to AminoLink® Coupling Resin (Thermo Scientific) following the 'Antibody Immobilisation' protocol in Pierce® Co-Immunoprecipitation (Co-IP) Kit (Thermo Scientific, 26149). The cell lysates were added to the myc-immobilized column (light-labelled in one, medium-labelled in another and heavy-labelled in the last) following the 'Co-IP' protocol in Pierce® Co-Immunoprecipitation (Co-IP) Kit (Thermo Scientific, 26149) and incubated for 2 h on a rotating wheel at 4°C. The flow-throughs were removed and stored at -80°C and the columns washed three times using the IP-Lysis/Wash buffer (Thermo Scientific). The beads from each of the three columns were resuspended in IP-Lysis/Wash buffer, transferred to separate LoBind Eppendorfs, centrifuged at 200 x *g* and the supernatant discarded. The beads were then boiled in 30 µl of Laemmli buffer at 95°C for 5 min. The samples were centrifuged at 200 x *g*, the supernatant transferred to a fresh LoBind Eppendorf and the boiling process repeated. The protein samples were run on a pre-cast 12% SDS-PAGE gel (BioRad) for 10 min at 200 V. The SDS-PAGE gel was stained with Instant Blue (Expedeon) and destained using H₂O.

The protein bands in the destained SDS-PAGE gel were excised by cutting into 1x1 mm squares using a scalpel. At this point the light-, medium- and heavy- labelled protein bands were mixed with one another. The gel pieces were transferred to a LoBind Eppendorf and

the gel completely destained using ammonium bicarbonate (100 mM) and 100% acetonitrile at a 1:1 (v/v) ratio for 30 min at room temperature with occasional vortexing, until the gel pieces were completely clear. The supernatant was discarded, 750 μ l of 100% acetonitrile added and the sample incubated at room temperature with occasional vortexing for approximately 10 min until the gel pieces shrunk and turned white in colour. The supernatant was discarded and enough ammonium bicarbonate (100 mM, with 10 mM DTT added) was added to completely cover the gel pieces. The sample was incubated for 30 min at 56°C then the supernatant discarded. The sample was allowed to cool to room temperature then 500 μ l of 100% acetonitrile was added, incubated for 10 min at room temperature and the supernatant discarded. Enough MS-IG-IAA solution was added to completely cover the gel pieces and the sample incubated in the dark at room temperature for 20 min. The supernatant was then discarded and 500 ml of 100% acetonitrile added, incubated for 10 min at room temperature then the supernatant discarded.

The proteins within the gel pieces were then digested with trypsin. Tryptic digestion was carried out using Trypsin Singles, Proteomics Grade, Enzyme (Sigma Aldrich, T7200). Enough trypsin buffer was added to completely cover the gel pieces, the volume added was recorded and the samples incubated at 4°C for 30 min. After the incubation, more trypsin buffer was added if necessary (if the gel pieces absorbed any of the trypsin buffer, more was added to ensure that the gel pieces were completely submerged in trypsin buffer) and the volume recorded, and the samples incubated for a further 90 min at 4°C. After the incubation, ammonium bicarbonate (10 mM, containing 10% [v/v] acetonitrile) was added to the digest and the final volume recorded. The tryptic digest samples were sealed with parafilm and incubated at 37°C for 18 h.

After the tryptic digestion, the digested peptides were extracted from the gel pieces using the MS-IG-Extraction buffer. Enough MS-IG-Extraction buffer was added so that there was approximately a 1:2 ratio between the digest volume and the MS-IG-Extraction buffer. The samples were then incubated on an orbital shaker at 37°C for 15 min and the supernatant transferred to a fresh LoBind Eppendorf. The peptides were then desalted using a C18 microspin column (as described in Section 2.35.1) and the eluted peptides diluted with 500 μ l of autoclaved H₂O and stored overnight at -80°C. Frozen samples were then lyophilised in preparation for analysis by mass spectrometry.

2.36 Liquid chromatography-mass spectrometry

All lyophilised samples were sent to the FingerPrints Proteomics Facility at the College of Life Sciences, University of Dundee, for analysis by liquid chromatography tandem mass spectrometry (LC-MS/MS).

2.37 Analysis of mass spectrometry data

The data obtained from the LC-MS/MS was analysed based on the methods and parameters described by Urbaniak et al., (2012). The data was processed using MaxQuant (Cox and Mann, 2008) version 1.5.2.8 which uses the Andromeda search engine (Cox et al., 2011). Proteins identified by LC-MS/MS were matched against the *T. brucei brucei* 927 annotated protein sequence database (Release 26, downloaded from TriTrypDB (Aslett et al., (2010); <http://www.tritryp.org>) and a list of known frequently observed contaminants (e.g. trypsin, human keratins). Search parameters specified a MS tolerance of 5 ppm, a MS/MS tolerance at 0.5 Da and full trypsin specificity, allowing for up to three missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionines, N-terminal protein acetylation and N-pyroglytamate were allowed as variable modifications. Peptides were required to be at least 6 amino acids in length, and false discovery rates of 0.01 were calculated at the levels of peptides, proteins and modification sites based on the number of hits against the reversed sequence database. SILAC ratios were calculated using only peptides that could be uniquely mapped to a given protein.

2.38 Analysis of protist FeFe-hydrogenases

The full length sequences of the FeFe-hydrogenases were analysed using InterPro: Protein Sequence Analysis & Classification (Mitchell et al., 2015) (www.ebi.ac.uk/interpro/), which predicted protein domains and functional sites, and provided information on the protein superfamily, family and sub-family levels. The query sequence was entered in FASTA format and submitted using the default settings. The protein architecture predictions made by InterPro were cross-referenced against SMART (Simple Modular Architecture Research Tool, http://smart.embl-heidelberg.de/smart/set_mode.cgi) (Letunic et al., 2015; Schultz et al., 1998) which predicted protein architecture and Pfam domains, signal peptides, internal repeats, intrinsic protein disorder, outlier homologues and homologues with known

structure. The query sequence was entered in FASTA format and all of the aforementioned search fields selected before submission. Cartoon structures of the FeFe-hydrogenases, illustrating the domain architecture of each, were then created using Microsoft PowerPoint 2010.

2.39 Analysis of microarray data

A microarray dataset complete with three biological replicates (Fritz-Laylin and Cande, 2010) showing RNA levels during *Naegleria gruberi* differentiation was downloaded from the NCBI Gene Expression Omnibus (Edgar et al., 2002), GEO Series accession number GSE21527. Raw data was imported into Microsoft Excel 2010 according to the time point (0 to 80 min in 20 min increments) and the repeat number (total of three repeats present in the dataset), arranged into columns and labelled (see Figure 2.1 A).

Genes of interest could then be identified within the MS Excel spreadsheet according to their JGI accession number (found in the JGI Genome Portal in the published *Naegleria gruberi* genome (Fritz-Laylin et al., 2010b) e.g. for NirK it is JGItrs50708_1). Utilising the find option under 'look in' of MS Excel, 'values' must be selected from the dropdown menu in order to find the gene of interest via its accession number. The data for both blocks 1 and 2 for specific genes were selected from columns A to R, copied, and pasted in a new spreadsheet (Figure 2.1 A).

Subsequently, an average of both blocks had to be calculated for each time point and each repeat. For example NirK (see Figure 2.1 B) using Time0_Rep1 (in column E), the formula in cell E6 " $=\text{SUM}(E2,E3)/2$ ". This formula was reproduced along row 6 to calculate the average for both blocks across all time points and repeats. Next, an overall average for each time point had to be calculated from the 3 biological replicates. For example NirK (see Figure 2.1 B) the 3 replicates for 'Time0_Rep1' are found in columns E, F and G, with their block averages calculated in row 6 (described above); the formula to calculate the overall average in cell E10 being " $=\text{SUM}(E6,F6,G6)/3$ ". This formula was used to calculate the overall average for each time point, 0, 20, 40, 60 and 80. Error bars showing the variability across the repeats next had to be calculated. The maximum value in the data for e.g. NirK 'Time0' in cell E12 uses the formula " $=\text{MAXA}(E2,F2,G2,E3,F3,G3)$ ". The maximum error bar is then calculated by subtracting the average from the maximum value e.g. NirK 'Time0' in cell E13 using the formula " $=\text{SUM}(E12-E10)$ ". These formulas were then used for each time point, 0,

20, 40, 60 and 80. The minimum value in the data for *e.g.* NirK 'Time0' in cell E14 uses the formula "`=MINA(E2,F2,G2,E3,F3,G3)`" and the minimum error bar in cell E15 using the formula "`=SUM(E10-E14)`". Finally a bar graph was produced using the clustered column chart type and the overall average data selected from the spreadsheet *e.g.* for NirK, cells E10 to I10. Vertical error bars were added using the 'Custom' option and specifying the values *e.g.* for NirK, cells E13 to I13 for the maximum bars and cells E15 to I15 for minimum bars.

2.40 Assembling an Access database for analysis of TbHP40 and TbHP30 evolution

The full length sequences of *Tb927.11.2920* (*TbHP40*) and *Tb927.11.2910* (*TbHP30*) were subjected to protein BLAST (basic local alignment search tool) searches using the non-redundant protein sequences database and the blastp (protein-protein BLAST) algorithm in order to identify candidate homologues of both proteins. A wide variety of protists and fungi were searched for the presence of likely homologues and orthologues. Specific organism searches were recorded in a Microsoft Access 2010 database (see Supplementary Data provided on CD1). Information of any candidate orthologues/homologues identified by the BLAST searches was also stored in this database. This information included details on whether it was a homologue to *TbHP40* or *TbHP30*, the organism it was found in, the kingdom, phylum, class and Tax ID of the organism, the e-value, the length of the protein in amino acids, its accession number and the amino acid sequence. Proteins identified as candidate orthologues/homologues were then subjected to reciprocal protein BLAST searches to ensure the top hit of the search was either *TbHP40* or *TbHP30*. Queries against the database could then be generated to sort through the data collected as required. These queries could be exported into Microsoft Excel 2010 and Notepad as needed.

Sequences were aligned using MUSCLE (Multiple Sequence Comparison by Log-Expectation, www.ebi.ac.uk/Tools/msa/muscle) (Edgar, 2004a, b). The query sequences were entered in FASTA format and the default settings used, with the output format being ClustalW. The output alignment was copied and pasted into Microsoft Word 2010.

Multiple sequence alignments were shaded using Boxshade (version 3.21, http://www.ch.embnet.org/software/BOX_form.html). The output alignments from MUSCLE were copied and pasted into Boxshade and the following options selected for all alignments: 'output format – postscript landscape'; 'font size – 8'; 'consensus line – no consensus line'; 'fraction of sequences (that must agree for shading) – 1.0'; 'input sequence format – ALN'.

'Run BOXSHADE' was then clicked and the output saved as Adobe PDF files. Any further formatting was done using Adobe Photoshop CS3.

2.41 Protein targeting predictions

Protein localisations were predicted using a variety of online proteomics tools:

Mitoprot – This program only predicted N-terminal mitochondrial targeting sequences and potential cleavage sites (Claros and Vincens, 1996). Query sequences were entered in FASTA format then submitted.

Predotar 1.03 – This program predicted N-terminal targeting sequences to the mitochondria and plastids (Small et al., 2004). Query sequences were entered in FASTA format either 'plant sequences (possibility of plastid targeting)' or 'animal or fungal sequences (no possibility of plastid targeting)' were selected and the query submitted.

TargetP 1.1 – This program predicted the subcellular localisation of eukaryotic proteins (Emanuelsson et al., 2000). Query sequences were entered in FASTA format and 'plant or non-plant' selected under the 'organism group' option, depending upon the origin of the sequence. The default 'no cutoffs: winner-takes-all' option was used and the sequence submitted.

WoLF PSORT – This program predicted subcellular localisation for animal, plant and fungal sequences (Horton et al., 2007). Query sequences were entered in FASTA format, either 'animal, plant or fungi' selected and then submitted.

2.42 Homology modelling

The homology modelling for *Tb*927.11.2920 (*Tb*HP40) was carried out following a protocol detailed in Szpara et al., (2014), using MOE software.

The following solved structures from the histidine phosphatase protein families PDB were identified as the closest structural matches to *Tb*HP40 and selected for use in the homology modelling of *Tb*HP40: Chain A of TIGAR from *Danio rerio* (PDB code 3E9D.A) (Li and Jogl, 2009); PGAM5 from *Homo sapiens* (PDB code 3MXO) (Unpublished) and PhoE from *Bacillus stearothermophilus* (PDB code 1H2E) (Rigden et al., 2002). Homology models for

Tb927.11.2920 (TbHP40) were constructed against each of the selected solved structures using MOE 2014.09 (Chemical Computing Group, Montreal), as follows. First, an initial proposed partial geometry was copied from the template chains in the solved structures, using all co-ordinates where residue identity was conserved. Otherwise, only backbone coordinates were used. Based on this initial partial geometry, Boltzmann-weighted randomized modelling (Levitt, 1992) was employed with segment searching for regions that could not be mapped onto the initial partial geometry (Fechteler et al., 1995). 1000 models were constructed. On completion of segment addition, each model was energetically minimized in the AMBER99 force field (Wang et al., 2000b). The highest scoring intermediate model was then determined by the Generalized Born/Volume Integral (GB/VI) methodology (Labute, 2008). Molecular surfaces were created using the method of Connolly (Connolly, 1983), as applied within MOE. Attempts were made to model *Tb927.11.2910 (TbHP30)*, but no homologous solved structures could be identified, therefore modelling using this method was not possible. Attempts were also made to model *TbHP40* homologues from *Leishmania major* (XP_001685971), *Bodo saltans* (BS92235), *Naegleria gruberi* (XP_002676814), *Acanthamoeba castellanii* (ELR24091) and *Arabidopsis thaliana* (NP_172369), but no viable models were produced.

Figure 2.1

A.

J	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	JCI	AccessionBlock	ID	Time0_Rep1	Time0_Rep2	Time0_Rep3	Time20_Rep1	Time20_Rep2	Time20_Rep3	Time40_Rep1	Time40_Rep2	Time40_Rep3	Time60_Rep1	Time60_Rep2	Time60_Rep3	Time80_Rep1	Time80_Rep2	Time80_Rep3
1	JG1rs1008_1	BLOCK1	JG1DB:Naegr1_1008	4723.2677	4058.5566	3225.1686	3047.9225	3741.5345	3749.5631	3530.2612	3474.9914	3380.1498	1731.0889	2743.0819	1935.4659	1887.8115	1903.5999	2273.2471
2	JG1rs1008_2	BLOCK2	JG1DB:Naegr1_1008	4215.9698	4371.7128	3067.7817	2973.6264	3789.4141	4118.2614	3506.1632	2995.9867	2998.2314	1815.002	2351.8082	2041.069	1698.5552	1821.281	1839.2861
3	JG1rs103_1	BLOCK1	JG1DB:Naegr1_103	1175.5473	1172.701	1376.2732	476.4231	816.5057	652.6866	197.9744	117.0188	136.1929	36.9043	34.3292	71.2672	77.0745	68.5208	149.6381
4	JG1rs103_2	BLOCK2	JG1DB:Naegr1_103	1065.2248	1272.5983	1314.6877	531.0499	759.4988	609.5048	117.6789	154.3169	151.7703	24.5781	35.1648	71.7897	76.7588	72.9141	157.4681
5	JG1rs103_1	BLOCK1	JG1DB:Naegr1_103	485.4874	675.7876	1050.3331	385.9228	1298.6254	1342.0702	664.4666	1329.6695	1409.6432	742.6894	1110.6156	1278.8687	485.4466	636.1737	736.0771
6	JG1rs103_2	BLOCK2	JG1DB:Naegr1_103	443.564	682.2016	1060.7412	357.4677	1372.5863	1378.5333	2067.713	1460.7869	1460.7869	782.7047	1188.0436	1177.2262	484.4441	604.8169	754.7201
7	JG1rs1063_1	BLOCK1	JG1DB:Naegr1_1063	4898.2092	2815.7715	2688.0862	3570.7527	1735.8139	1284.2519	2067.713	691.8733	422.1429	362.0637	348.4345	274.9756	491.0444	594.4647	726.5811
8	JG1rs1063_2	BLOCK2	JG1DB:Naegr1_1063	4765.6131	2958.8497	2791.674	3639.9904	1729.2902	1296.0718	1979.3464	744.3722	524.4109	381.6536	321.5643	280.7272	495.4052	704.6882	706.4481

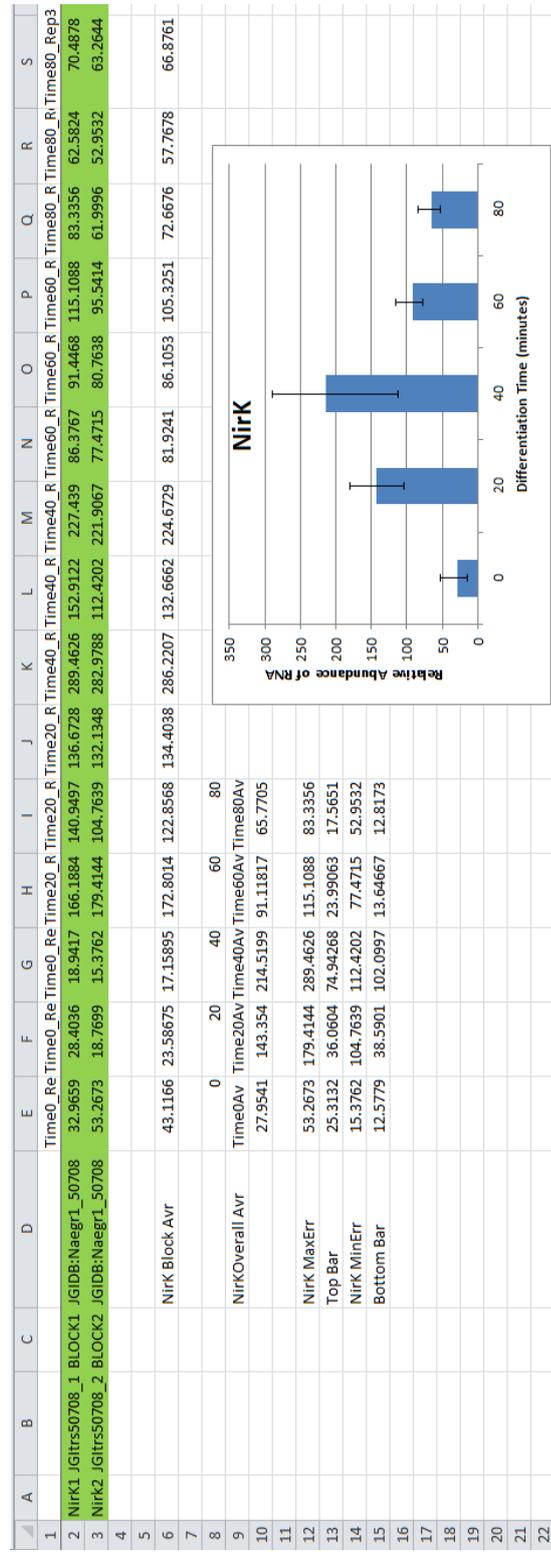


Figure 2.1. Print screens from the Excel spreadsheet used to produce the microarray mRNA expression level graphs.

- A. Print screen of the imported raw data from the *Naegleria gruberi* microarray, showing the arrangement of the time points, repeats, blocks and JCI accession number.
- B. Print screen of the working data for the *Naegleria gruberi* NirK gene, showing how the averages and error bars were calculated in order to produce the microarray graph.

Chapter 3

Bioinformatics analysis of novel, paralogous histidine phosphatases

A theme of mitochondrial metabolism is recurrent throughout this thesis. In this chapter, I begin my characterisation of a paralogous pair of histidine phosphatases identified by a previous student (Brown, 2011), each of which contains a putative mitochondrial leader sequence. These histidine phosphatases were not subject to extensive analysis by this or any previous students in the group. Traditionally, adaptation to parasitism is linked to metabolic stream-lining (Ginger, 2006). Thus, it could be speculated that these putative histidine phosphatases retained by trypanosomatid parasites serve important functions at some point during the life cycle. To begin my investigation, I started with a bioinformatic analysis of *Tb927.11.2920* (henceforth referred to as *TbHP40*) and *Tb927.11.2910* (henceforth referred to as *TbHP30*) and also looked at where they localised within procyclic *T. brucei*.

As I will show below (Section 3.1-3.2), the distribution of these histidine phosphatases is patchy across eukaryotes for which genome sequences are available. But one paralogue, *TbHP30*, has lost its active site in all of the trypanosomatid orthologues. *TbHP40* and *TbHP30* are neighbouring genes, located on chromosome 11 of *T. brucei*, and may have arisen due to a tandem duplication event. Both contain a PGAM domain (clade 1 - IPR013078), which denotes them as members of the histidine phosphatase superfamily. Clade 1 histidine phosphatases function due to a conserved essential histidine residue that forms part of a 'RHG' (arginine-histidine-glycine) catalytic triad. *TbHP40* retains this 'RHG' triad, whereas the essential histidine residue in *TbHP30* has degenerated into an asparagine - 'RNG'. This suggests that *TbHP30* has lost the ability to function catalytically.

3.1 An overview of the histidine phosphatase superfamily

The eighth enzyme of the glycolysis reaction pathway is cofactor-dependent phosphoglycerate mutase (dPGM), which catalyses the interconversion of 2-phosphoglycerate and 3-phosphoglycerate (Rigden, 2008), and is the archetypal member of the histidine phosphatase superfamily. Most members of this superfamily display phosphatase activity rather than the mutase activity seen in dPGM, making dPGM an anomaly within the family. But due to dPGM being one of the best studied members it often leads to mis-annotation of other histidine phosphatases as mutases (Rigden, 2008). Another

atypical histidine phosphatase is Sts-1 (suppressor of T-cell receptor signalling 1), which despite having all the characteristics of a *bona fide* histidine phosphatase (both structurally and biochemically, in terms of the presence of a PGAM domain and within it a functional histidine residue), can exhibit tyrosine phosphatase activity as a negative regulator of T-cell antigen receptor signalling pathways (Sadatomi et al., 2013).

Histidine phosphatases can be divided into two branches. The functions of the enzymes found in branch 1 are extremely diverse, and include dPGM, TIGAR, Sts-1 and PhoE. Branch 2 consists of acid phosphatases such as phytase (Jedrzejewski, 2000; Rigden, 2008). When comparing the amino acid sequences between branch 1 and 2 enzymes, there is very little conservation across the two branches. Histidine phosphatases from both branches have a wide variety of functions, for example some are involved in intermediary metabolism or metabolic regulation, signalling cascades and regulatory roles, whereas others function in extracellular metabolism, notably the scavenging of phosphate molecules (Rigden, 2008). TIGAR (tumour protein 53-induced glycolysis and apoptosis regulator) is thought to have a role in controlling the levels of fructose-2,6-bisphosphate within a cell, in turn contributing to the regulation of the rates of glycolysis and gluconeogenesis and the generation of apoptosis-inducing reactive oxygen species (Bensaad et al., 2006; Li and Jogl, 2009). Although a recent study showed that human TIGAR would also take 2,3-bisphosphoglycerate, 2-phosphoglycerate and phosphoenolpyruvate as potential substrates (Gerin et al., 2014).

Despite the striking differences in sequence homology, function and substrate specificity, all members of the superfamily contain an arginine-histidine-glycine (RHG) motif at the N-terminus, the histidine residue of which is phosphorylated and dephosphorylated during catalysis (Rigden, 2008). During the reaction mechanism, the phospho- group from the substrate is transferred to the essential histidine residue with the assistance of surrounding residues. These include well conserved flanking arginine, glycine and histidine residues which stabilise the transfer by hydrogen bonding (Wang et al., 2006) and form the 'phosphate pocket' (Figure 3.1) (Rigden, 2008). The N- and C- terminal regions tend to differ between various histidine phosphatases with different roles in different species, but the catalytic core structures are usually quite similar across both branch 1 and 2 histidine phosphatases, for which known solved structures are available (Figure 3.2).

The rare family members that display mutase activity, for example dPGM, differ from the reaction mechanism shown in Figure 3.1 because their essential histidine is already 'primed'

with a phospho- group ready for transfer to either the 2- or 3- position of 2-phosphoglycerate or 3-phosphoglycerate. The intermediate 2,3-phosphoglycerate form re-orientates within the catalytic site before one of the phospho- groups is removed, the substrate released and the 'primed' mutase enzyme is regenerated (Rigden, 2008).

Enzyme specificity can differ massively depending upon the enzyme in question. Some histidine phosphatases take very small substrates, for example phosphoglycerate, and others are specific for large phosphorylated proteins such as the histidine-containing phospho-transfer domain of ArcB (an *Escherichia coli* sensory kinase) (Ogino et al., 1998; Rigden, 2008). Modelling of the catalytic site of a histidine phosphatase enzyme can provide clues as to its preferred substrate and therefore predictions can be made as to its potential physiological function within a cell.

The trypanosomatid parasites have an unusually large array of novel or uncharacterised histidine phosphatases, compared to those found in other unicellular organisms such as *Giardia lamblia* (no readily recognisable homologues), *Naegleria gruberi* (one homologue), *Plasmodium falciparum* (one homologue) and *Entamoeba histolytica* (two homologues) (Brown, 2011). All four putative *T. brucei* histidine phosphatase homologues are annotated as being putative phosphoglycerate mutases in GeneDB. One shares homology with a broad specificity phosphatase found in *Bacillus stearothermophilus* named PhoE (Brown, 2011); another shares homology with the biochemically characterised phosphoglycerate 5 (PGAM5) found in humans (Brown, 2011), although in humans it shows no phosphoglycerate mutase activity and has been shown to have mitochondrial function (Takeda et al., 2009). It has also been shown that PGAM5 is another atypical branch 1 histidine phosphatase, which can display serine/threonine protein phosphatase activity, and is involved in the activation of apoptosis signal-regulating kinase 1 (Sadatomi et al., 2013; Takeda et al., 2009). The final two putative histidine phosphatases in *T. brucei* are annotated as being 'glycerolphosphate mutases' and none of the homologues identified in other organisms, which were found using NCBI BLAST, have been biochemically characterised.

3.2 *TbHP40* and *TbHP30* are mitochondrial proteins

A Clustal sequence alignment of *TbHP40* and *TbHP30* using MUSCLE shows that these proteins only share 35% identity. *TbHP30* is 91 amino acids longer than *TbHP40* (383 amino acids vs. 291 amino acids respectively), principally because of a long N-terminal extension (Figure 3.3 C).

TbHP40 and *TbHP30* have been detected previously in mitochondrial proteomes (Acestor et al., 2009; Fisk et al., 2013; Niemann et al., 2013; Panigrahi et al., 2009; Urbaniak et al., 2013) (Figure 3.3 A-B), but contamination of organellar proteomes with proteins from other cellular sites is not unusual. Thus, before subjecting *TbHP40* and *TbHP30* to more extensive bioinformatic analysis, I first confirmed the candidature of mitochondrial localisation using C-terminal epitope tags.

Previously, the coding sequences of *TbHP40* and *TbHP30* were fused in-frame at the C-terminus with three tandemly-repeated myc epitopes (henceforth known as *TbHP40::myc₃* and *TbHP30::myc₃*). The genes encoding these myc-tagged proteins were cloned into the pDex377 expression vector (Kelly et al., 2007) in place of the luciferase reporter gene. These plasmid constructs were made with a previous undergraduate student (Steven Barlow). The plasmids were linearised with NotI and transfected into procyclic 427 *T. brucei* for homologous recombination into a minichromosomal locus and constitutive expression. Once stable transformants had been selected, protein samples were harvested for Western blot analysis and slides made for immunofluorescence imaging.

The protein samples for both *TbHP40::myc₃* and *TbHP30::myc₃* were separated using SDS-PAGE, blotted onto an Amersham Hybond-P membrane and the membrane was probed with anti-myc antibody. Protein was detected in both transformed populations (Figure 3.4) with a distinct band at ~34 kDa in lane 1 that corresponds with the predicted size of *TbHP40*, and another distinct band was detected at ~43 kDa in lane 2 that corresponds with the predicted size of *TbHP30*. The smaller bands detected in the *TbHP30* sample are likely to be the result of proteolytic degradation; a significant amount of protein was expressed within cells. Cells were incubated with MitoTracker Red, settled onto coverslips, fixed with *para*-formaldehyde and decorated with anti-myc antibody. This showed mitochondrial localisations for both *TbHP40::myc₃* and *TbHP30::myc₃* (Figure 3.4).

Having provided confirmation of mitochondrial localisation of both *TbHP40* and *TbHP30*, which is also supported by presence of peptides in mitochondrial proteomes for both

bloodstream and procyclic life stages of *T. brucei*, and indicates constitutive expression of these proteins (Fisk et al., 2013; Niemann et al., 2013; Panigrahi et al., 2009; Urbaniak et al., 2012; Urbaniak et al., 2013), I attempted to address the following points using bioinformatic approaches:

- A. To explore the phylogenomics of these novel histidine phosphatases in concordance with current views on eukaryotic phylogeny.
- B. To understand how the paralogy and degeneration of the active site of *TbHP30* might have evolved.
- C. Modelling experiments to get an understanding of what sort of substrate these proteins might act upon; whether they be other proteins or small metabolites.

All these analyses were designed to gain insight into what the physiological and cellular function of *TbHP40* and *TbHP30* may be. Results from additional experimental studies are described in Chapters 4 and 5.

3.3 Phylogenomic distribution of *TbHP40* and *TbHP30* orthologues

TbHP40 is highly conserved across the trypanosomatids and orthologues were found in all of the trypanosomatids whose genome sequence is accessible at TritrypDB. Very little difference was observed between *TbHP40* orthologues identified in *Trypanosoma* species when the sequences were aligned using MUSCLE (Figure 3.5). The *T. brucei* and *T. evansi* genes are identical in sequence, and in the other *Trypanosoma* species, there was slight variation at the N- and C- terminus, and a small region in the middle of the protein sequence that differed from one-another.

Orthologues to *TbHP40* were also found in all of the *Leishmania* sequences analysed. Many of the *Leishmania* orthologues were almost identical to each other (*L. major* and *L. infantum* share 99.6% sequence identity) and there is less divergence around the N- and C- terminus, compared to the orthologues found in the *Trypanosoma* species such as *T. cruzi* (Figure 3.5). However, when the *Leishmania* sequences were aligned with *TbHP40* and other *Trypanosoma* orthologues, it was obvious that *Leishmania* sequences were longer, with an extended C-terminus and a large insertion in the PGAM domain (Figure 3.5). Looking within the *Leishmania* species, more amino acid differences were present between species within the insertion and the *Leishmania*-specific C-terminal extension than within the PGAM domain itself. This suggested these regions of *Leishmania* HP40 proteins were under less

selective pressure than the PGAM domain. Thus, HP40 proteins in *Leishmania* species conform to the general consensus that *Leishmania* proteins tend to be ~30% longer than orthologues found *T. brucei* or *T. cruzi* (El-Sayed et al., 2005).

Many other kinetoplastids were also found to contain orthologues to *TbHP40*, including the plant pathogen *Phytomonas* sp., the insect pathogens *Crithidia fasciculata*, and *Leptomonas seymouri*, and *Endotrypanum monterogeii*, which is a close relative of *Leishmania* and a parasite of sloths. When these sequences were aligned with *Leishmania* and *TbHP40*, again the two distinct blocks of conservation were apparent (Figure 3.5). All of the orthologues to *TbHP40* were longer than the *Trypanosoma* sequences, and all resembled the *L. major* sequence, with a long divergent C-terminus and internal insertion between the two regions of conservation. The free-living kinetoplastid, *Bodo saltans*, is the closest relative to the trypanosomatid family for which a nuclear genome sequence is available (Jackson et al., 2015). It also contained an orthologue of *TbHP40* and its architecture more closely resembled the *Trypanosoma* HP40 proteins than it did *Leishmania* HP40 proteins – *i.e.* the *B. saltans* HP40 lacked a *Leishmania*-style insert within the PGAM domain and had only a short C-terminal sequence extending beyond the PGAM domain (Figure 3.5).

Like *TbHP40*, *TbHP30* is also highly conserved within the trypanosomatid family; *TbHP30* orthologues can be found in all trypanosomatids, for which genome sequence is available. In sequence alignments *e.g.* Figure 3.6, trypanosomatid HP30 proteins principally differ only with respect to the amino acid sequence N-terminal to the PGAM domain and *Leishmania*, *Crithidia* and *Phytomonas* orthologues contain an insertion within the PGAM domain, which is not present in *Bodo* or *Trypanosoma* proteins. The degeneracy of the catalytic site is evident in all *TbHP30* orthologues, but the remainder of the PGAM domain is otherwise well conserved.

To look at the phylogenomic distribution of *TbHP40* and *TbHP30*, the amino acid sequences of both proteins were used as query sequences to search draft or complete genome sequences for a range of eukaryotes spanning all known eukaryotic super-groups, as shown in Figure 3.7. Homologous or ‘candidate orthologous’ proteins were found in only a few eukaryotes. In reciprocal BLAST analyses, *TbHP40* was always recovered as the top BLAST hit, with *TbHP30* orthologues only within kinetoplastid organisms. Even in *Euglena gracilis*, which together with the kinetoplastids form the Euglenozoa, only a candidate orthologue to *TbHP40* was found. From the heterolobosean group that is sister to the Euglenozoa; two proteins that identify *TbHP40* as the top reciprocating hit were found. All of the candidate

TbHP40 orthologues or homologues shown in Figure 3.7 contain the canonical catalytic 'RHG' triad, characteristic of all histidine phosphatases. An alignment of candidate HP40 proteins found in *T. brucei*, *B. saltans*, *E. gracilis* and *N. gruberi* are shown in Figure 3.8. Unfortunately, the *Naegleria* protein that more closely resembles *TbHP40* lacks a predicted start codon. However, many *Naegleria* protein-coding genes contain introns and gene model predictions for this organism are not always robust (Fritz-Laylin et al., 2010b).

Across the rest of the eukaryotic tree, the distribution of homologues to *TbHP40* is patchy (Figure 3.7). Amongst the Amorpha (formerly known as the Unikonts (Adl et al., 2012; He et al., 2014)) the distribution of homologues to *TbHP40* is particularly sparse. Only one homologue was found in the Holozoa, which belonged to *Capsaspora owczarzaki*. The distribution of homologues in fungi is also sparse, with only *Fusarium oxysporum* and *Trichophyton rubrum* (both of which belong to the phylum Ascomycota) containing homologous HP40 sequences. No homologues containing the 'RHG' catalytic triad were found in other Ascomycota fungi such as *Candida albicans* and *Aspergillus niger* and only two sequences, both with minimal homology to *TbHP40*, were found in *Saccharomyces cerevisiae*. No homologues to *TbHP40* were identified in any members of the Basidiomycota phylum analysed (52 species sampled, see Supplementary CD 1, represented by *Serpula lacrymans*, *Ustilago maydis* and *Puccini* sp. in Figure 3.7) nor any of members of the Microsporidia phylum analysed (22 species sampled, see Supplementary CD 1, represented by *Encephalitozoon cuniculi* in Figure 3.7). No homologues to *TbHP40* were found in the representative member of the Apusozoa, *Thecamonas trahens*, which is viewed as one of the closest neighbours to the Opisthokonts and evolutionarily speaking a divergent, enigmatic protist (Cavalier-Smith and Chao, 2010; Kim et al., 2006). Several sequences with limited homology to *TbHP40* that contained the 'RHG' catalytic triad were identified in the Amoebozoa, although the quality of the amino acid sequence for the HP40 homologue in *Polysphondylium pallidum* was questionable. The *Acanthamoeba castellanii* homologue had a 38 residue N-terminal extension but aligned well with *TbHP40* around the active site region and then sporadically along the rest of the length of sequence (data not shown).

Compared to the Amorpha, it appeared that homologues to *TbHP40* occurred more frequently amongst the land plants, green algae and red algae, but the distribution of these homologues was still patchy. Multiple homologues to *TbHP40* are found across a variety of land plants. The model organism *Arabidopsis thaliana*, which has a relatively small genome for a flowering plant (Swarbreck et al., 2008), contained two homologues (or potential paralogues), *Solanum lycopersicum*, commonly known as the tomato, contained five

homologues (or potential paralogues) and *Populus euphratica*, commonly known as the desert poplar tree, contained one sequence with strong homology to *TbHP40* and another with moderate homology to *TbHP40*. Homologous sequences to *TbHP40*, containing the 'RHG' catalytic triad, were identified in several of the green algae. Epitope-tagged *TbHP40* and *TbHP30* each localise to the mitochondrion in procyclic *T. brucei*. The analysis of the protein sequences using mitochondrial targeting prediction programs predict with varying degrees of confidence mitochondrial localisation of these trypanosome proteins (Table 3.1). It is recognised that mitochondrial targeting prediction programs are often not robust in correctly predicting mitochondrial targeting (Calvo and Mootha, 2010). However, mitochondrial targeting predictions have been trained against cohorts of *bona fide* plant mitochondrial proteins (Claros and Vincens, 1996; Emanuelsson et al., 2000; Horton et al., 2007; Small et al., 2004) and the type of N-terminal transit or sorting motifs that direct many proteins into plants chloroplasts are also relatively well defined. Thus, to perhaps gain insight into whether candidate *TbHP40* orthologues are mitochondrial or plastid targeted in other eukaryotes, candidate HP40 orthologues from plants were subject to analysis by organellar targeting prediction programs (Table 3.2). In only a few instances was a confident prediction of mitochondrial or chloroplast localisation apparent. Searches through mitochondrial and chloroplast proteomes also failed to reveal the presence of candidate HP40 orthologues from *Arabidopsis*. Thus, insight into function of these candidate plant HP40 proteins needs experimental analysis.

3.4 Homology modelling

To create a model for *TbHP40* using MOE, the full length amino acid sequence for *TbHP40* was used as the query sequence to search the protein database (PDB) within MOE for homologous sequences with solved structures that are publically available. These results were then cross-referenced for the most up-to-date results by searching PDB directly with the *TbHP40* amino acid query sequence (http://www.rcsb.org/pdb/secondary.do?p=v2/secondary/search.jsp#search_sequences).

The highest hit from both searches in the PDB database was the solved structure for the histidine phosphatase TIGAR (TP53-induced glycolysis and apoptosis regulator) from the freshwater zebrafish (*Danio rerio*, Protein Data Bank code 3E9D.A) (Li and Jogl, 2009). Although, despite TIGAR being the top hit, the *TbHP40* and TIGAR sequences only share 20% sequence identity with one another (and an E value of $6.7e^{-006}$). The solved structure for

TIGAR consists of two chains, which interact to form a homodimer. Once this structure was imported into MOE, one of the two chains was deleted, and the homology model for *TbHP40* based on chain A only. The *TbHP40* and TIGAR amino acid sequences were then aligned using MOE, in preparation for the modelling of the tertiary structure of *TbHP40*. *TbHP40* contained a 14 amino acid long N-terminal extension that was not present in TIGAR, therefore the 'N-terminus outgap modelling' option was selected. In this mode, MOE does *ab initio* modelling based on the secondary structure of the N-terminus region, which does not align with the template TIGAR sequence, to produce a model of the full length of *TbHP40*. 1000 different models for *TbHP40* were constructed during modelling, consisting of 100 main-chain models with 10 side chains. The final structure was then energy minimized using AMBER99 force-field algorithms (Wang et al., 2000a) and the final output rendered in MOE using the ribbon backbone option (Figure 3.9).

At first glance, the homology model for *TbHP40* contains the six-stranded β -sheet core with the active RHG triad located towards the top of β 1, as do the structures of TIGAR and the closely related histidine phosphatase PhoE, from the bacterium *Bacillus stearothermophilus* (Rigden et al., 2002). However, when closely scrutinised, β 4 of *TbHP40* is not part of the $\alpha/\beta/\alpha$ sandwich that contains the active site. β 4 is in fact homologous to the β 5 strand of the TIGAR structure upon which the model was based on (Figure 3.11). The 5 core β -sheets in *TbHP40* are flanked by α -helical regions, α 12 and α 9 on one side, α 3 and α 5 on the other side, forming an $\alpha/\beta/\alpha$ structure, which resembles the $\alpha/\beta/\alpha$ structure also present in the TIGAR and PhoE structures, albeit one β -strand short (Figure 3.11).

The Ramachandran plot for the *TbHP40* homology model shows the bond angles for all of the amino acids present in the sequence. The majority of amino acids are in a favourable conformation (represented by green circles in Figure 3.10). At least 70 amino acids are in a generally acceptable conformation (yellow circles in Figure 3.10), but 7 amino acids are outliers (red crosses in Figure 3.10). Attempts to manually adjust the conformations of these residues proved unsuccessful in altering them to a more favourable conformation. However, none of these outlier residues are located anywhere near the catalytic $\alpha/\beta/\alpha$ structural region of the molecule.

In attempts to gain insight about the nature of the substrate the *TbHP40* protein may dephosphorylate, surface models were created based upon the homology model backbone of *TbHP40*. Firstly the surface of the whole molecule was rendered (the grey region in Figure 3.12 C), followed by the specific modelling of the active 'RHG' triad. This showed that the

'RHG' catalytic triad is found in a fairly spherical recessed pocket at the opposite end of the protein to the N-terminus (the orange region in Figure 3.12 C). For comparison purposes with histidine phosphatases with known substrates, the solved structures of TIGAR and PhoE were also surface modelled using MOE (Figure 3.12 A and B respectively). The *TbHP40* active site more closely resembles the TIGAR active site, which is slightly more exposed than *TbHP40* but still located within a recessed pocket. The active site of PhoE is located within a deep groove even narrower than that of *TbHP40*. PhoE is known to be a broad specificity phosphatase, which acts upon relatively small molecules including 3-phosphoglycerate, α -naphthylphosphate and p-nitrophenylphosphate (Rigden et al., 2001; Rigden et al., 2002). Therefore the structure of the *TbHP40* active site suggests that potential substrates of *TbHP40* are more likely to be smaller molecules rather than large proteins, which would encounter greater steric hindrance when trying to access an active site located within a narrow deep groove.

Homology modelling was also attempted using *TbHP30* as the query sequence, but no model could be constructed since no homologous sequences could be identified in PDB. This indicates that the *TbHP30* amino acid sequence has diverged significantly from all other histidine phosphatases for which solved structures are available and perhaps supports the interpretation that it is a catalytically dead protein – *i.e.* that it has no intrinsic catalytic activity.

Attempts were also made to model a variety of other sequences that were identified as orthologues or homologues to *TbHP40* (Section 3.3). The HP40 orthologue from *Trypanosoma cruzi* identified TIGAR (PDB: 3E9D.A) as the highest hit (with an E value of $4.2e^{+000}$), but the resultant homology model only had 4 β -strands in the catalytic $\alpha/\beta/\alpha$ region. No realistic model was obtained for the *Leishmania major* HP40 orthologue despite SixA (PDB: 1UJB) being identified as the top hit (E value of $4.2e^{+000}$) due to the sequence alignment being poor. The HP40 orthologue identified in *Bodo saltans* also identified TIGAR (PDB: 3E9D.A) as the sequence with the highest identity (E value of $1.6e^{-007}$), but the resultant model was unfeasible since the active site was buried deep within the molecule and there were only 2 β -strands in the catalytic $\alpha/\beta/\alpha$ region. Again, no model could be constructed for the HP40 orthologue found in *Naegleria gruberi* due to no homologous sequences being identified in PDB. The model for the HP40 homologue from *Acanthamoeba castellanii* was based on the human PGAM5 solved structure (PDB: Q96HS1) with an E value of $2.7e^{+000}$ and the model for the HP40 homologue with the highest sequence identity to *TbHP40* (40%) from *Arabidopsis thaliana* was based on the phosphoglycerate mutase

structure from *Borrelia burgdorferi* (PDB: 4EMB). Both of the resultant models were unfeasible due to the 'RHG' catalytic triad being buried deep within the models structure with no obvious access to the surface of the molecule, and an $\alpha/\beta/\alpha$ region with less than 4 β -strands.

Although the model presented here for *TbHP40* appears to lack the one of the six β -strands required for the $\alpha/\beta/\alpha$ catalytic core, this might not necessarily mean that it is not a functional histidine phosphatase. In PhoE β -strands A, F, C D, G and H form the six-stranded mixed core, flanked by α -helices 2 and 3 on one side and 6 and 7 on the other; in TIGAR β -strands 1, 4, 2, 3, 6 and 7 form the mixed core, flanked by α -helices 8 and 7 on one side and 2 and 3 on the other. These two solved structures can be superimposed upon one another showing strong conservation of the phosphatase core fold despite PhoE functioning as a monomer and TIGAR as a dimer. When comparing *TbHP40* to PhoE and TIGAR, the core fold looks very similar, β -strands 1, 3, 2, 6 and 5 are flanked by α -helices 12 and 9 one one side, and 3 and 5 on the other. The region where the additional β -strand should be present has an α -helix (α_6 in Figure 3.11 C) and a turn in place instead. When reviewing the Ramachandran plot, of the 8 residues in this region (residues 96-103), 4 are classified as accepted but lie outside the core regions. Some of these residues that lie outside of the core areas on the plot may be responsible for the alternative tertiary structures seen in this region. In a model with less steric hindrance or in the actual *TbHP40* structure *in vivo*, there is the possibility that some of these 'accepted' residues would form the additional core β -strand structure.

It is also interesting to note that MOE 2014.09 does not display small regions of both TIGAR and PhoE structure correctly, in accordance to the published solved structures. In PhoE, two short anti-parallel β -strands are reported in the crystal structure, (labelled β_2 and β_5 in Figure 3.11 B) yet MOE represents this as a turn for what should be β_2 and a turn with a small α -helical segment for what should be β_5 . With TIGAR, β -strand 5 is not shown when a TIGAR monomer is visualised using MOE (labelled β_5 in Figure 3.11 A). Since the aforementioned *TbHP40* equivalent of the missing β -strand in the catalytic core is represented in MOE as a turn with a small α -helical region (see Figure 3.11 C), it could be that this is actually a very small β -strand region since this is the same misrepresented pattern seen in PhoE, which MOE appears to have difficulties in rendering.

Additionally, there is the suggestion that *TbHP40* may function as a heterodimer in conjunction with *TbHP30* (see data shown in Chapter 5), and so *in vivo*, significant structural changes could occur in *TbHP40* when bound in the active conformation with *TbHP30*.

Experimental data (see subsequent chapters) suggests that *TbHP40* does not necessarily function as a monomer *in vivo*, therefore a monomer model of *TbHP40* alone may not be accurate. Since there is no solved structure that is homologous to *TbHP30*, the software is currently incapable of building a heterodimeric model. Also, β -strand 5 in TIGAR, which is the equivalent of β -strand 4 in *TbHP40*, is essential for the dimer formation in TIGAR and is critical for the co-ordination of the metal ions that stabilise the dimer interface (Li and Jogl, 2009). Having this β -strand present in the predicted structure of *TbHP40* could suggest that this area of the molecule may be involved in dimer formation.

Figure 3.1

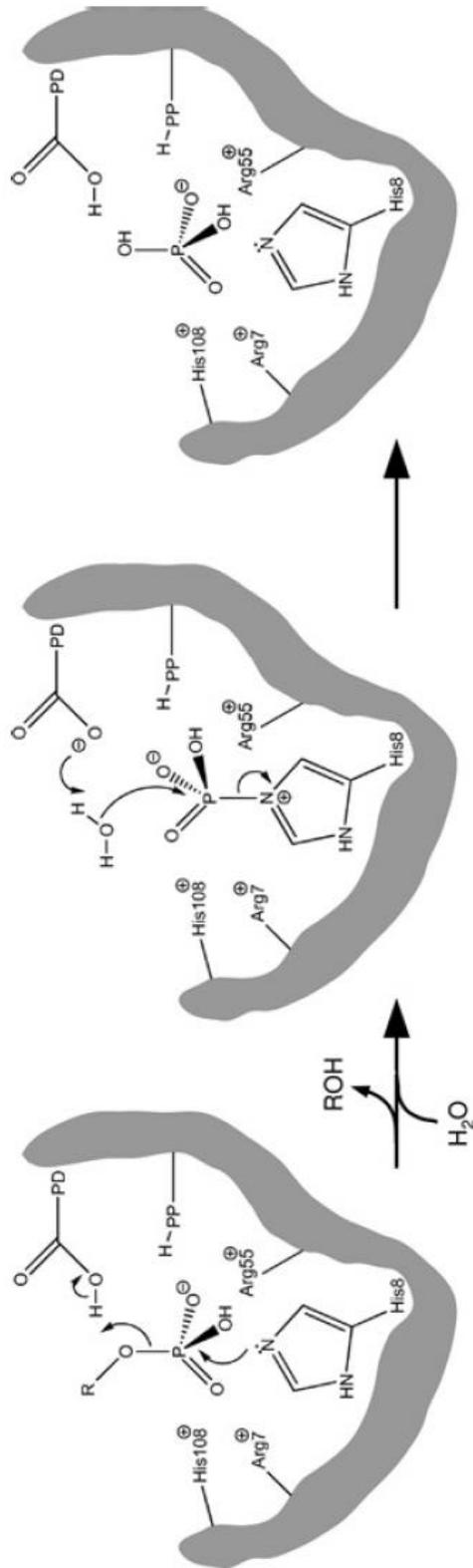


Figure 3.1. A schematic diagram detailing the mechanism of catalysis of the histidine phosphatase superfamily. Amino acids are numbered according to the *E. coli* SixA histidine phosphatase. The crucial His⁸ residue is phosphorylated and dephosphorylated during the reaction mechanism, with the other surrounding residues in the catalytic core reacting electrostatically with the phospho group to assist transfer. PP represents additional neutral or positive residues that may hydrogen bond with the phospho group and PD is a proton donor, usually aspartate or glutamate that varies in position depending upon the family the histidine phosphatase belongs to. Reproduced from (Rigden, 2008).

Figure 3.2

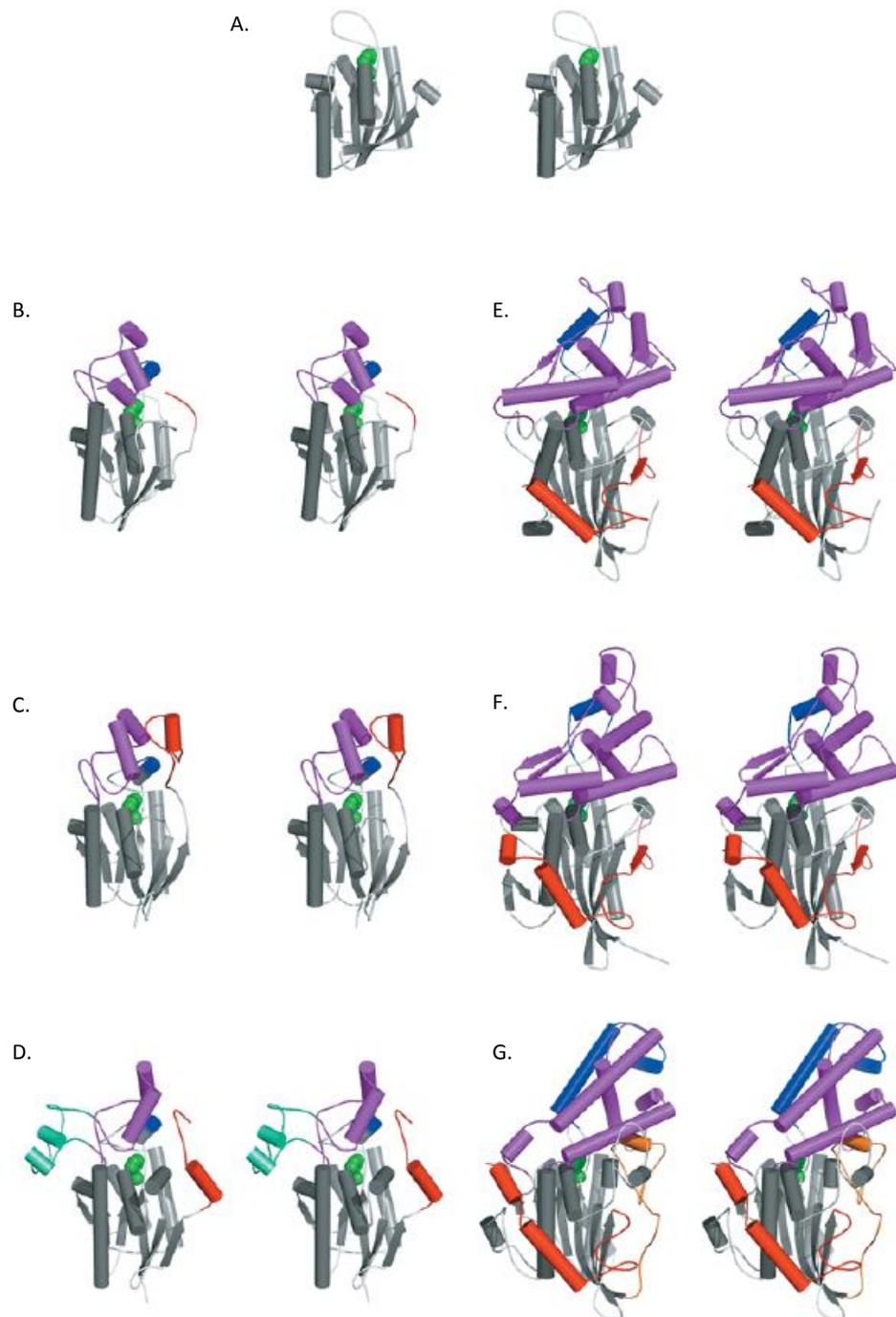


Figure 3.2. Cartoon representations of a selection of solved histidine phosphatase structures. The core structures are conserved across all histidine phosphatases, the main differences between one another occur at the N- and C- terminal regions. Core structures shown in grey; orange shows N-terminal region; red shows C-terminal region; blue highlights insertions in the $\beta 1$ - $\alpha 1$ region; turquoise shows dPGM-specific insertion. Green shows the essential conserved histidine in the active site A. *E. coli* SixA (PDB code 1UJC); B. *G. stearothermophilus* PhoE (PDB code 1H2E); C. Human liver F26BPase (PDB code 1K6M); D. *E. coli* dPGM (PDB code 1E58); E. *E. coli* glucose-1-phosphate (PDB code 1NT4); F. *E. coli* phytase (PDB code 1DKQ); G. *A. fumigatus* phytase (PDB code 1QWO). Reproduced from Rigden (2008).

Figure 3.3

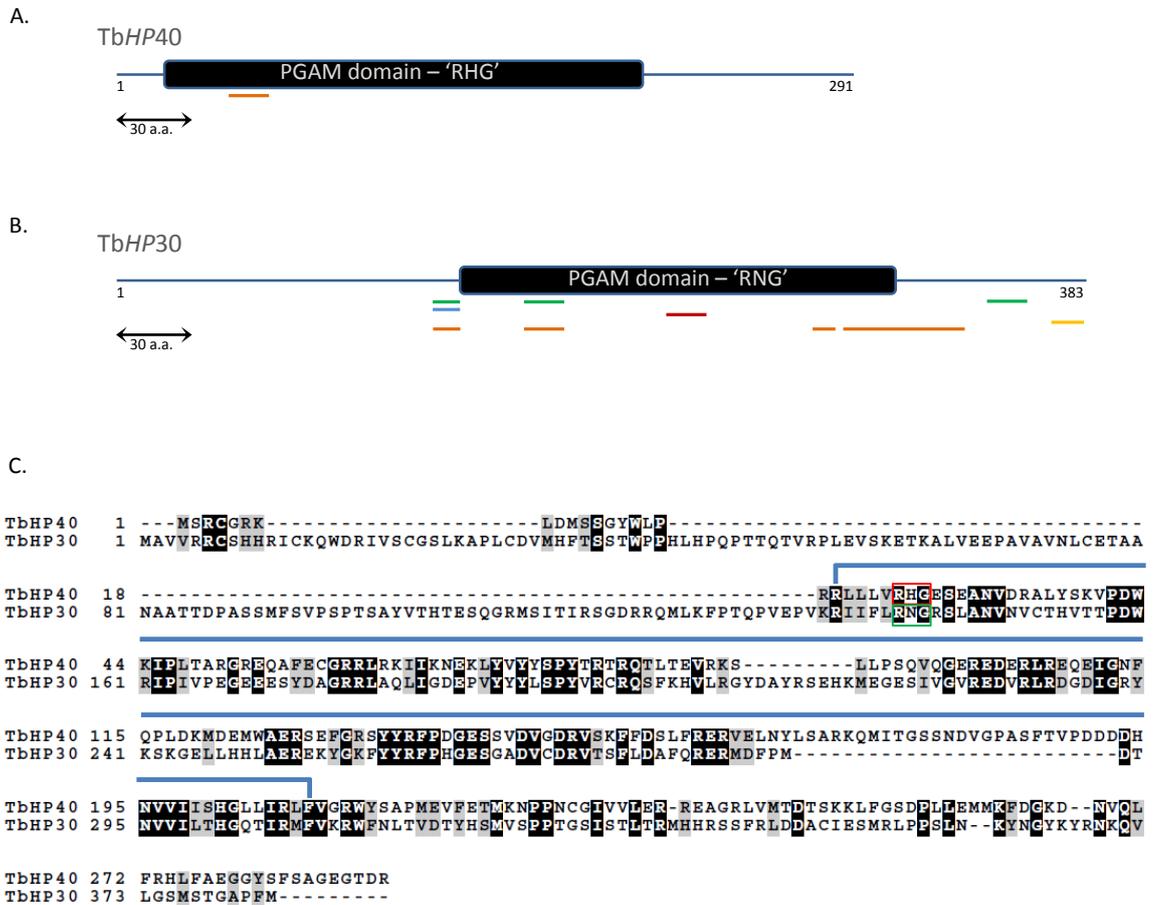


Figure 3.3. Domain architecture and conserved regions of *TbHP40* and *TbHP30*

A-B. Cartoon schematics showing the domain architecture of *TbHP40* and *TbHP30*, and evidence for expression based on peptide fragments detected by mass spectrometry during proteomics screens. Orange is from fractionated mitochondrial membrane and ER proteomes (Niemann et al., 2013); Blue is from a procyclic form mitochondrial proteome (Panigrahi et al., 2009); Red is from a mitochondrial methylarginine proteome (Fisk et al., 2013); Yellow is from a comparative phosphoproteomes of bloodstream and procyclic forms (Urbaniak et al., 2012); green is from a bloodstream form enriched kinome (Urbaniak et al., 2013).

C. A ClustalW sequence alignment of the full length sequences for *TbHP40* and *TbHP30*. Sequences were aligned using MUSCLE and coloured using BoxShade. Conserved residues are shown in black, similar residues are shown in grey. The active 'RHG' site in *TbHP40* is highlighted with a red box, the degenerative 'RNG' inactive site of *TbHP30* is highlighted with a green box. The blue lines indicate the positions of the PGAM domains.

Figure 3.4

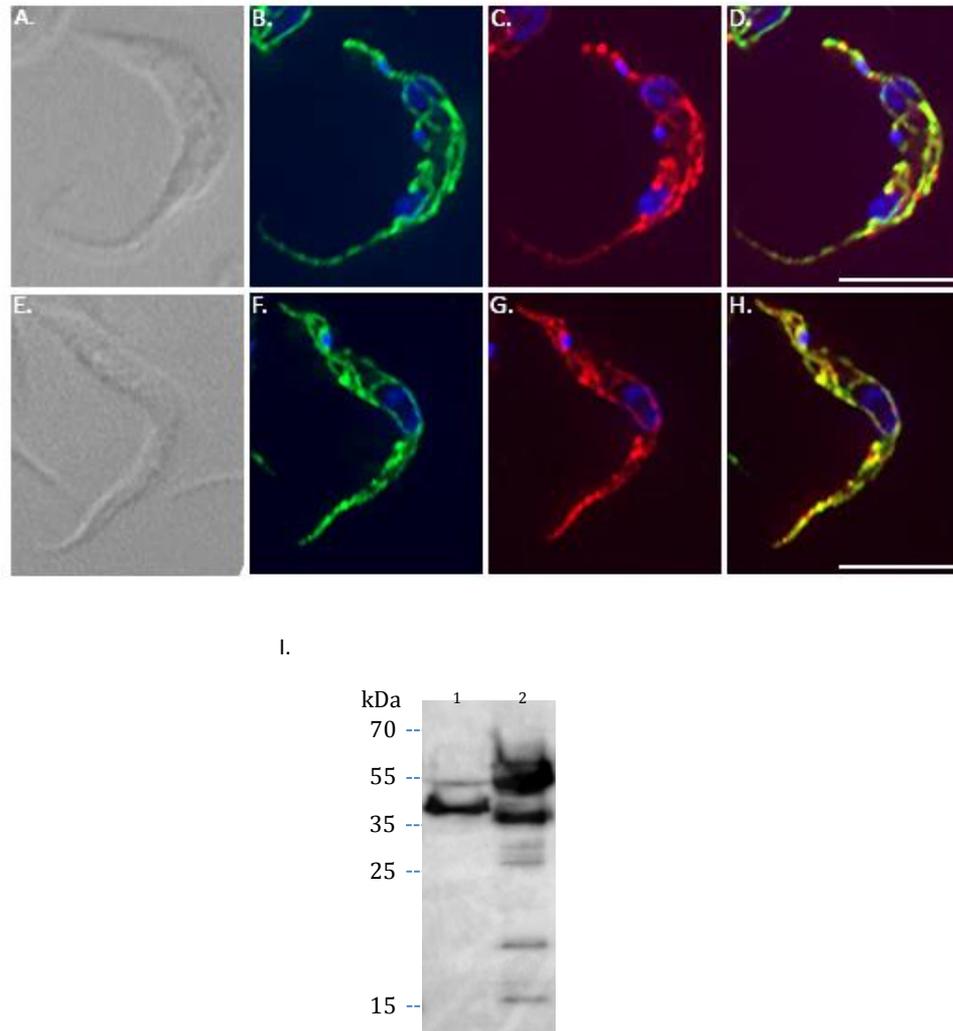


Figure 3.4. Expression and localisation of *TbHP40::myc₃* and *TbHP30::myc₃* in procyclic 427 *T. brucei*. A-D. Localisation of *TbHP40::myc₃* to the mitochondrion of *T. brucei*. Indirect Immunofluorescence was used to image the myc-tagged *TbHP40*. A: DIC; B: *TbHP40::myc₃* decorated with the anti-myc antibody at a dilution of 1/1000 in blocking buffer B, merged with DAPI; C: MitoTracker merged with DAPI; D: Merged. E-H. Localisation of *TbHP30::myc₃* to the mitochondrion of *T. brucei*. Fluorescence microscopy was used to image the myc-tagged *TbHP30*. E: DIC; F: *TbHP30::myc₃* decorated with the anti-myc antibody at a dilution of 1/1000 in blocking buffer B, merged with DAPI; G: MitoTracker merged with DAPI; H: Merged. I. Western blot analysis of whole cell lysates from cells expressing either *TbHP40::myc₃* (lane 1) or *TbHP30::myc₃* (lane 2) and decorated with anti-myc antibody at a dilution of 1/1000 in blocking buffer A.

Figure 3.5

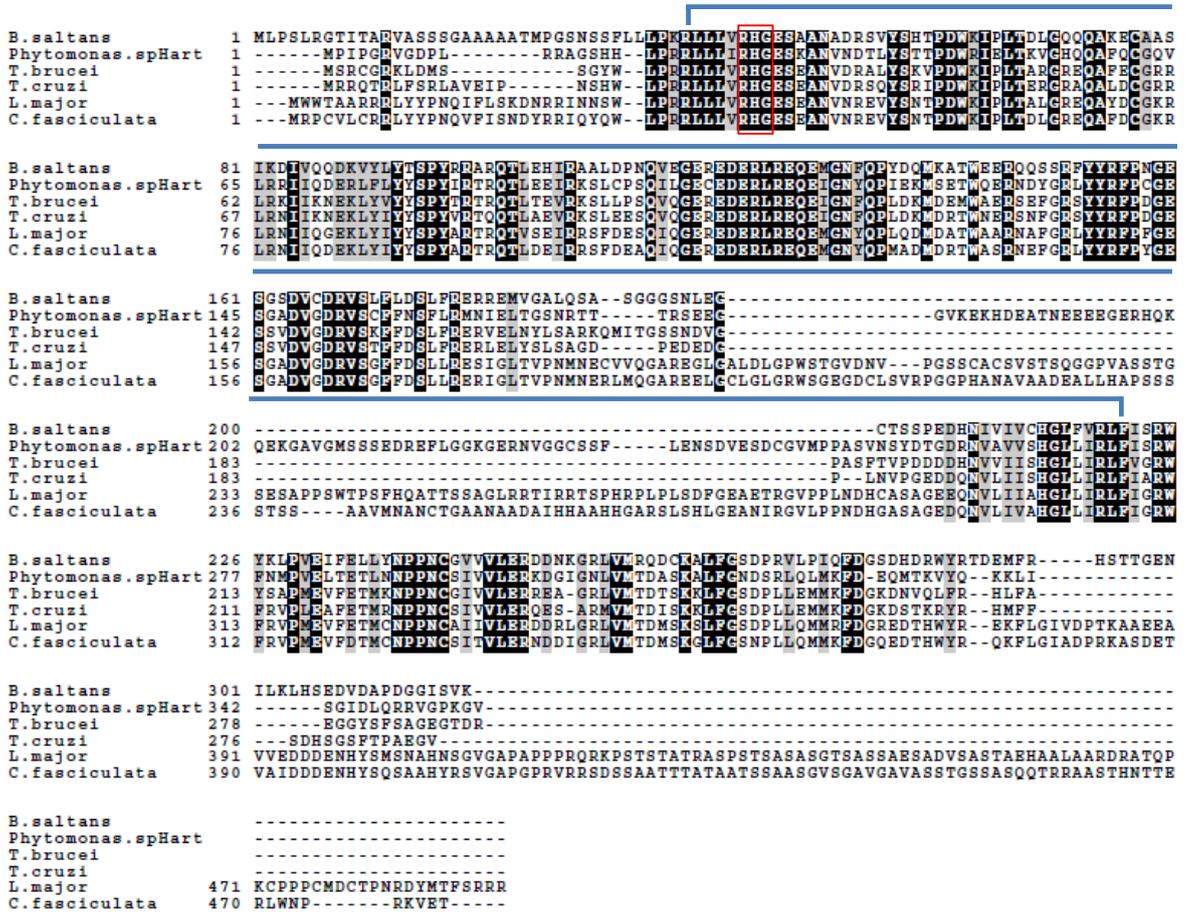


Figure 3.5. Comparison of *TbHP40* sequences from different trypanosomatids.

A ClustalW sequence alignment of the full length amino acid sequences of *TbHP40* in *T. brucei*, and orthologues identified in *Trypanosoma cruzi*, *Leishmania major*, *Crithidia fasciculata*, *Phytomonas* sp. Hart1 (Note, *TbHP40* orthologue also found in *Phytomonas* sp. EM1) and *Bodo saltans*. The alignment was performed using MUSCLE and coloured using BoxShade. Conserved residues are shown in black, similar residues are shown in grey. The highly conserved active 'RHG' sites are highlighted with a red box. The blue lines indicate the positions of the *TbHP40* PGAM domain.

Figure 3.6

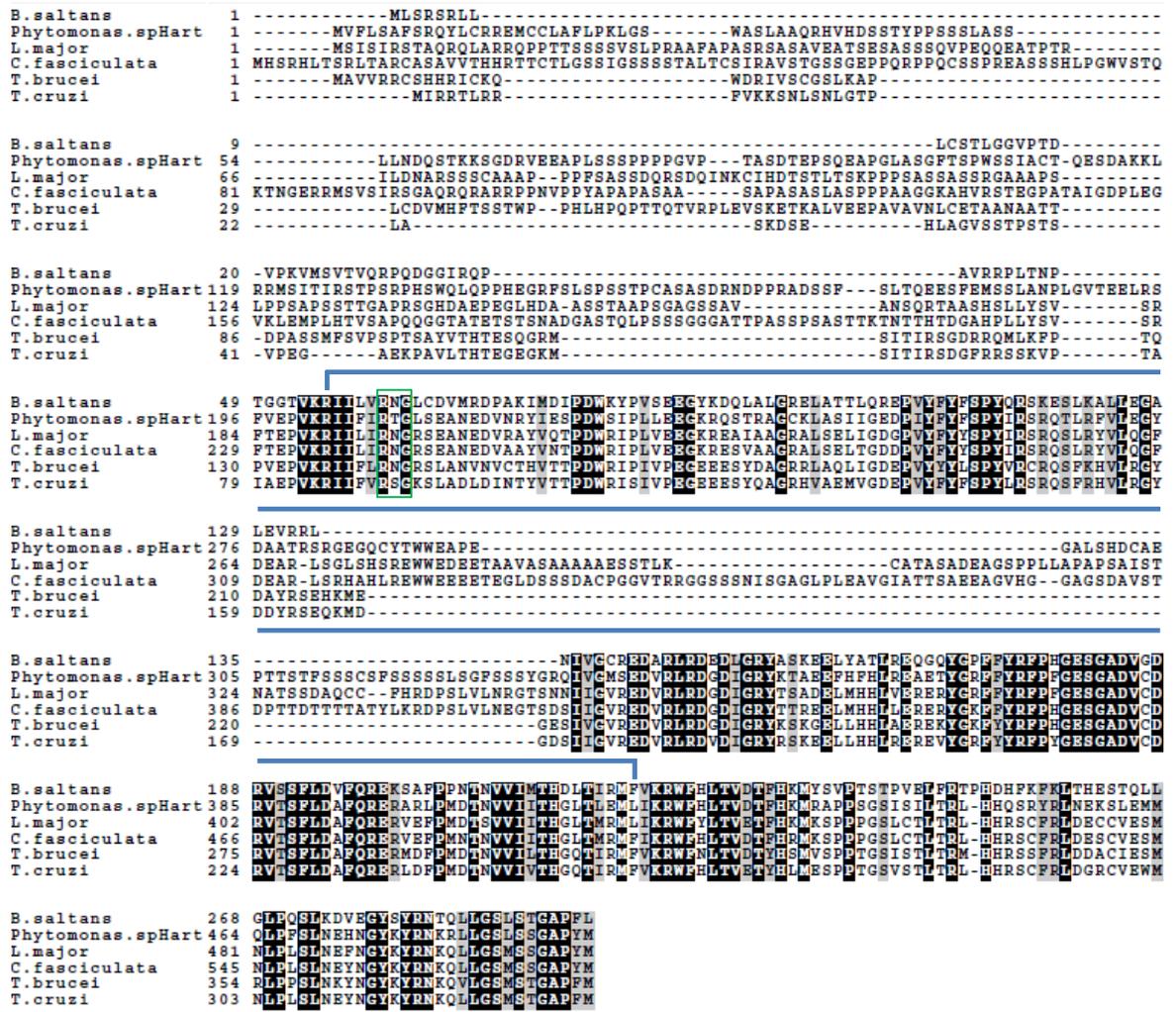
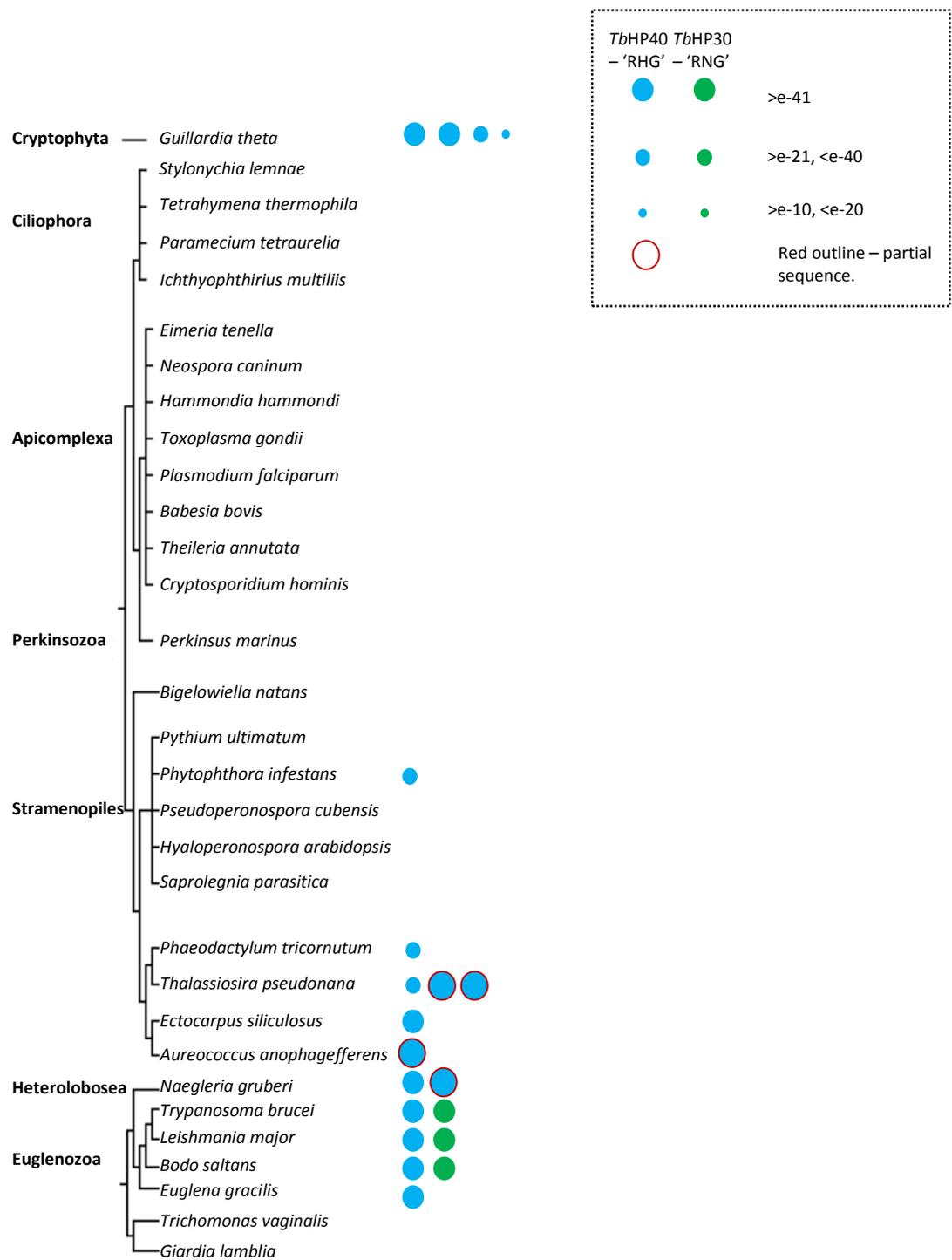


Figure 3.6. Comparison of *TbHP30* sequences different trypanosomatids.

A ClustalW sequence alignment of the full length amino sequences of *TbHP30* in *T. brucei*, and orthologues to *TbHP40* identified in *Trypanosoma cruzi*, *Leishmania major*, *Crithidia fasciculata*, *Phytomonas* sp. Hart1 (Note, *TbHP30* orthologue not found in *Phytomonas* sp. EM1) and *Bodo saltans*. The alignment was performed using MUSCLE and coloured using BoxShade. Conserved residues are shown in black, similar residues are shown in grey. The degenerated 'RXG' sites are highlighted with a green box, showing the loss of the active histidine residue in all orthologues. The blue lines indicate the positions of the *TbHP40* PGAM domain.

Figure 3.7



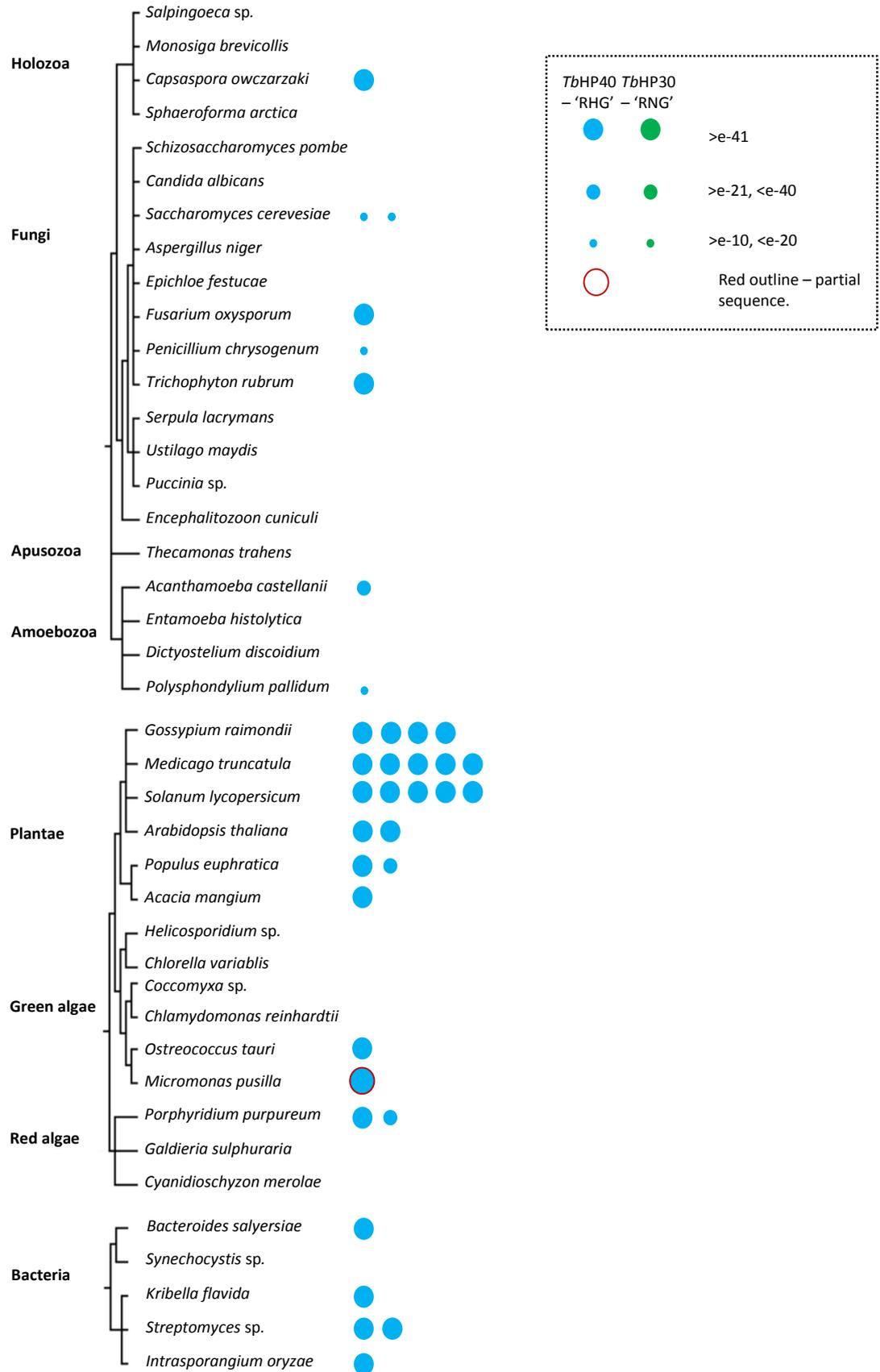


Figure 3.7 Distribution of *TbHP40* and *TbHP30* orthologues/homologues across a range of eukaryotes.

Figure 3.8

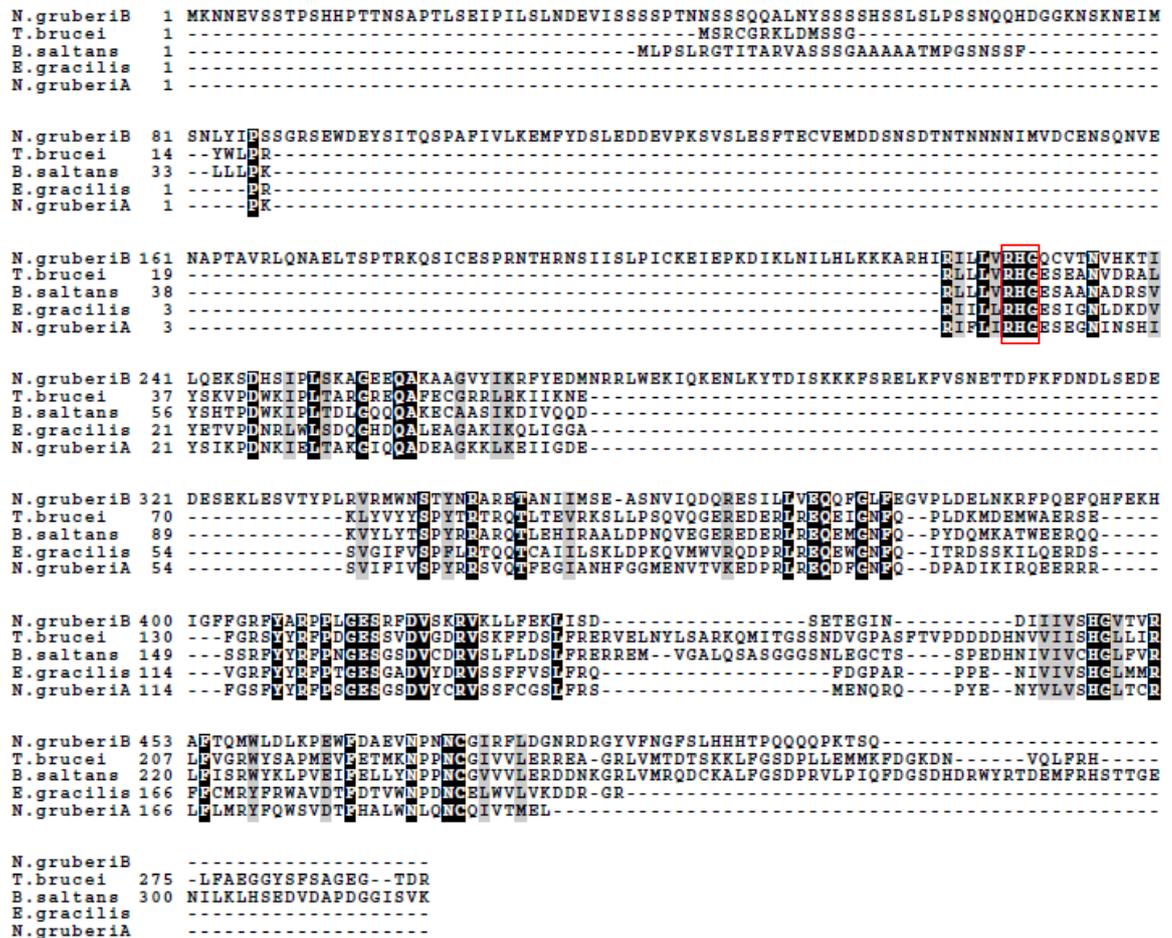


Figure 3.8. Sequence alignment of candidate *TbHP40* orthologues from several discicrinate excavates.

A ClustalW amino acid sequence alignment of full length *TbHP40* with candidate orthologues found in *Bodo saltans*, *Euglena gracilis* and *Naegleria gruberi*. In *N. gruberi*, A and B, recover *TbHP40* as top reciprocating BLAST hits. The alignment built in MUSCLE using the ClustalW algorithm and coloured using BoxShade. Conserved residues are shown in black, similar residues are shown in grey. The highly conserved active ‘RHG’ sites are highlighted with a red box.

Table 3.1

Table 3.1. Probability predictions for mitochondrial localisation of *TbHP40* and *TbHP30* using EXPASY Proteomics Server tools: MitoProt; Predotar v. 1.03; TargetP v.1.1.1.
 +, weak mitochondrial localisation prediction; ++, fair mitochondrial localisation prediction, +++ strong mitochondrial localisation prediction. Tools predicted mitochondrial localisation for all gene products with varying degrees of confidence. Output from MitoProt includes a predicted cleavage site: residue 15 for *TbHP40* and residue 63 for *TbHP30*.

Mitochondrial Targeting Program					
Protein	MitoProt	Predotar	WoLF PSORT	Target P. v.1.1	
<i>TbHP40</i>	++ (predicted cleavage product: MSRCGRKLDMSGGYW)	++	+++	+++	
<i>TbHP30</i>	+++ (predicted cleavage product: MAVVRRCSHHRICKQWDRIVSCGSLKAPLCDVMHFTSSTWPP HLHPQPTTQTVRPLEVSKETK)	+	+++	++	

Table 3.2

Table 3.2. Probability predictions for mitochondrial and/or plastid localisation of homologues to *TbHP40* from a variety of land plant species using ExpASY Proteomics Server tools: MitoProt; Predotar v. 1.03; TargetP v.1.1.1. +, weak mitochondrial localisation prediction; ++, fair mitochondrial localisation prediction, +++, strong mitochondrial localisation prediction, blank cells, no predicted mitochondrial/plastid localisation. Tools predicted mitochondrial localisation for some gene products with varying degrees of confidence. Output from MitoProt includes a predicted cleavage site: residue 28 for *P. euphratica*, residue 14 for *S. lycopersicum* and residue 14 for *S. tuberosum*.

Plastid/Mitochondrial Targeting Program						
TbHP40 homologue from organism	MitoProt (mitochondrial prediction)	Predotar (plastid prediction)	Predotar (mitochondrial prediction)	WoLF PSORT (plant chloroplast prediction)	Target P. v.1.1 (plant chloroplast prediction)	
<i>M. truncatula</i>	+	+++	+	+	++	
<i>P. euphratica</i>	++ (predicted cleavage product: MIAHSLTTSPASSLLPPRKLKSSFIQ)	+++		++	+++	
<i>S. lycopersicum</i>	+		+	++	+++	
<i>S. tuberosum</i>	++ (predicted cleavage product:MGNNSSNRTKERQ)		+	+	+	
<i>A. thaliana</i>	+		+	+	+	
<i>G. raimondii</i>	+		+	+	+	
<i>A. mangium</i>	+			+	+	

Figure 3.9

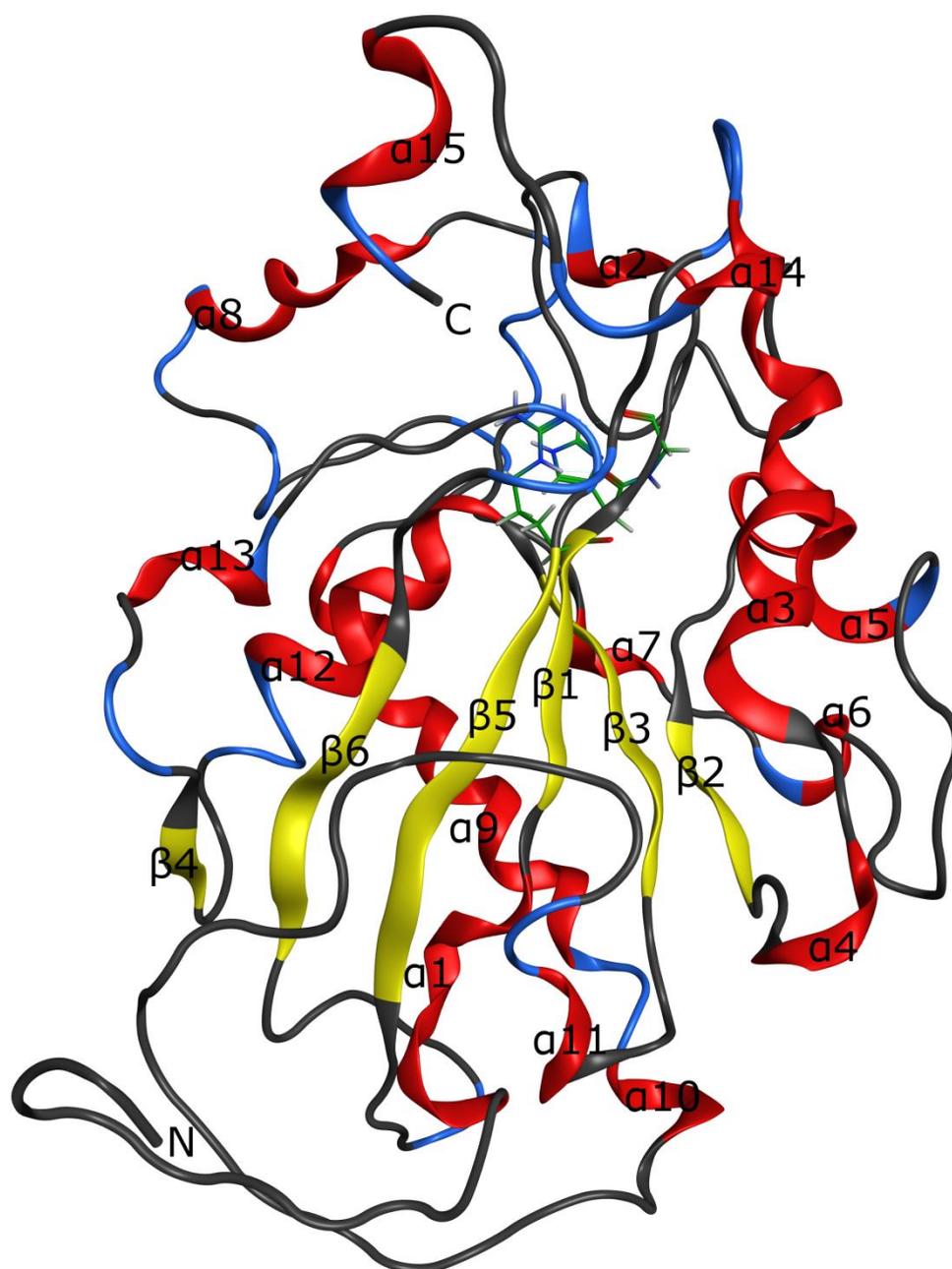


Figure 3.9. Homology model of the structure of *TbHP40*, based on sequence alignments with the solved structures of *D. rerio* TIGAR (PDB code 3E9D.A) and *B. stearrowthermophilus* PhoE (PDB code 1H2E) (Rigden, 2002).

Black shows the backbone; regions in red are α -helices, numbered 1-15; regions in yellow are β -sheets, numbered 1-6; blue areas are turns; N- N-terminus; C- C-terminus. The active 'RHG' residues are represented by green stick structures.

Figure 3.10

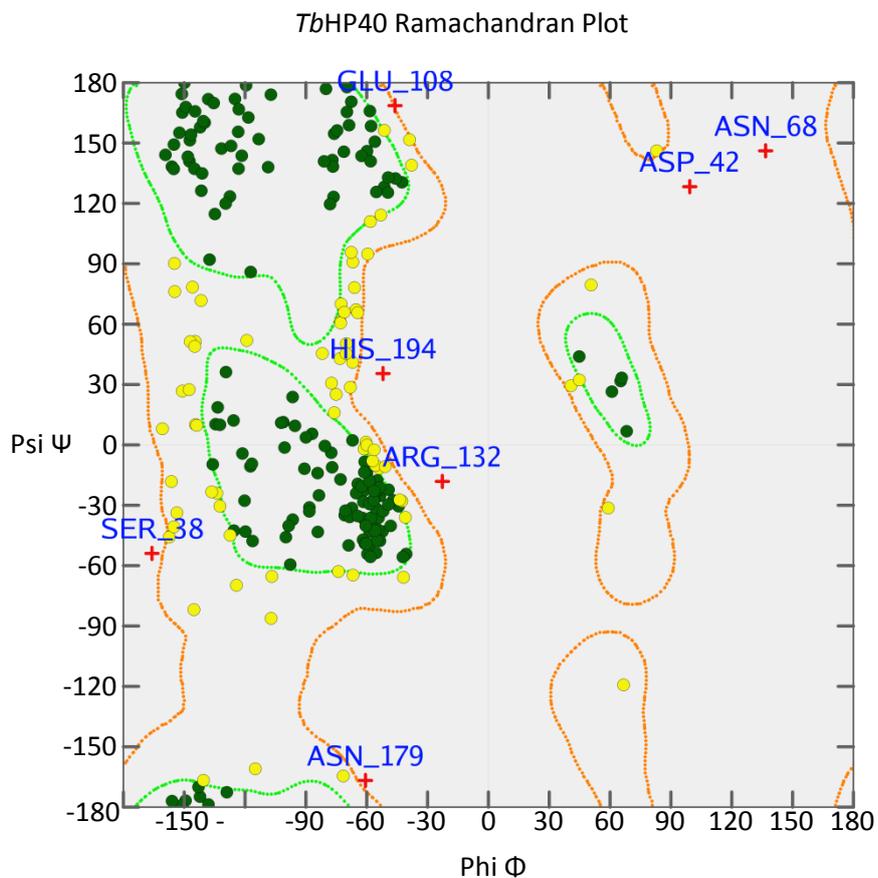


Figure 3.10. A Ramachandran plot showing the bond angles of the residues in the homology model for *Tb*HP40 (Figure 3.9).

Green circles show favoured conformations; yellow circles show accepted conformations; red crosses represent outliers, these residues are additionally labelled in blue. SER_38 – serine, position 38; ASP_42 – aspartic acid, position 42; ASN_68 – asparagine, position 68; GLU_108 – glutamic acid, position 108; ARG_132 – arginine, position 132; ASN_179 – asparagine, position 179; HIS_194 – histidine, position 194 (none of these outlier residues are associated with the active site region).

Figure 3.11

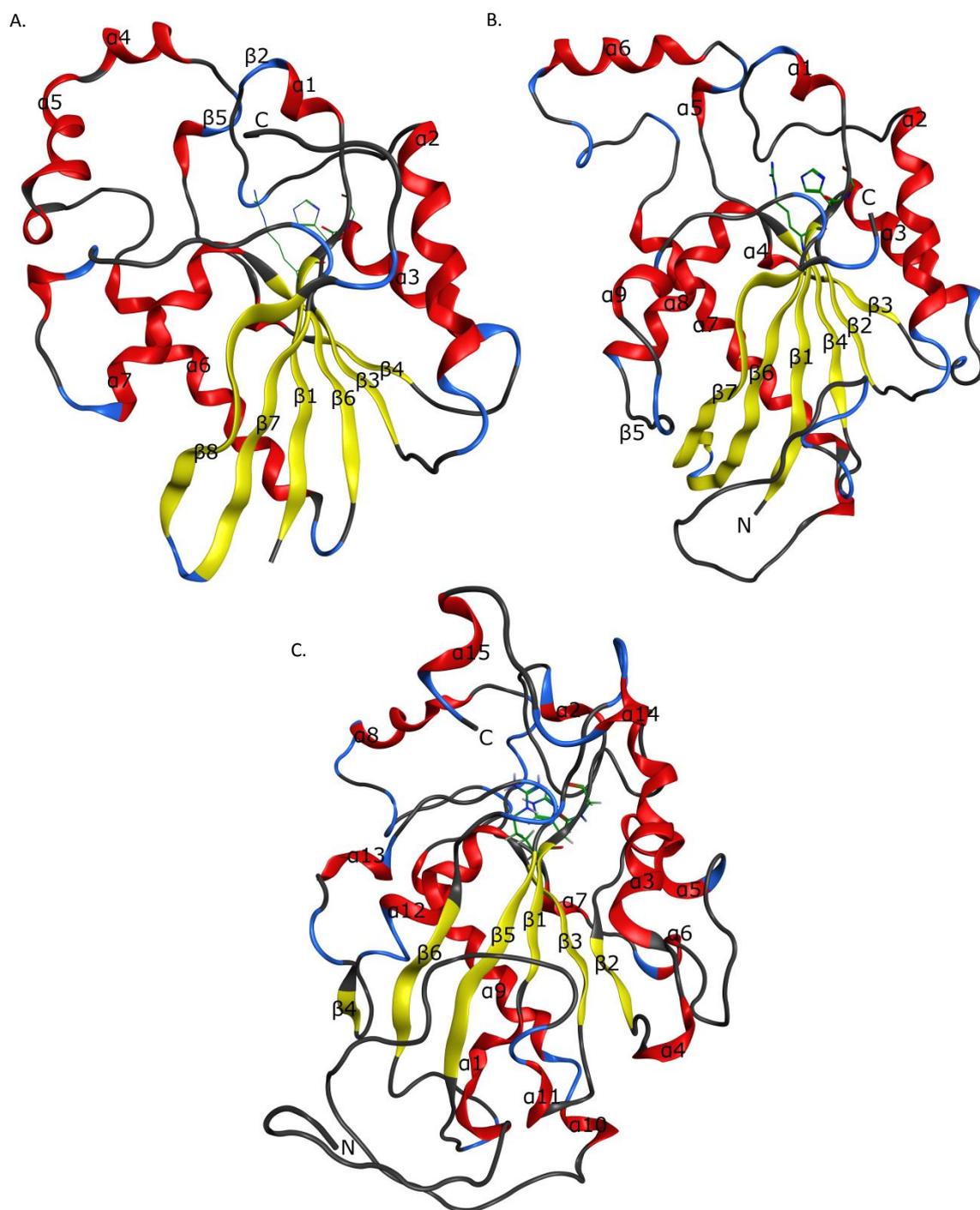


Figure 3.11. The solved structures of TIGAR and PhoE compared against the homology model of TbHP40 from *T. brucei*.

A. The structure of *D. rerio* TIGAR (PDB code 3E9D.A)

B. The structure of *B. stearotherophilus* PhoE (PDB code 1H2E) (Rigden, 2008)

C. The homology model of TbHP40 from *T. brucei*

Black shows the backbone; Regions in red are α -helices; regions in yellow are β -sheets; blue areas are turns; N – N-terminus; C- C-terminus. The active 'RHG' residues are represented by green stick structures. All rendered using MOE.

Figure 3.12

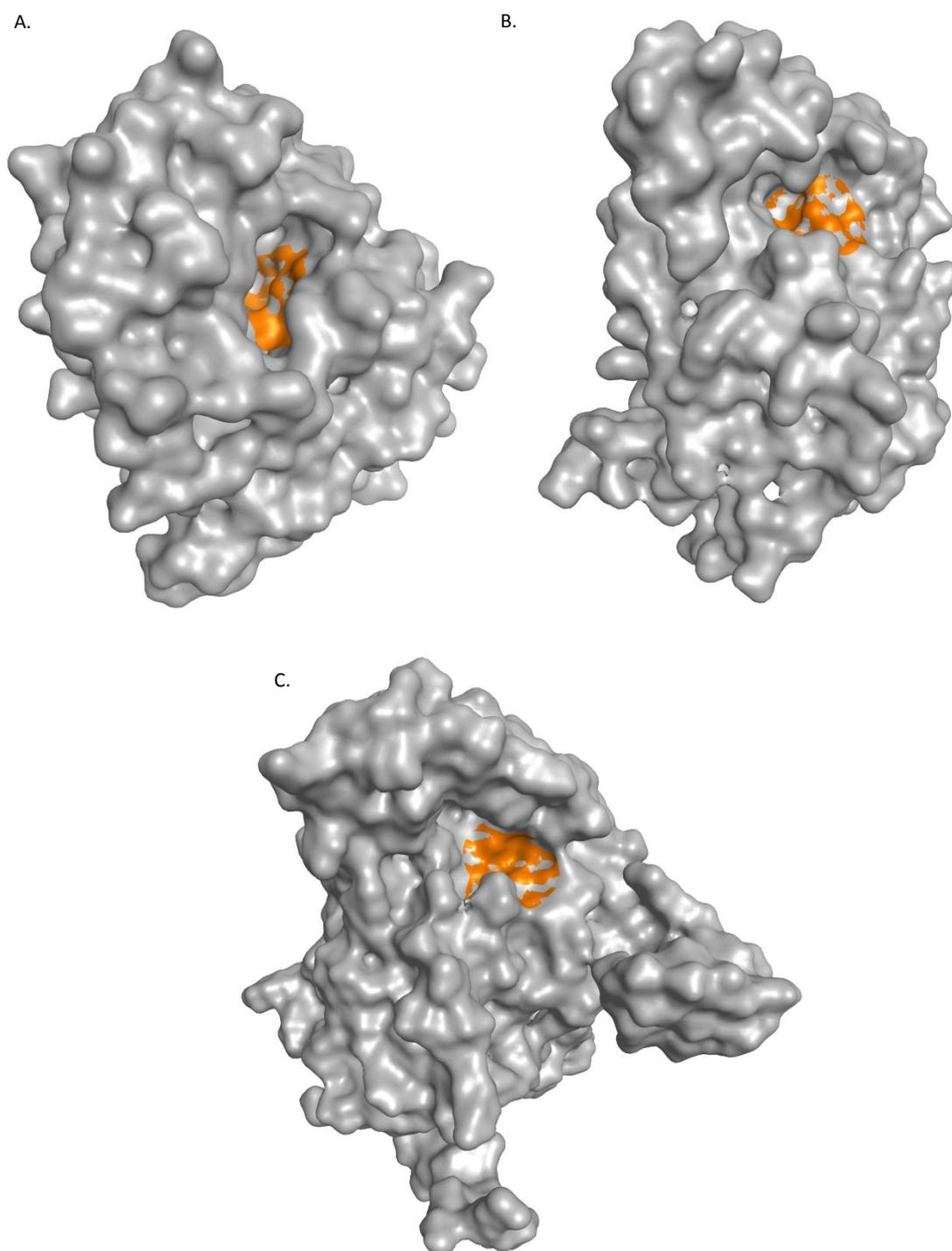


Figure 3.12. Surface modelling the solved structures of TIGAR and PhoE and the homology model of *TbHP40* from *T. brucei*.

A. The structure of *D. rerio* TIGAR (PDB code 3E9D.A)

B. The structure of *B. stearothersophilus* PhoE (PDB code 1H2E) (Rigden, 2008)

C. The homology model of *TbHP40* from *T. brucei*

The active 'RHG' residues for each structure are shown in orange. Each active site sits within a deep pocket towards one end of the protein.

Chapter 4

Evidence for the essentiality of *TbHP40* and *TbHP30*

From the investigations reported in Chapter 3, I learnt that *Tb927.11.2920* (*TbHP40*) has a patchy evolutionary distribution and I got the impression that it is catalytically active. In contrast, *Tb927.11.2910* (*TbHP30*) encodes a kinetoplastid-specific protein and is likely to be catalytically inactive. Both *TbHP40* and *TbHP30* were predicted to have N-terminal mitochondrial targeting sequences and both were shown to localise to the mitochondrion of procyclic *T. brucei*. *TbHP40* and *TbHP30* had also been identified in an array of mitochondrial proteome studies (Acestor et al., 2009; Fisk et al., 2013; Niemann et al., 2013; Panigrahi et al., 2009; Urbaniak et al., 2013), and thus both are constitutively expressed in *T. brucei*, despite the striking stage-specific differences in the extent of mitochondrial metabolism in African trypanosomes (Bringaud et al., 2006). However, beyond this, an understanding of the specific cellular function of *TbHP40* and *TbHP30* remained lacking. Thus, I set out to determine whether *TbHP40* and/or *TbHP30* were essential in procyclic *T. brucei*. The results of that investigation are presented in this chapter.

4.1 The generation of a *TbHP40/TbHP30*^{+/-} heterozygous cell line

TbHP40 and *TbHP30* are neighbouring genes on chromosome 11. Thus, as a first step to testing potential essentiality of *TbHP40* and/or *TbHP30*, one allele of the paralogous gene pair was targeted for deletion from diploid procyclic *T. brucei*.

Initially, plasmids encoding genes conferring resistance to either phleomycin or blasticidin S HCl and flanked at the 5' and 3' coding sequences by tubulin or actin intergenic sequences, respectively, (to allow appropriate mRNA processing and expression of the phleomycin *Sh ble* gene product or blasticidin deaminase) were used as PCR templates. Forward and reverse oligonucleotide primers that annealed to the tubulin or actin intergenic sequences and which contained at their 5'-ends 30 bp of homology to the start of the *TbHP40* open reading frame (forward primer) or 30 bp of homology to the end of the *TbHP30* coding sequence (reverse primer) were used in PCR as shown in Figure 4.1 A. PCR products from this reaction were then used as templates for a second PCR where new forward and reverse primers annealed to the first and last 20 nucleotides of the first PCR amplicon, which was now used as the DNA template. Each primer in the second PCR possessed at the 5' end either an additional 30 bp

matching sequence at the 5' end of *TbHP40* (forward primer) or the end of *TbHP30* (reverse primer). Thus a final PCR product was produced in which the phleomycin resistance-conferring *Sh ble* gene or blasticidin deaminase gene were flanked by appropriate mRNA processing signals and 60 bp homology flanks to allow integration of the drug resistance cassette into the *T. brucei* genome via homologous recombination.

PCR products (~3 µg of DNA) were electroporated into procyclic *T. brucei* and stable transformants were selected by the addition of blasticidin S HCl (10 µg ml⁻¹) or phleomycin (3 µg ml⁻¹). From drug resistant populations genomic DNA (gDNA) was isolated, digested with *EcoRI* and separated according to size by agarose gel electrophoresis. The digested DNA was then transferred onto an Amersham Hybond-N membrane and the membrane hybridised against an alkaline phosphatase labelled DNA sequence gene probe that corresponded to an upstream intergenic region (UIR) (*i.e.* a probe that lies in the intergenic region between *TbHP40* [*Tb11.02.2920*] and the next upstream gene located on chromosome 11 [*Tb11.02.2930*]). The membrane was then washed and incubated with a chemiluminescent substrate before visualisation on x-ray film. The parental 427 gDNA, when digested with *EcoRI* and hybridised with the UIR probe, was expected to result in one band ~10 kB in size on the Southern blot (Figure 4.1 B). The gDNA from the populations resistant to either blasticidin S HCl or phleomycin, when digested with *EcoRI* and hybridised with the UIR probe, would yield two bands - one ~10 kB in size which would correspond to a wild-type allele (Figure 4.1 A), and another ~8 kB in size, which would correspond to the correct integration of the the blasticidin S HCl or phleomycin resistance marker (Figure 4.1 C).

Southern blot analysis showed that the *EcoRI* digested wild-type 427 gDNA was as expected. The presence of two hybridising bands in the gDNA extracted from cells transfected with the gene encoding blasticidin deaminase indicated disruption, as expected, of one allele of *TbHP40/TbHP30* (Figure 4.1 E). However, there was no evidence for the correct integration of the phleomycin resistance cassette, since only one band of ~10 kB was detected in that lane and this corresponded to the expected size of a wild-type allele (Figure 4.1 E, middle lane). Thus, the results from the Southern blot were consistent with generating a procyclic *T. brucei* population that was resistant to blasticidin S HCl and also haploid for *TbHP40/TbHP30* (henceforth referred to as 4030KO^{+/-blast}). Further confirmation was sought by PCR; here, primers corresponding to the intergenic sequence upstream of *TbHP40* (forward primer) and the intergenic sequence downstream of *TbHP30* (reverse primer) were used to amplify two bands in the 4030KO^{+/-blast} gDNA (this primer combination allowed the amplification of both a

wild-type allele and a disrupted allele). One band was ~1.8 kB in size, and the other was ~3.8 kB. This 2 kB reduction in size was indicative of the successful integration of the blasticidin resistance cassette. As expected, in the wild-type 427 control gDNA, only one band ~3.8 kB in size was amplified. Thus, the PCR analysis provided further evidence that 4030KO^{+/-blast} was heterozygous for TbHP40/TbHP30.

4.2 Generation of a TbHP40/TbHP30^{-/-} cell line – attempt 1

No further attempts to produce a TbHP40/TbHP30^{-/-} population resistant to phleomycin were made. Instead, my efforts initially focused on attempts to use 4030KO^{+/-blast} cells to generate a mutant null for both TbHP40 and TbHP30.

Initially, a synthetic DNA construct was purchased in which a puromycin N-acetyltransferase gene was flanked at the 5' end of its coding sequence with 192 bp of intergenic sequence immediately upstream of TbHP40, and 192 bp of intergenic sequence immediately downstream of TbHP30 at its 3' coding end. This puromycin resistance cassette was transformed into procyclic 427 *T. brucei* as a control to show appropriate integration could occur, thereby generating heterozygous 4030KO^{+/-puro}. Integration of the puromycin resistance conferring cassette resulted in a size reduction of ~1.8 kB (compared to wild-type 427 gDNA) when EcoRI digested gDNA was hybridised against the UIR probe used for the analysis shown in Figure 4.1 (Figure 4.2 A). Two independent transfections of the puromycin resistance cassette into 4030KO^{+/-blast} cells were carried out. Following the overnight recovery of transfected cells, one transfection was selected in quadruplet at the population level by the addition of puromycin (2 µg ml⁻¹) or puromycin (2 µg ml⁻¹) and blasticidin S HCl (10 µg ml⁻¹) to 1x10⁶ cells in 10 ml of fresh SDM-79 medium. Stable transformants were then allowed to grow through independently. This resulted in four candidate null (-/-) populations for which gDNA was made, digested with EcoRI and screened by Southern blot using the UIR probe. All of these candidate TbHP40/TbHP30 null mutant populations still contained a wild-type allele. The second independent transfection was selected both at a population level in the presence of blasticidin and puromycin, and at a clonal level using two 96-well plates in which 200 µl of culture (at 2x10⁶ cells ml⁻¹) was plated per well. One plate was selected in the presence of only puromycin, the other with puromycin and blasticidin. Stable transformants were obtained from the population selection. Seven wells from the 96-well plate grown with puromycin yielded viable populations of cells. Clonal transformants were transferred to

separate flasks containing 1 ml of fresh SDM-79 medium, and incubated with both puromycin and blasticidin. Once this 1 ml culture reached stationary phase, a further 9 ml of fresh SDM-79 medium was added. Only six of these wells continued to grow in the presence of both drugs. Only two wells from the plate incubated with puromycin and blasticidin gave populations of viable cells following 11 days of selection. These populations were treated in the same way as candidate clonal lines from the 96-well plate incubated with puromycin. Both of the new candidate clonal lines remained resistant to both puromycin and blasticidin drugs. All nine candidate nulls (-/-) were screened by Southern blot using the UIR probe, but a wild-type allele was still detected in all transformants selected at both clonal and population levels.

Thus, from two independent transformations followed by population and/or clonal level selection, no cell lines completely null for *TbHP40/TbHP30* were isolated. This provided preliminary evidence to suggest that *TbHP40* and/or *TbHP30* are essential in procyclic *T. brucei*.

4.3 Generation of a *TbHP40/TbHP30*^{-/-} cell line – attempt 2

Following the preliminary indication that *TbHP40* and/or *TbHP30* were essential I undertook additional experiments. Here, the aim was to take the 4030KO^{+/-blast} population to ectopically express either *TbHP40::myc₃* or *TbHP30::myc₃* on the +/- genetic background and to then attempt the deletion of the remaining *TbHP40* and *TbHP30* allele.

Constructs permitting expression of *TbHP40::myc₃* or *TbHP30::myc₃* (as described in Chapter 3) were linearised, and transfected separately into 4030KO^{+/-blast} cells. Stable transformants expressing *TbHP40::myc₃* or *TbHP30::myc₃* were selected by the addition of hygromycin (50 µg ml⁻¹). Whole cell lysates from populations expressing either *TbHP40::myc₃* or *TbHP30::myc₃* were analysed by Western blotting, and cover slips containing cells fixed with *para*-formaldehyde were permeabilised and processed for indirect immunofluorescence microscopy. The primary antibody used in these analyses was anti-myc. From the Western analysis of both transformed populations, one population expressed a myc-tagged protein ~34 kDa in size, which corresponded to the predicted size of *TbHP40::myc₃*. The other population expressed myc-tagged protein where the dominant protein band was ~43 kDa in size, which corresponded to the predicted size of *TbHP30::myc₃* (Figure 4.3 A). In the latter, smaller fainter bands detected on the Western blot were most likely caused by proteolytic

degradation of *TbHP30::myc₃*. Expression of myc-tagged *TbHP40* and *TbHP30* was further confirmed using immunofluorescence, which showed ~100% of cells constitutively expressed either *TbHP40::myc₃* or *TbHP30::myc₃*, which, as expected, localised to the mitochondrion of *T. brucei* in both populations (Figure 4.3 B-C).

Independent transfections of the puromycin resistant cassette into 4030KO^{+/-blast} expressing either *TbHP40::myc₃* or *TbHP30::myc₃* were then carried out. The continued expression of either *TbHP40::myc₃* or *TbHP30::myc₃* was confirmed visually by indirect immunofluorescence before each transfection. In the first of two independent sets of transfections, stable transformants were clonally selected using puromycin and two 96-well plates (one plate for transformants expressing *TbHP40::myc₃* and one plate for transformants expressing *TbHP30::myc₃*). Only two wells from the plate in which 4030KO^{+/-blast}/*TbHP40::myc₃* cells were subject to puromycin selection grew out. These clonal lines were transferred to separate 1 ml flasks of fresh SDM-79 and grown in the presence of puromycin. Again, upon reaching stationary phase, the culture volume was expanded to 10 ml. When the culture volume was increased, the cultures were also split into one fraction cultured in the presence of puromycin, the other cultured in the presence of puromycin and blasticidin. Both clonal lines survived exposure to blasticidin and puromycin. Genomic DNA was made from all four cultures and analysed by Southern blot using the hybridisation UIR probe (Figure 4.2 B). All four transformants still retained a copy of the wild-type allele. To verify the result of a rough-looking blot, the *EcoRI* digests of the gDNA was repeated and a new Southern blot membrane hybridised with a probe that corresponded to the open reading frame of *TbHP40* (40ORF). If the deletion of the second allele had been successful in any of the candidate null (-/-) cell lines, a wild-type allele ~10 kb in size would not be detected. The Southern blot hybridised against the 40ORF probe did show the presence of a wild-type allele in all of the candidate null (-/-) cell lines. A second hybridisation signal corresponding to *TbHP40::myc₃* (~4 kb in size) was detected in all cell lines apart from the wild-type 427 and 4030KO^{+/-blast} cells.

In a second set of transfections, stable transformants were clonally selected using puromycin or puromycin and blasticidin, resulting in four 96-well plates. Any wells that survived selection were transferred to 1 ml cultures of SDM-79 containing puromycin, blasticidin and hygromycin on a 24-well plate (hygromycin being the drug selection marker for the expression of *TbHP40::myc₃* or *TbHP30::myc₃*). Transformants grew through in three waves. The initial wave resulted in fourteen clonal lines from 4030KO^{+/-blast}/*TbHP30::myc₃* (thirteen

originating from the 'puromycin' plate, one from the 'puromycin and blasticidin' plate), only ten of these clones survived selection in the presence of all three drugs (nine originating from the 'puromycin' plate, one from the 'puromycin and blasticidin' plate). Twelve clonal lines were obtained from the $4030\text{KO}^{+/-\text{blast}}/\text{TbHP40}::\text{myc}_3$ selection (ten originating from the 'puromycin' plate, two from the 'puromycin and blasticidin' plate) and only seven of these survived selection in the presence of all three drugs (five originating from the 'puromycin' plate, two from the 'puromycin and blasticidin' plate). The next wave of transformants saw no clonal lines from the $4030\text{KO}^{+/-\text{blast}}/\text{TbHP30}::\text{myc}_3$ selection plates, and nine from the $4030\text{KO}^{+/-\text{blast}}/\text{TbHP40}::\text{myc}_3$ selection (all nine originating from the 'puromycin' plate); eight clones survived selection with all three drugs. The final wave of transformants to emerge again saw no growth from the $4030\text{KO}^{+/-\text{blast}}/\text{TbHP30}::\text{myc}_3$ selection plates, and four clones from the $4030\text{KO}^{+/-\text{blast}}/\text{TbHP40}::\text{myc}_3$ selection (all four again originating from the 'puromycin' plate); all four survived selection with all three drugs.

Since this experiment yielded such a high number of cell lines, initial screening was done using PCR with the intention of further analysing of any candidate mutants null for both *TbHP30* and *TbHP40* by Southern blotting. Genomic DNA was harvested from all cell lines and then used as DNA template in separate PCR reactions, using the primers detailed in Table 2.3. Two PCR reactions were carried out on each DNA template with primer pairs designed to amplify either the *TbHP40* open reading frame or the *TbHP30* open reading frame. Since the cell lines expressed either *TbHP40* or *TbHP30*, one of the two PCR reactions always acted as a control to make sure that the gDNA harvested provided a template amenable to PCR, so a band of ~1 kB was always expected in one of the two reactions (dependent on whether *TbHP40*::*myc*₃ or *TbHP30*::*myc*₃ were expressed and which primer pair was used). If no bands were detected from the other PCR, then this would be indicative of a candidate mutant null for the endogenous loci containing *TbHP30* and *TbHP40*, assuming the remaining wild-type *TbHP40*/*TbHP30* allele would have been replaced with the puromycin resistance cassette.

All cell lines obtained from these analyses still retained a wild-type allele, which was amplified by PCR as described above. A representative agarose gel showing six PCR reactions, one using a template from $4030\text{KO}^{+/-\text{blast}}/\text{TbHP30}::\text{myc}_3$ background and two using templates from a $4030\text{KO}^{+/-\text{blast}}/\text{TbHP40}::\text{myc}_3$ background are shown in Figure 4.4 B. If the deletion had been successful in the $4030\text{KO}^{+/-\text{blast}}/\text{TbHP30}::\text{myc}_3$ cell line, a band ~1 kB in size would have been detected only in lane 2 (which used the *TbHP30* ORF primer pair, and would amplify

the gene encoding *TbHP30::myc₃*) and not in lane 1, since this reaction used the *TbHP40* ORF primer pair. Since a band ~1 kB in size was detected in lane 1, this showed that a wild-type allele of *TbHP40/TbHP30* was still present in the genome.

Again, from two independent transformations followed by clonal level selection, no double-knockout cell lines were confirmed, irrespective of whether cells expressed either *TbHP40::myc₃* or *TbHP30::myc₃*. This analysis provided further evidence to suggest that both *TbHP40* and *TbHP30* are essential in procyclic *T. brucei*.

4.4 Generation of a *TbHP40/TbHP30*^{-/-} cell line – attempt 3

Another strategy was devised to test the essentiality of *TbHP40* and *TbHP30*, whereby the 4030KO^{+/-blast} population was manipulated further to express both *TbHP40::myc₃* and *TbHP30::myc₃*. This cell line was then subjected to a final round of genetic manipulation to delete the remaining locus encoding *TbHP30* and *TbHP40*. If both *TbHP40* and *TbHP30* are essential in procyclic *T. brucei* (perhaps because the gene products interact with one another in order to function properly), then this may explain why the deletion of the second allele had proved so difficult to achieve when only either *TbHP40::myc₃* or *TbHP30::myc₃* were ectopically expressed in a *TbHP40/TbHP30* heterozygous (+/-) background.

To carry out this approach DNA containing the coding sequence of *TbHP40::myc₃* was isolated from the pDex377 vector by restriction digest using BamHI and NsiI and then cloned into the pDex477-Y2 expression vector, which had been digested with BamHI and NsiI prior to ligation. Following selection of transformants with phleomycin this would permit expression of myc-tagged *TbHP40*. Ligated plasmids were transformed into XL-1 blue competent *E. coli* and plasmid DNA prepared in the usual way using the GeneJET™ Plasmid Miniprep kit (ThermoFisher Scientific). Plasmid DNA was screened for the presence of insert using BamHI and NsiI restriction digests and a plasmid containing the correct insert was confirmed by DNA sequencing. This new *TbHP40::myc₃* construct was linearised using the restriction enzyme NotI and then transfected into the existing '4030KO^{+/-blast} expressing *TbHP30::myc₃*' cell line and also wild-type procyclic 427 cells. Since immunofluorescence would not be able to distinguish between the two different mitochondrial myc-tagged proteins expressed in the same cell, evidence of successful transformation and expression of both *TbHP40::myc₃* and *TbHP30::myc₃* was sought only by Western blot for the transformed 4030KO^{+/-blast}/*TbHP30::myc₃* cell line. Western blots probed with anti-myc antibody (Figure

4.5 A) detected a single protein band, which corresponded to the expected size of *TbHP40::myc₃* (~34 kDa), in the transfected 427 background (lane 1; Figure 4.5 A). For proteins blotted from the whole cell lysate prepared from 4030KO^{+/-blast} cells, co-expressing both *TbHP40::myc₃* and *TbHP30::myc₃* two protein-containing bands were detected, one that corresponded to the expected size of *TbHP40::myc₃* (~34 kDa) and the other that corresponded to the expected size of *TbHP30::myc₃* (~43 kDa) (lane 2; Figure 4.5 A).

Following the confirmation by Western blot, two independent transfections of the puromycin resistance-conferring KO cassette into 4030KO^{+/-blast} cells expressing both *TbHP40::myc₃* and *TbHP30::myc₃* were carried out. For the first transfection, stable transformants were selected on a population level, by the addition of puromycin (2 µg ml⁻¹) and blasticidin (10 µg ml⁻¹) to electroporated cells. Genomic DNA was isolated from the drug resistant population and screened by Southern blot using a hybridisation probe that corresponded to the open reading from of *TbHP30* (30ORF). If the second wild-type allele for *TbHP40* and *TbHP30* had been successfully replaced by the puromycin resistance cassette, then there would be no detection of a wild-type band ~10 kB in size and only a band of ~3.8 kB, which would correspond to the ectopic expression of *TbHP30::myc₃*. Unfortunately, the presence of a wild-type band was still detected in the candidate null (-/-) population.

For the second transfection, first the continued expression of *TbHP40::myc₃* and *TbHP30::myc₃* was confirmed by Western blot before the puromycin resistance cassette was transfected into 4030KO^{+/-blast} cells expressing both *TbHP40::myc₃* and *TbHP30::myc₃*. Stable transformants were selected clonally using various drug combinations: puromycin ('puro') (2 µg ml⁻¹); blasticidin ('blast') (10 µg ml⁻¹); hygromycin ('hyg') (50 µg ml⁻¹); and phleomycin ('phleo') (3 µg ml⁻¹). Four 96-well plates were therefore set up for the selection. Any wells that survived the initial drug selections were transferred into 1 ml cultures of SDM-79 containing 'puro/blast/hyg/phleo' on a 24-well plate and viable cell populations expanded to 10 ml cultures for the isolation of gDNA. Again, transformants grew through in three waves. Wave one resulted in eighteen clonal cell lines (five from the 'puro' plate, six from the 'puro/blast' plate, five from the 'puro/blast/hyg' plate and two from the 'puro/blast/hyg/phleo' plate). Only two of these candidate clonal lines did not survive in the presence of all four drugs (one originating from the 'puro' plate and the other from the 'puro/blast' plate). The second wave of emerging transformants resulted in two clonal cell lines (one from the 'puro/blast' plate and the other from the 'puro/blast/hyg' plate); both survived quadruple drug selection. The third 'wave' resulted in only one more clonal cell line

(from the 'puro/blast/hyg' plate), which also grew in the presence of all four drugs. Genomic DNA was made from all nineteen cell lines, digested with EcoRI, and all candidate nulls were screened by Southern blot, using the 3OORF probe. A representative Southern blot of these screens can be seen in Figure 4.5 B, where fourteen of the mutant candidate nulls for *TbHP30* and *TbHP40* were screened. All cell lines appeared to still retain a wild-type allele when probed with the 3OORF probe. However, the clarity of this exposure was not perfect but the hybridisation signals in some lanes gave a hint of mutants potentially null for both endogenous alleles of *TbHP40/TbHP30*. Thus, gDNA from these clones (corresponding to lanes 5, 6 and 7 in blot shown in Figure 4.5 B) were re-analysed by Southern blot (Figure 4.5 D). This Southern blot was again probed with the 3OORF probe, and a ~10 kB band that corresponded to the expected size of a wild-type allele was readily detected in all lanes. The small ~3.8 kB band was only detected in the three candidate null mutants expressing *TbHP30::myc₃*.

Despite being able to survive and grow in the presence of all four drugs, all nineteen candidate null (*TbHP30/TbHP40*^{-/-}) cell lines still retained a copy of the wild-type allele. Thus, there was never any convincing evidence that a null mutant for *TbHP40* and *TbHP30* was ever obtained. This again pointed to essentiality of both *TbHP40* and *TbHP30* in procyclic *T. brucei*.

4.5 Progress with the generation a conditional *TbHP40*^{-/-} cell line

A final approach was taken to assess the possible essentiality of *TbHP40* and *TbHP30*. Here, I attempted to make a conditional gene knock out, where only the disruption of *TbHP40* occurred. The experiments here used the 29-13 strain of procyclic *T. brucei*, in which cells express both T7 RNA polymerase (not relevant here) and a tetracycline repressor protein to facilitate inducible expression of genes targeted for conditional knock out. For the conditional knock out, a first allele of *TbHP40/TbHP30* was targeted for deletion using the puromycin resistance cassette. Subsequently, *TbHP40::myc₃* encoded within pDex477-Y2 was expressed in these cells, thereby facilitating ectopic expression of *TbHP40::myc₃*, and then deletion of the remaining gene copy of *TbHP40* was attempted.

The puromycin resistance cassette was transfected into 29-13 cells and stable transformants were selected both at the population level and clonally with puromycin. Three candidate clonal lines and the population survived selection with puromycin and gDNA was made from

each cell line, digested with EcoRI and screened by Southern blot using the UIR probe. The Southern blot detected two bands in clonal lines C and D, a wild-type ~10 kB band and a smaller band ~8.2 kB which corresponded to the expected size reduction that would be observed following the successful integration of the puromycin resistance cassette (Figure 4.6 B).

The attempt to delete one *TbHP30/TbHP40* allele in a 29-13 background using the puromycin resistance cassette had thus been successful, and one clonal line was taken forward for further genetic manipulation (henceforth referred to as 4030KO^{+/-puro}C). DNA encoding *TbHP40::myc₃* was then transfected into 4030KO^{+/-puro}C cells and stable transformants were selected using phleomycin (*TbHP40::myc₃* in the pDex377 vector could not be used, since the selection marker for this vector is hygromycin and the 29-13 cell line used for conditional gene KO was already resistant to hygromycin). Phleomycin-resistant cells were next incubated in the presence or absence of doxycycline for 24 hours, to induce the expression of *TbHP40::myc₃*. Cover slips were made, fixed with *para*-formaldehyde, permeabilised and processed for indirect immunofluorescence. When incubated with doxycycline, only ~10% of the population exhibited high-level inducible expression of *TbHP40::myc₃* and the remainder of the cells showed very minimal or no expression (*cf* constitutive expression of *TbHP40::myc₃*, in a 427 background where 100% stably express myc-tagged *TbHP40* at relatively homogenous levels).

In order to generate a conditional gene knock out of *TbHP40*, the final round of selection for stable transformants had to be done in the presence of doxycycline, to maintain expression of *TbHP40::myc₃* during the selection process. Since only ~10% of cells expressed *TbHP40::myc₃*, the likelihood of obtaining a null mutant in which both wild-type alleles of *TbHP40/TbHP30* have been disrupted was likely to be low. Despite this, 4030KO^{+/-puro}C/*TbHP40::myc₃* cells were transfected with a synthetic construct that would confer blasticidin resistance and which targeted the remaining copy of *TbHP40* for KO – *i.e.* I attempted to produce a conditional *TbHP40* KO in which an endogenous copy of *TbHP30* remained. Transformants were selected clonally in the presence of doxycycline, puromycin and blasticidin. Unfortunately, no transformants from the final round of transfection were obtained, and given the low number of cells that could be induced to express *TbHP40::myc₃* no further transfections were carried out.

Although it is perhaps somewhat unsatisfactory in that I never got a cell line or population in which both alleles encoding the paralogous *TbHP40* and *TbHP30* gene pair were lost, the fact

that nine independent transfections were carried out in which there was potential, albeit not realised potential, for the second allele encoding TbHP30 and TbHP40 to be disrupted, nonetheless provides likely indication that at least TbHP40, if not TbHP40 and TbHP30 are essential in procyclic *T. brucei*.

Figure 4.1

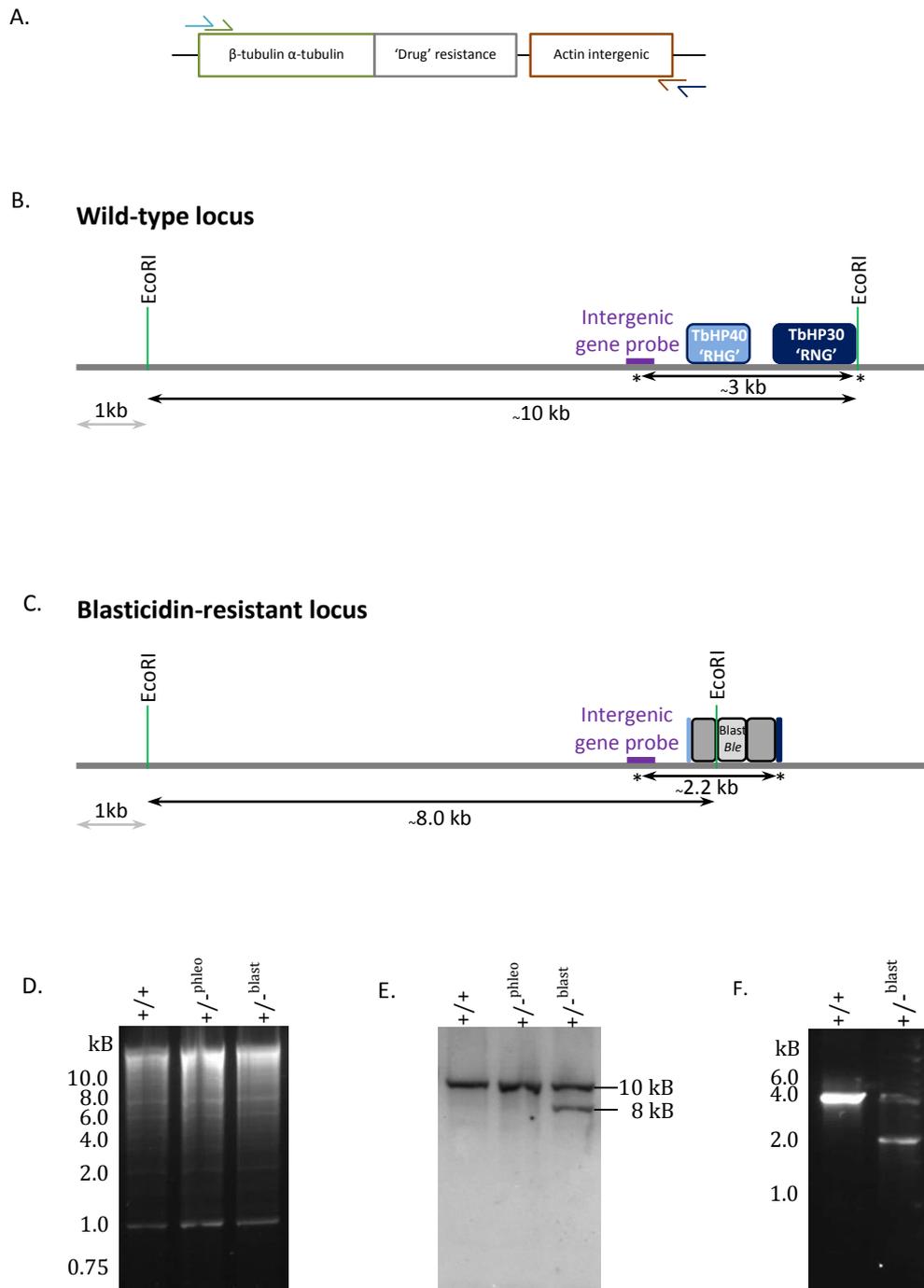


Figure 4.1. Southern analysis of candidate *TbHP40*/*TbHP30*^{+/-} populations.

A. A cartoon schematic of the blasticidin or phleomycin resistance cassettes. The gene for either blasticidin deaminase or the phleomycin *Sh ble* gene product is flanked at the 5' end with tubulin $\beta\alpha$ intergenic sequence and at the 3' end with actin intergenic sequence. The primer pairs used at the 5' end (detailed in Table 2.3) were 4030KOA (shown in dark green) and 4030KOC (highlighted in light blue). At the 3' end the primer pair used was 4030KOB (shown in dark orange) and 4030KOD (shown in dark blue). Not drawn to scale.

B. A restriction map showing the locations of the *Eco*RI digest sites upstream and downstream of *TbHP40* and *TbHP30* in a wild-type allele. Sequence corresponding to upstream intergenic region (purple) provided the DNA probe used for Southern analysis.

C. A restriction map showing the locations of the *Eco*RI digest sites upstream and downstream of the *TbHP40* and *TbHP30* gene locus after the successful integration of the blasticidin deaminase or phleomycin *sh ble* resistance cassette. The blasticidin or phleomycin resistance cassette (described in more detail in Figure 4.1 A) is represented by the light and dark grey boxes. The remaining (following integration of the resistance cassette) short 5' end of *TbHP40* and the remaining short 3' end of *TbHP30* are represented by light and dark blue boxes respectively.

D. A post-stained 0.7% TBE agarose gel showing the *Eco*RI digested gDNA from wild-type 427 (+/+), and candidate blasticidin S HCl-resistant (+/-^{blast}) or phleomycin-resistant (+/-^{phleo}) *T. brucei* populations.

E. The Southern blot of the gel shown in Figure 4.1 D, hybridised with the upstream intergenic region probe. In the wild-type 427 (+/+) lane only one band ~10 kB in size is detected when probed. In the blasticidin S HCl-resistant population (+/-^{blast}) two hybridisation signals ~10 kB and ~8 kB in size were detected. This indicated that one of the two alleles had been deleted and replaced with the blasticidin resistance cassette.

F. A post-stained 0.8% TAE agarose gel showing the PCR amplicons resulting from PCR using the primer combination UpIG0445 (forward primer) and 3040KO6 (reverse primer) (see Table 2.3, also shown as astrisks in Figure 4.1 B and C) and genomic DNA isolated from either 427 (lane 1) or 4030KO^{+/-blast} (lane 2) cells.

Figure 4.2

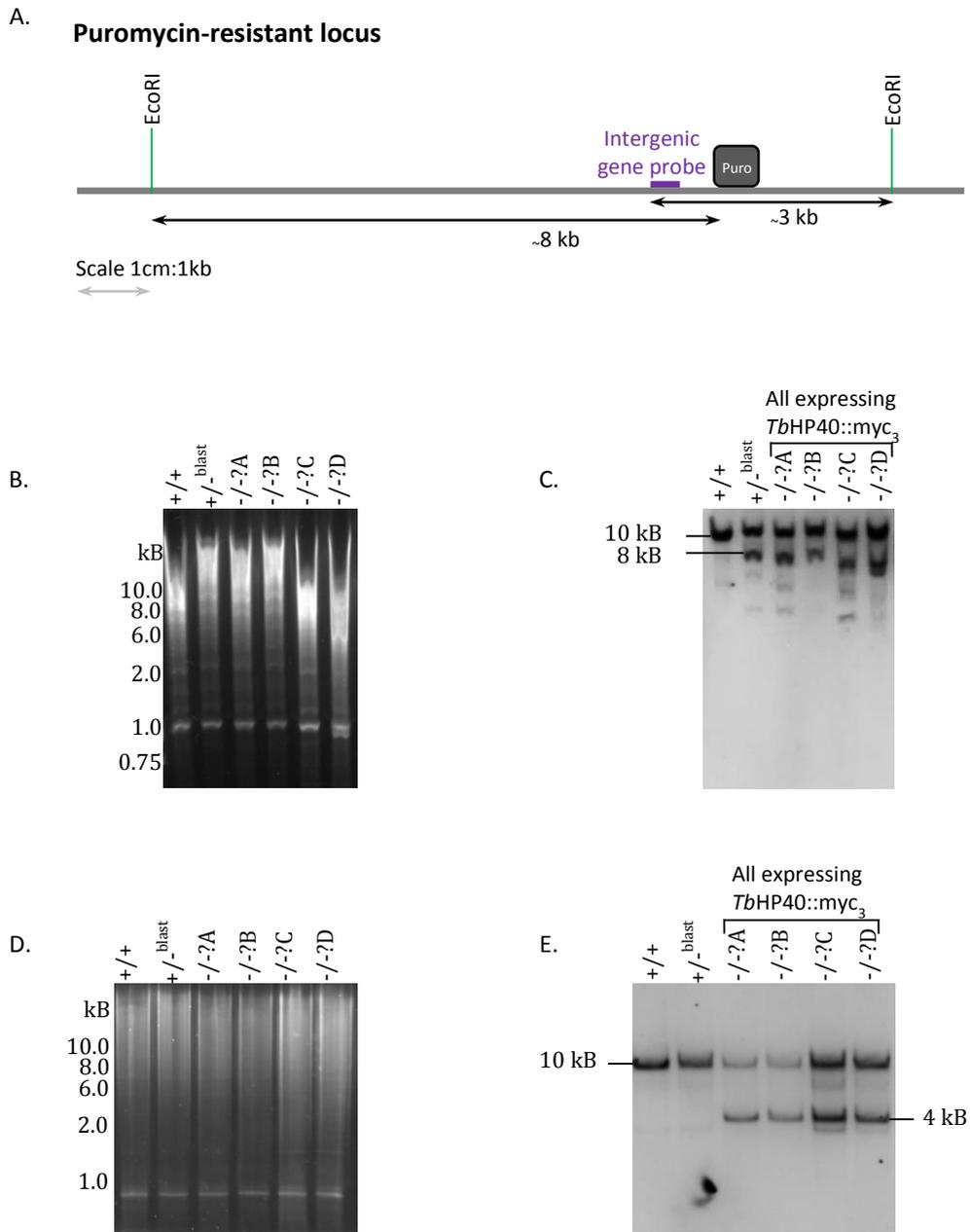


Figure 4.2. Southern analysis for the disruption of the *TbHP40* and *TbHP30* gene locus through homologous recombination using a puromycin N-acetyltransferase gene.

(4030KO^{+/-blast} cells expressing *TbHP40::myc₃* or *TbHP30::myc₃*)

A. A restriction map showing the locations of the EcoRI digest sites upstream and downstream of the *TbHP40* and *TbHP30* gene locus after the successful integration of the puromycin N-acetyltransferase resistance cassette. The puromycin N-acetyltransferase is represented by the grey box and labelled 'puro'.

B. The post-stained 0.7% TBE agarose gel showing the EcoRI digested gDNA from wild-type 427 (+/+), the blasticidin S HCl-resistant (+/-^{blast}) and four candidate null (-/-) *T. brucei* populations. (-/-)A was initially grown in the presence of puromycin, then split and one half grown in the presence of puromycin and blasticidin [(-/-)B]. (-/-)C was also initially grown in the presence of puromycin then split and one half grown in the presence of puromycin and blasticidin [(-/-)D].

C. The Southern blot of the gel shown in B, hybridised with the UIR probe. The wild-type 427 (+/+) lane detected only one band ~10 kB in size, which was also detected in all of the candidate null (-/-) populations which indicated the puromycin resistance cassette had not successfully replaced the remaining *TbHP40* and *TbHP30* gene locus. The ~8 kB band detected in the (+/-^{blast}), showing the successful integration of the blasticidin resistance construct, was also detected in all of the candidate null (-/-) populations. Smaller, fainter bands of unknown origin were detected in most of the lanes, so the Southern blot was repeated to clarify the results observed here (see below).

D. The post-stained 0.7% TBE agarose gel showing the EcoRI digested gDNA, the lanes are the same as described in B.

E. The Southern blot of the gel shown in D, hybridised with the *TbHP40* open reading frame (40ORF) probe. A wild-type allele ~10 kB was detected in all of the candidate null populations. The smaller ~4 kB band detected only in the candidate null (-/-) populations corresponded to the ectopic locus for expression of *TbHP40::myc₃*.

Figure 4.3

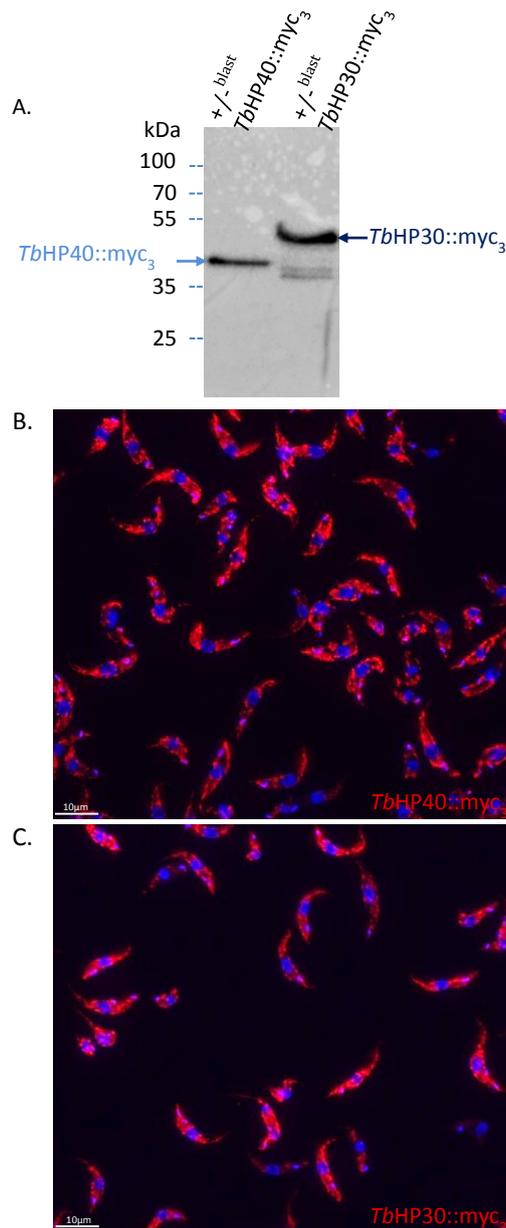


Figure 4.3. Expression of *TbHP40::myc₃* or *TbHP30::myc₃* in 4030KO^{+/-blast} cells.

A. Western blot detection of *TbHP40::myc₃* or *TbHP30::myc₃* in a 4030KO^{+/-blast} background. Whole cell lysates were run on a 10% SDS-PAGE gel and transferred onto an Amersham Hybond-P membrane. The Western blot was decorated with the anti-myc antibody at a dilution of 1/1000 in blocking buffer.

B. Representative indirect immunofluorescence image panel of 4030KO^{+/-blast} cells expressing *TbHP40::myc₃*, illustrating ~100% of cells constitutively expressed the *TbHP40::myc₃*. The merged image shows both *TbHP40::myc₃* expression and DAPI-stained DNA in para-formaldehyde fixed cells decorated with the anti-myc antibody at a dilution of 1/1000 in blocking buffer B (Section 2.30).

C. As in B, except 4030KO^{+/-blast} cells expressing *TbHP30::myc₃* were imaged.

Figure 4.4

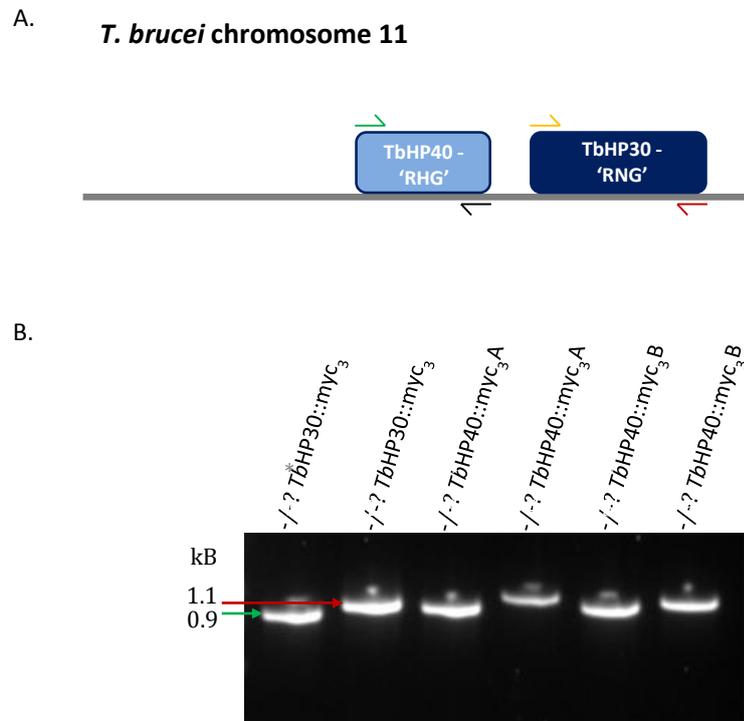


Figure 4.4 Screening candidate *TbHP40*/*TbHP30* null (-/-) mutants by PCR.

A. A cartoon schematic of the *TbHP40* and *TbHP30* gene loci on chromosome 11 of *T. brucei*. The positions of the primer pairs used in screening the gDNA from candidate null (-/-) cell lines are highlighted with arrows above *TbHP40* (forward primer shown in green, named 40ORF F and detailed in Table 2.3. The reverse primer shown in black, named 40ORF R and detailed in Table 2.3) and *TbHP30* (forward primer shown in yellow, named 30ORF F and detailed in Table 2.3). The reverse primer shown in red, named 30ORF R and detailed in Table 2.3).

B. A representative agarose gel showing the PCRs using gDNA template from three of the candidate null (-/-) cell lines expressing either *TbHP40::myc₃* or *TbHP30::myc₃* using two different sets of primer pairs (40ORF F&R and 30ORF F&R). The PCR reactions in lanes 1-2 use gDNA from a candidate null expressing *TbHP30::myc₃* and the 30ORF F&R primer pair. If the disruption of the second gene locus using the puromycin resistance cassette had been successful, a PCR product would not be observed in lane 2. The PCR reactions in lanes 3-6 use gDNA from two different candidate null expressing *TbHP40::myc₃* and the 40ORF F&R primer pair. If the disruption of the second gene locus using the puromycin resistance cassette had been successful, a PCR product would not be observed in lanes 3 and 6.

Figure 4.5

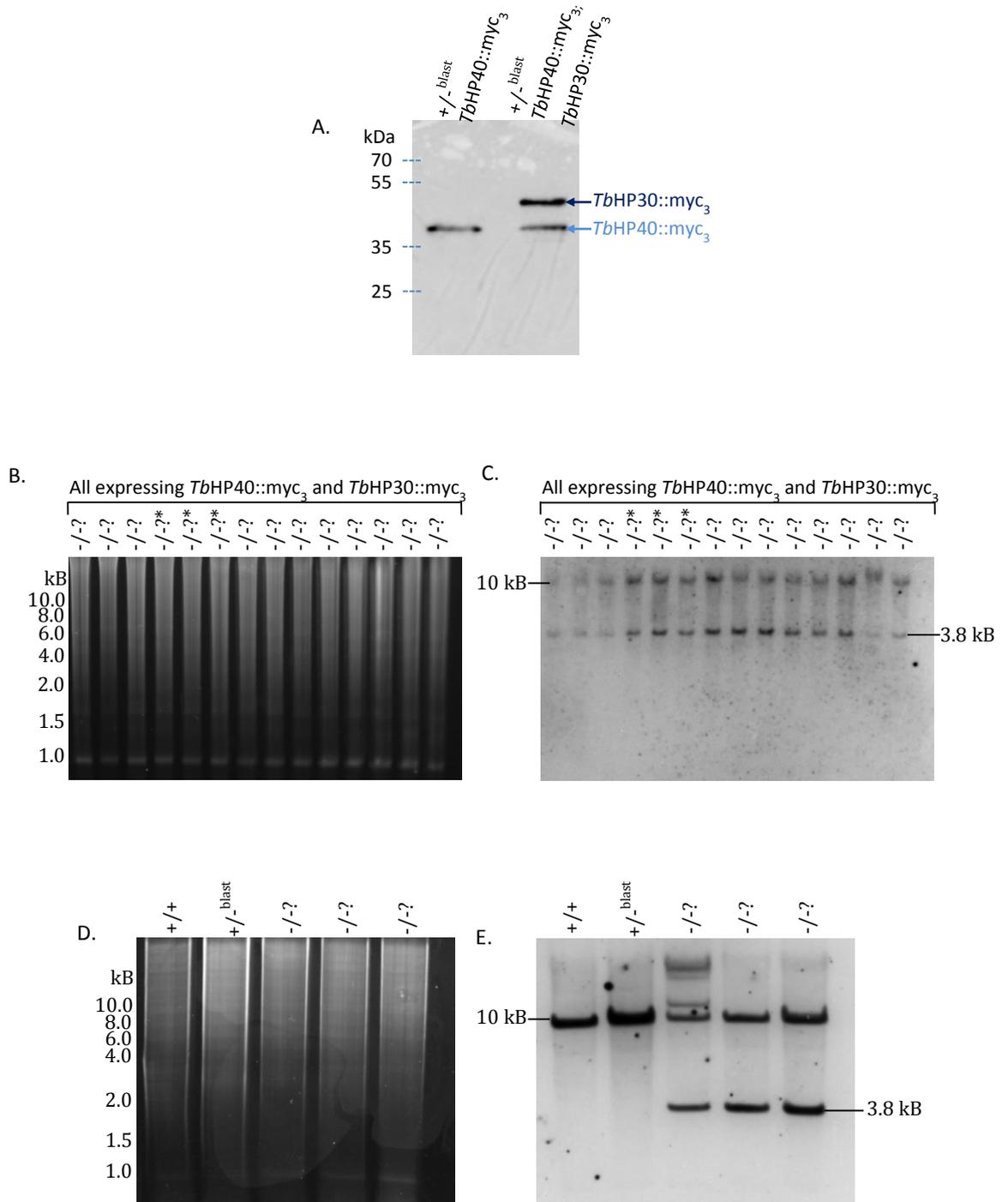


Figure 4.5. Southern analysis of candidate *TbHP40/TbHP30* null (-/-) mutants expressing *TbHP40::myc₃* and *TbHP30::myc₃*

A. The Western blot of whole cell lysates from procyclic 427 *T. brucei* expressing *TbHP40::myc₃* and 4030KO^{+/-blast} cells expressing both *TbHP40::myc₃* and *TbHP30::myc₃*.

B. The post-stained 0.7% TBE agarose gel showing the EcoRI digested gDNA from fourteen of the nineteen candidate *TbHP40/TbHP30* null (-/-) clonal lines, all of which expressed both *TbHP40::myc₃* and *TbHP30::myc₃*.

C. The Southern blot of the gel shown in Figure 4.5 B, hybridised with sequence that corresponded to the open reading frame of *TbHP30* (3OORF). A hybridisation signal which corresponded to the expected size of a wild-type gene locus, (~10 kB), was detected in all of the candidate null clonal lines. The ~3.8 kB band corresponded to the ectopic expression of *TbHP30::myc₃*.

D. The post-stained 0.7% TBE agarose gel showing the EcoRI digested gDNA from wild-type 427 (+/+), the blasticidin S HCl-resistant (+/-^{blast}) population and the three clonal candidate *TbHP40/TbHP30* null (-/-) cell lines that are highlighted with an asterisk in Figure 4.5 C.

E. The Southern blot of the gel shown in Figure 4.5 D, hybridised with the 3OORF probe. The presence of a wild-type gene locus was observed in all of the candidate null (-/-) cell lines.

Figure 4.6

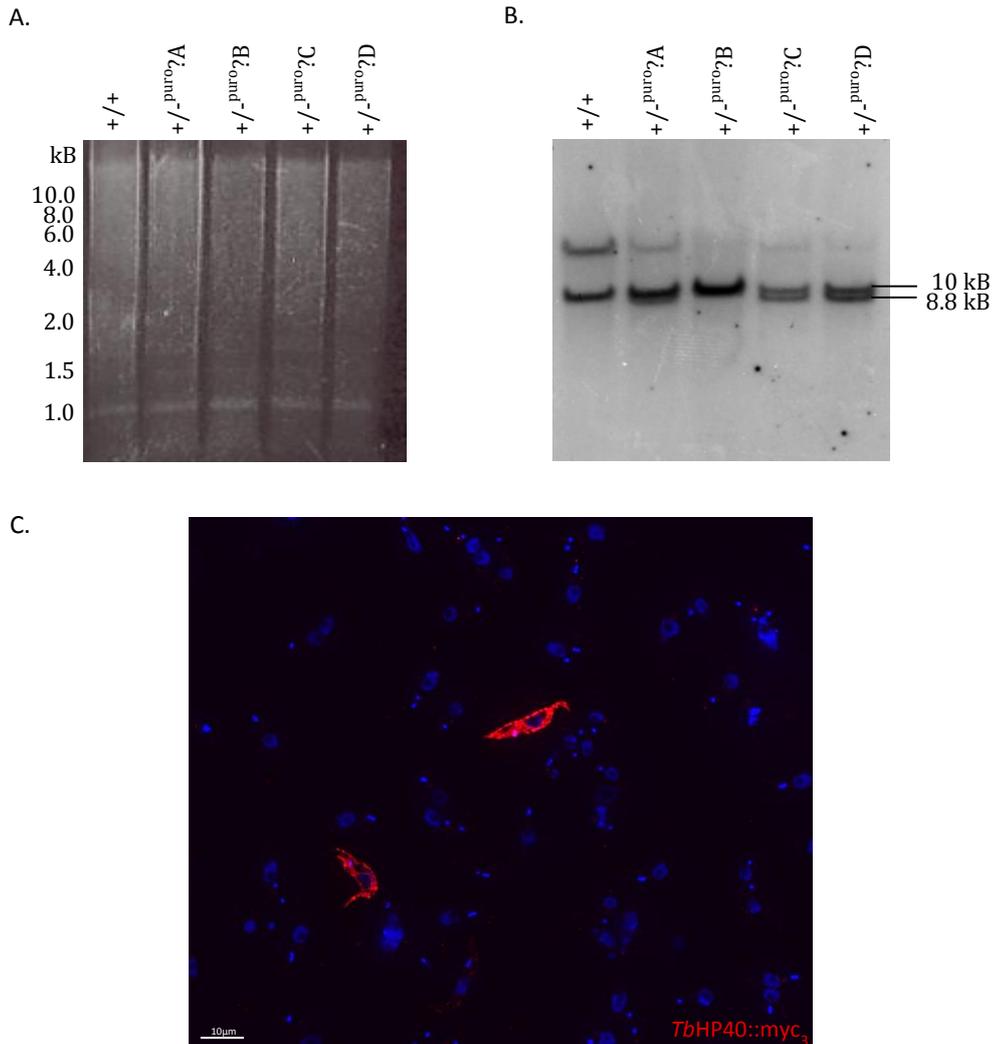


Figure 4.6. Southern analysis of candidate *TbHP40/TbHP30*^{+/-puro} cells and expression of *TbHP40::myc₃* in a 29-13 background of procyclic *T. brucei*.

A. A post-stained 0.7% TBE agarose gel showing the EcoRI digested gDNA from wild-type 427 (+/+), and candidate puromycin-resistant (+/-^{puro}) cells.

B. The Southern blot of the gel shown in Figure 4.6 A, hybridised with the UIR probe. A hybridisation signal which corresponded to the expected size of a wild-type gene locus, (~10 kB), was detected in all lanes. However in two of the puromycin-resistant clonal cells (+/-^{puro} C-D), an additional signal with a size decrease of ~1.8 kB was detected. This indicated that one of the two gene loci had been successfully replaced with the puromycin resistance cassette.

C. Representative indirect immunofluorescence image panel of 4030KO +/-^{puro} C cells expressing *TbHP40::myc₃*, after 24 h incubation with doxycycline. This illustrated that only ~10% of cells constitutively expressed the *TbHP40::myc₃*. The image shows a merge of *TbHP40::myc₃* induced-expression and DAPI-stained DNA in *para*-formaldehyde fixed cells decorated with the anti-myc antibody at a dilution of 1/1000 in blocking buffer B.

Chapter 5

Identifying interacting partners for the novel histidine phosphatases *TbHP40* and *TbHP30*

Identification of the potential interacting partners for a protein of unknown function can be very informative in regard of determining the cellular role(s) of that protein. Several recent examples include the purification of three different super-complexes associated with BRCA1/BARD1 that form in response to acute DNA damage sites in human HeLa cells (Greenberg et al., 2006); the identification of novel deubiquitinating enzymes in the parasitic nematode *Trichinella spiralis* (White et al., 2011); the identification of highly divergent Origin Recognition Complex orthologues in *Trypanosoma brucei* (Tiengwe et al., 2012); the characterisation of novel immunogenic antigens isolated from *Trypanosoma congolense*-infected cattle (Pillay et al., 2013) and the identification and characterisation of a novel class of lysosomal targeting receptors, known as the cysteine protease binding protein family, in *Entamoeba histolytica* (Marumo et al., 2014). However, identifying partners for novel proteins can be challenging when proteins have enzymatic function(s), as opposed to structural or some regulatory roles, since interactions between enzymes and their partner proteins may tend to be transient or easily disrupted during experimental processing procedures (Chien and Gierasch, 2014; Mohammed and Carroll, 2013; Rees et al., 2015; Roux et al., 2012)

The two putative histidine phosphatases *TbHP40* and *TbHP30* are of unknown function. Localisation experiments for *TbHP40* and *TbHP30* indicated both are mitochondrial in procyclic *T. brucei* (see Chapter 3.2). A mitochondrial localisation for *TbHP40* and *TbHP30* was also suggested from organellar proteome inventories generated for procyclic *T. brucei* (Fisk et al., 2013; Niemann et al., 2013; Panigrahi et al., 2009; Urbaniak et al., 2013). However, the function of neither *TbHP40* nor *TbHP30* was known prior to the start of my experiments. Compartmentalisation of *TbHP40* and *TbHP30* within the mitochondrion potentially limits the range of metabolic pathways that these proteins may function in, assuming that the PFAM domain (IPR013078) defining both proteins accurately points to an enzymatic function. In an attempt to further characterise *TbHP40* and *TbHP30*, two techniques to identify potential interacting partners were utilised, and the results from the application of these techniques are discussed in this chapter.

A first approach revolved around the use of 'proximity-dependent biotinylation identification' (known as BioID) which permits the identification of candidate interacting partners or proteins in close proximity to the protein of interest *in vivo* - in my case *TbHP40*. This technique was first reported in conjunction with its application in mammalian cells where it was used to identify multiple proteins that are associated with/in close proximity to an intermediate filament protein that is a component of the inner nuclear membrane (Roux et al., 2012). Subsequently BioID was adapted for use in *T. brucei* and used to reveal novel protein components of the essential golgi-associated 'bi-lobe' (Morriswood et al., 2013). More recently the approach has been applied further in protists, resulting in the discoveries of novel mitochondrial proteins in *Giardia lamblia* and further novel cytoskeletal proteins in *T. brucei* and *Toxoplasma gondii* (Chen et al., 2015; Martincova et al., 2015; McAllaster et al., 2015). BioID exploits the unusually strong affinity between biotin and streptavidin, which has a dissociation constant (K_d) of $\sim 10^{-14}$ M, and is the strongest non-covalent interaction known in nature. Thus, a protein of interest (in my experiments, *TbHP40*) is fused in-frame with a mutant version of biotin ligase from *Escherichia coli* that is 35 kDa in size and is known more widely as BirA*. BirA* behaves in the same manner as the unmodified BirA enzyme in that it forms a reactive intermediate, biotinoyl-5'-AMP, from ATP and biotin. However BirA* has an extremely low affinity for this intermediate and rather than retaining the intermediate within its active site until a lysine residue on a protein substrate is encountered, the intermediate is released thereby leading to indiscriminate biotinylation of any protein in the immediate vicinity (Roux et al., 2012). The hypothetical labelling radius that BirA* is capable of *in vivo* is thought to be ~ 10 nm (Kim et al., 2014). Proteins biotinylated via the BioID approach may be readily affinity purified from cell lysates using streptavidin coated beads and subsequently analysed by mass spectrometry.

The second approach described in this chapter used a more traditional co-immunoprecipitation approach to study potential protein-protein interactions *in vivo*, albeit that the approach was adapted to make use of stable isotope labelling by amino acids in cell culture (SILAC). The urea-based lysis conditions used in the BioID approach are not readily compatible with co-immunoprecipitation experiments because with the latter, stability of the physiological interaction(s) between proteins must be maintained.

5.1 A proximity-dependent biotinylation approach to identifying candidates interacting partners for Tb927.11.2920

A synthetic gene corresponding to full length *TbHP40* minus its stop codon was purchased, thereby negating the internal HindIII site present in the coding sequence of native *TbHP40*. Within the synthetic gene a silent mutation at base-pair 750 removed the internal HindIII site. HindIII and NdeI sites were added upstream and a HindIII site downstream of the *TbHP40* coding sequence in order to facilitate easy in-frame sub-cloning into plasmid pLew100v5b1d-HYG. This plasmid vector contains the mutant BirA* plus a single myc-epitope coded in-frame and upstream of BirA* (Morriswood et al., 2013). Henceforth the fusion of *TbHP40* and BirA* known as *TbHP40::BirA** (although more formally it might have been written *TbHP40::myc::BirA**).

The commercially synthesised *TbHP40* was excised from the standard cloning vector pEX-A in which it was supplied by digestion with HindIII. It was then cloned into the pLew100v5b1d-HYG-myc-BirA* vector that had also been digested with HindIII, and then treated with Shrimp Alkaline Phosphatase to dephosphorylate the linearised DNA and minimise the potential for vector re-ligation. Ligation mixtures were transformed into XL-1 blue competent *E. coli* and plasmid DNA extracted from colonies grown overnight in LB broth using a miniprep kit. Plasmid DNA was screened for the presence of the *TbHP40::BirA** insert in the appropriate in-frame orientation using an XbaI restriction digest as shown in Figure 5.1. A digest in which insert was correctly orientated resulted in visualisation of three DNA bands following treatment with XbaI: a 7666 bp molecule corresponding to vector backbone; a 1908 bp insert corresponding to much of the *TbHP40::BirA** gene fusion; and a 578 bp corresponding to a DNA molecule spanning an XbaI site upstream of the HindIII site in pLew100v5b1d-HYG used for cloning and an XbaI site occurring just before the 5' end of the *TbHP40* coding sequence (Figure 5.1 B). A digest with insert in the wrong orientation also resulted in the detection of three DNA molecules, but this time with sizes of 7666 bp (vector), 1032 bp (which predominantly contained myc-tagged BirA* coding sequence) and a 1454 bp fragment consisting of *TbHP40* coding sequence plus vector sequence (spanning up to the XbaI restriction site at position 2333 bp in the pLew100v5b1d-HYG). Plasmid DNA containing the *TbHP40* insert that restriction mapped correctly was then sent for sequencing prior to transfection into procyclic *Trypanosoma brucei*.

Following the confirmation of plasmid sequence, the *TbHP40::BirA** construct was linearised using NotI then transfected into SmOx-P9 (Poon et al., 2012) and 29-13 (Wirtz et al., 1999)

cell lines (in these cells lines ectopic protein expression is induced by the addition of doxycycline) and into wild-type 427 *T. brucei* (where expression would be constitutive). Despite several independent attempts, stable transformants could only be obtained in the SmOx-P9 cell line. Evidence for expression of the construct was sought by Western blotting and immunofluorescence. The Western blot for the *TbHP40::BirA** SmOx-P9 population when probed with the anti-myc antibody shows detection of a protein ~70 kDa in size in the lane corresponding to whole cell lysate prepared from cells incubated for 24 hours with doxycycline to induce protein expression. The size of the protein detected matches closely the predicted size of the *TbHP40::BirA** fusion (~71 kDa) and importantly this band was also absent from the lane where the whole cell lysate was from cells not incubated with doxycycline (and therefore not induced for *TbHP40::BirA**) (Figure 5.2 A).

The *TbHP40::BirA** and the SmOx-P9 parental cell lines were subsequently grown in SDM-79-SILAC media (Ong et al., 2002; Urbaniak et al., 2012; Urbaniak et al., 2013). The purpose of this was to do an initial trial experiment in which cells were grown in medium and processed as though for mass spectrometry. The SILAC technique provides an *in vivo* approach for labelling all proteins synthesised within a cell that can be readily identified using mass spectrometry. Stable isotopes of particular essential amino acids are added to a media which is deficient of the natural form of those specific amino acids. Therefore the isotopic amino acids are incorporated into all proteins as the cultures grow and divide (Ong et al., 2002). The isotopic labels can be detected by mass spectrometry. Thus, combining SILAC with the BioID approach potentially provided a powerful tool for identifying candidate protein-protein or near-neighbour interactions since the use of the SILAC approach reduces the amount of noise detected by the mass spectrometer, and makes identification of significant results (*i.e.* likely true interacting protein(s)) more obvious when analysing the data.

The SmOx-P9 parental cell line was grown in SDM-79-SILAC containing the light labels arginine (R₀) and lysine (K₀). The *TbHP40::BirA** cell line was grown separately in SDM-79-SILAC containing the medium labels arginine (R₆) and lysine (K₄) and containing heavy labels arginine (R₁₀) and lysine (K₈). The SDM-79-SILAC appeared to have no detrimental effect on *T. brucei* growth (data not shown). All three cell lines were grown in the presence of doxycycline for 24 hours, then biotin and additional doxycycline were added, and the cultures incubated for a further 24 hours. In this way, all cell lines had undergone a minimum of 7 divisions in labelling conditions prior to harvest to ensure that all native forms of the amino acids are replaced with a labelled isotope. The cultures were harvested and soluble

and insoluble protein samples were prepared using PEME 1% NP-40 extraction buffer for analysis by Western blotting (Figure 5.2). Cells were also settled onto coverslips and fixed using *para*-formaldehyde, permeabilised and processed for immunofluorescence microscopy. Western blot membranes probed with streptavidin showed a complexity of biotinylated proteins in the soluble fraction, and in particular two intense signals were detected at ~70 kDa and ~34 kDa (Figure 5.2 lane A). In contrast, when identical membranes were probed with the anti-myc antibody only a single band at ~34 kDa was detected within the soluble proteins (Figure 5.2 lane B). These results suggested that *TbHP40* was a protein found exclusively in the soluble fraction, but proteolytic cleavage of the construct may have occurred as ~34 kDa corresponded with the predicted size of the myc::*BirA** fragment alone. The difference in potential proteolytic sensitivity of *TbHP40*::*BirA** between the analysis shown in Figure 5.2 A and Figure 5.2 B is that in my first analysis the cell pellet was transferred to boiling hot Laemmli buffer, whereas in the experiments for Figure 5.2 B samples were subject to a longer and different processing protocol.

Thus, the detergent-extraction of SILAC-labelled cells induced for *TbHP40*::*BirA** expression in the presence of excess exogenous biotin was repeated, but now with the presence of a cocktail of protease inhibitors (Halt Protease Inhibitor single-use cocktail [ThermoFisher], TLCK, benzamidine and PMSF) in the extraction buffer. Extracts of soluble proteins were again subjected to Western blotting. When the membrane was probed with streptavidin (Figure 5.2 lane D) a dramatic reduction in the complexity of the biotinylated protein profile was observed, although a significant signal was still detected at ~70 kDa. There was no obvious detection of signal at ~34 kDa. A Western blot membrane probed with anti-myc antibody (Figure 5.2 lane E) now detected a band of the correct expected size for *TbHP40*::*BirA** (~70 kDa). The analysis therefore demonstrated the importance of the presence of an anti-protease inhibitor cocktail in the extraction buffer, and this cocktail was used in all subsequent experiments. When coverslips containing fixed, permeabilised cells (induced for *TbHP40*::*BirA** expression) were decorated for indirect immunofluorescence using the anti-myc antibody and imaged using the DeltaVision microscope, a mitochondrial signal for *TbHP40*::*BirA** was observed (Figure 5.3): only ~20% of the population exhibited high-level, inducible *TbHP40*::*BirA** expression. Another ~20-30% of cells exhibited a low accumulation of inducibly expressed *TbHP40*::*BirA**, and the remainder of the cells showed minimal or no expression. When coverslips were decorated with streptavidin conjugated to TRITC, mitochondrial localisation of biotinylated proteins was again evident (Figure 5.4).

Since the *TbHP40Bir** construct could be detected by Western blot and fluorescence microscopy it was decided to proceed to MS analyses despite variable expression of *TbHP40::BirA** within my cell line. Even though the variable expression was a source of problems (~10⁹ cells were considered necessary for processing for analysis by MS), because the cells that expressed *TbHP40::BirA** appeared to express the protein at relatively high abundance, it was tentatively anticipated that sufficient protein would be present in the samples sent for MS analysis. Moreover, another member of the group who was also applying the BioID technique was working with proteins that at endogenous levels were expressed in relatively low abundance within a cell, even though 80-100% of her cells expressed *BirA** fusion proteins when gene expression was induced by the addition of doxycycline. So even though only a low percentage of my cells expressed *TbHP40::BirA** at varying levels when induced with doxycycline, a relatively high overall abundance of *TbHP40::BirA** potentially compensated for not having a homogenous level of protein expression across a majority of cells in a population.

Thus, the *TbHP40Bir** and *SmOx-P9* cell lines were grown again in SDM-79-SILAC media, harvested, and three pellets (*SmOx-P9-Light*, *TbHP40::BirA*-Medium* and *TbHP40::BirA*-Heavy*) mixed, before lysing the combined cell pellets with the PEME 1% NP-40 extraction buffer in the presence of the cocktail of inhibitors. The cell lysate was then solubilised in 8M urea 1% SDS solubilisation buffer and incubated with streptavidin-immobilised magnetic beads (Pierce™). The beads were next incubated in the dark with solubilisation buffer containing 50mM iodoacetamide, washed with solubilisation buffer, and then digested with trypsin overnight. Due to the presence of SDS and a high urea concentration in the trypsin-digested peptides, the peptides required purification before analysis by mass spectrometry. Purification was done by filter-aided sample preparation using a C18 microspin column (Wisniewski et al., 2011; Wisniewski et al., 2009). Peptide samples were finally lyophilised and sent for analysis by liquid chromatography tandem mass spectrometry (LC-MS/MS) (FingerPrints Proteomic facility, Dundee University). Data obtained was processed using MaxQuant (Cox and Mann, 2008) and peptides identified from the *T. brucei brucei* 927 protein database (Version 25, available from TriTrypDB) and common known contaminants. The output from MaxQuant was analysed using the Perseus framework as shown in Figures 5.5-5.8.

For a classical mass spectrometry analysis, any proteins of likely interest would be identified according the relative intensity (the y-axis in Figures 5.5-5.6, points above the 7.5 mark of

the intensity (Log¹⁰ scale). However, the incorporation of the SILAC into the experiment reduced the amount of noise in the results because the mass spectrometer can provide quantitative results in terms of protein abundance and can distinguish between the light-labelled parental line (SmOx-P9) and the the transgenic medium- and heavy-labelled *TbHP40::BirA** cell line even when the cells are mixed. This is because of the mass difference introduced by the labelled amino acids. Applying SILAC to the experiments provided an unbiased method to identify interacting partners of the bait protein (*TbHP40::BirA**), since non-specific proteins binding to the bait protein (grown in the presence of ‘medium’- and ‘heavy’- labelled amino acids) will do so equally as well with the protein in the parental control cells (grown in the presence of ‘light’-labelled amino acids), resulting in 1:1 ratios. However, if specific reactions occurred between true interacting partners of the bait protein, this would result in different ratios which is then detected by the mass spectrometer (Ong and Mann, 2005). In this way, molecules of potential interest as neighbouring or interacting proteins lie to the right of the zero line on the x-axis and are enriched more than 2-fold. When comparing the medium-labelled *TbHP40::BirA** cell line versus the light-labelled SmOx-P9 cell line, two proteins are enriched above the cut off threshold (Figure 5.5), the darker blue circle corresponds to two peptides that map to *TbHP40* (Figure 5.8 A-B), showing the *TbHP40::BirA** biotinylated itself, and another identification coloured in lighter blue corresponded to a single peptide, which mapped to three identical proteins: two proteins encoded by genes annotated as ‘KMP-11’ and another an ‘unspecified product’ (Figure 5.8C-D). This pattern was not reflected in the heavy-labelled *TbHP40::BirA** versus light-labelled SmOx-P9 (Figure 5.6) which acted as a built-in biological repeat within the system. Here *TbHP40* was the only enriched protein located above the cut off threshold, and KMP-11 was located back in the region categorised as noise. When comparing ‘heavy-labelled versus light-labelled’ against ‘medium-labelled versus light-labelled’ (Figure 5.7), it was clear that the only greatly enriched protein from both experiments is *TbHP40*, and KMP-11 was located near the zero line from the y-axis, so was most likely an anomalous result. Unfortunately, no other proteins were identified as potential interacting partners with *TbHP40* – only *TbHP40* itself was biotinylated, enriched and identified by LC-MS/MS. This was perhaps somewhat disappointing as the BiID approach applied elsewhere in the group was successful in identifying a small list of candidate-interacting proteins for the *T. brucei* cytoskeletal protein RP2.

5.2 Identifying potential interacting partners of TbHP40 using a co-immunoprecipitation approach

An alternative strategy to the BioID/SILAC approach was devised using small C-terminal epitope tags added to *TbHP40* and *TbHP30*. Initially, the aim was to determine whether *TbHP40* and *TbHP30* were likely to interact with one another in a multimeric protein complex (*e.g.* did they form hetero-dimers or hetero-tetramers), but the aim was subsequently broadened by the application of SILAC to see whether additional proteins might be immunoprecipitated from procyclic *T. brucei* engineered to express *TbHP40* with a C-terminal triple-myc tag (*TbHP40::myc₃*).

In order to express *TbHP40* and *TbHP30* with different C-terminal epitope tags some cloning was first required. Thus, the full length coding sequence for *TbHP40* with a C-terminal triple-myc tag (*TbHP40::myc₃*) previously used in the localisation experiments described in Section 3.2 was excised from its pDex377-based vector by digestion using the restriction enzymes BamHI and NsiI, and sub-cloned into pDex477-Y2 that had also been digested with BamHI and NsiI. pDex477-Y2 (Appendix II Figure B.3 B) contains a gene conferring resistance to phleomycin, and transfer of *TbHP40::myc₃* to this plasmid allowed expression of epitope-tagged *TbHP30* (with a non-myc tag; see below) from a pDex377-based construct in which stable transformants are selected by their resistance to hygromycin. Ligation mixtures containing BamHI-NsiI-digested pDex477-Y2 and *TbHP40::myc₃* insert were transformed into XL-1 blue competent *E. coli* and plasmid DNA prepared as mini-preps from clonal overnight cultures in the usual way; BamHI and NsiI restriction digests were used to screen for the presence of putative *TbHP40::myc₃*-containing insert and a plasmid predicted to contain correct insert was sent for DNA sequencing. Once verification of the pDex477-Y2-*TbHP40_{myc}* was confirmed by sequencing, plasmid DNA was linearised by digestion with NotI, and then transfected into procyclic 427 *T. brucei*. Expression of *TbHP40::myc₃* in phleomycin-resistant stable transformants was confirmed by indirect immunofluorescence and Western blotting (data not shown).

For expression of *TbHP30* with an alternative epitope tag, an epitope described by the amino acid sequence G-A-F-S-I-N-P-A-M was used. The GAFSINPAM epitope occurs in the human protein TDP-43 and is recognised by a commercially available monoclonal antibody 'anti-TDP-43 clone DB9' (anti-DB9) (Merck, Millipore). Both the definition of the epitope and its suitability for use epitope-tagging of trypanosome proteins – *i.e.* anti-DB9 does not recognise any *T. brucei* protein by immunofluorescence or Western blot – was shown by Dr Fiona

Benson and an undergraduate dissertation student. In the course of that work the GAFSINPAM epitope was sub-cloned into pDex377 such that excision of *TbHP30* coding sequence from pDex377-*TbHP30*_{myc} with BglIII and XhoI and ligation with BglIII-XhoI digested pDex377_{GAFSINPAM} would facilitate expression of *TbHP30* with a C-terminal GAFSINPAM epitope, following stable transformation of *T. brucei*. Plasmid DNA for pDex377-*TbHP30*_{GAFSINPAM} was thus isolated, screened for the presence of *TbHP30*-containing insert by restriction mapping with BglIII and XhoI, and a plasmid with appropriate insert confirmed by DNA sequencing. Subsequent transfection of procyclic 427 *T. brucei* expressing *TbHP40*::myc₃ with NotI-digested pDex377-*TbHP30*_{GAFSINPAM} followed by selection with hygromycin allowed the generation of procyclic cells constitutively expressing both *TbHP40*::myc₃ and *TbHP30*::GAFSINPAM.

Evidence for co-expression of *TbHP40*::myc₃ and *TbHP30*::GAFSINPAM was sought by indirect immunofluorescence and Western blotting. Coverslips containing *para*-formaldehyde-fixed, permeabilised cells were decorated with anti-myc; ~99% of cells imaged expressed *TbHP40*::myc₃ construct, although expression levels were variable between cells (Figure 5.9 A-C). Western blots of the whole cell lysate probed with anti-myc antibody detected a single protein of the predicted molecular size for *TbHP40*::myc₃ (39 kDa) (Figure 5.9 D, lane 1). Similarly, when coverslips containing *para*-formaldehyde-fixed, permeabilised cells were decorated with anti-DB9, ~99% of cells were seen to express *TbHP30*::GAFSINPAM although as with *TbHP40*::myc₃ expression, some cells contained less epitope-tagged protein than others (Figure 5.9 F-G). The same whole cell lysate as used in Figure 5.9 D were used in an independent Western blot that was decorated with anti-DB9, and a single protein of the size expected for *TbHP30*::GAFSINPAM (44 kDa) was detected (Figure 5.9 H). Henceforth, the cell line co-expressing *TbHP30*::GAFSINPAM and *TbHP40*::myc₃ is referred as 40_{MYC}/30_{GSP} and was used in all subsequent co-immunoprecipitation experiments.

All co-immunoprecipitation experiments were performed using the Pierce® Co-immunoprecipitation (Co-IP) Kit (Thermo Scientific). Since anti-myc antibody was most readily available to me and, to date, has been used more widely in a variety of applications, the myc antibody was chosen for immobilization to the AminoLink Plus Coupling Resin. Thus, myc-tagged *TbHP40* was immunoprecipitated, together with any associated protein complexes. Initially, the 40_{MYC}/30_{GSP} cell line was grown in SDM-79, harvested and cells lysed on ice in IP lysis/Wash buffer with a cocktail of inhibitors present (see Section 2.33). Following sample processing and protein elution from resin coupled to anti-myc antibody,

immunoprecipitated proteins were analysed by silver staining of SDS-PAGE gels and subjected to Western blotting using both anti-myc antibody and anti-DB9.

As anticipated, the Western blot probed with the anti-myc antibody detected protein of the expected size (~39 kDa) for *TbHP40::myc₃* in lanes containing 'eluted protein' and 'flow-through' (lanes 1 and 6-8 in Figure 5.10 A). The faint detection of *TbHP40::myc₃* in lane 1 likely corresponds to a small amount of *TbHP40::myc₃* which did not bind to the myc-tagged resin and remained in the flow-through. There was no detection of protein in lanes 2-5, suggesting no *TbHP40::myc₃* was eluted during the wash stages, but easy detection of *TbHP40::myc₃* in lanes 6-8, which are the elution steps, indicated *TbHP40::myc₃* was successfully eluted from the resin. In contrast, Western blots probed with the anti-DB9 were completely blank, with no protein even being detected in the flow-through or the washes (Figure 5.10 B). This was unexpected, but the same result was seen from three independent attempts at the immunoprecipitation.

To look more holistically for evidence of immunoprecipitation of multi-protein complexes containing *TbHP40::myc₃*, processed immunoprecipitates were analysed by silver staining of SDS-PAGE gels. Silver staining with silver nitrate was used rather than Coomassie blue staining because of the high sensitivity of the former and its ability to detect protein levels as little as <1 ng (Weiss et al., 2009). Silver stained SDS-PAGE gels showed a variety of different proteins were detectable at each step of co-immunoprecipitation experiments. For example from the gel shown in Figure 5.10 D Lane 1, which corresponds to lysate from 40_{MYC}/30_{GSP} cells mixed with IP Lysis/Wash buffer, looks almost identical to lane 2, which is the same sample but after 1 hour incubation with the myc-immobilised resin. Lane 3 shows the 'flow-through' from 40_{MYC}/30_{GSP} lysate after 4 hours incubation with the myc-immobilised resin. Lanes 4-5 show proteins from the 1st and 3rd washes of the resin, these lanes showed very different complexities to one another and when compared with the proteins detected in lane 3. Since the sample in lane 5 contained little in the way of protein (compared to lane 4) it indicated the wash steps were efficient. Lanes 6-8 show sample from three sequential elutions of protein from the resin-immobilized anti-myc antibody. Again, a different complexity to the proteins detected was noted when compared to the 1st and 3rd 'washes' (lanes 4-5). In lanes 6-7, two relatively intense bands were detected above the 35 kDa marker; these could correspond to proteins of the expected size of *TbHP40::myc₃* and *TbHP30::GAFSINPAM*. Finally, since very little protein is detected in lane 8, it appeared that the elution steps were efficient at removing proteins bound to the resin. A further

independent immunoprecipitate gave similar results when subjected to the same silver stain analysis outlined above (data not shown).

To look further at why there was no evidence for immunoprecipitation of *TbHP30::GAFSINPAM*, a troubleshooting experiment was carried out to analyse the stability of the GAFSINPAM epitope in the IP Lysis/Wash buffer solution. The Western blot in Figure 5.9 H showed that the GAFSINPAM epitope was detectable in the IP Lysis/Wash buffer immediately after lysis of the 40_{MYC}/30_{GSP} cell line. Here, cells were incubated on ice for no longer than 15 minutes before the protein samples were boiled at 100°C in non-reducing 'Lane Marker Sample Buffer' (Thermo Scientific) prior to analysis by SDS-PAGE and Western blotting. However, in the co-immunoprecipitation experiments, protein containing the GAFSINPAM epitope was incubated in the IP Lysis/Wash buffer for a much longer time. Thus, the 40_{MYC}/30_{GSP} cell line was grown in SDM-79, harvested and cells lysed on ice in the IP Lysis/Wash buffer containing the cocktail of inhibitors described in Section 2.33. This bait:prey protein mixture in the IP Lysis/Wash buffer was then incubated with the myc-immobilised resin on a column overnight at 4°C on a rotating wheel, with protein samples being taken immediately (zero hours), then at 1 hour intervals for 5 hours, and a final sample taken after 16 hours incubation. All samples collected were boiled immediately in non-reducing Lane Marker Sample Buffer and frozen at -80°C before analysis by Western blot. The Western blot probed with the anti-DB9 antibody showed very faint detection of a protein corresponding to the predicted size of *TbHP30::GAFSINPAM* in lanes 1 to 6, which corresponded with the 0-5 hour time points (Figure 5.10 C, the bands in question are indicated with a red arrow). Sometime between 5 and 16 hours this signal was lost although the detection of the GAFSINPAM is poor even at the zero hour time point. A positive control in lane 9 (recombinant TDP-43) confirmed the continued activity of anti-DB9 (Figure 5.10 C, black arrow). 40_{MYC}/30_{GSP} cells were also fixed and processed for immunofluorescence confirming the continued expression of both *TbHP40::myc₃* and *TbHP30::GAFSINPAM* (data not shown).

Since there appeared to be issues with the stability and detection of the GAFSINPAM epitope, I now took a slightly different approach to the co-immunoprecipitation experiment. This approach incorporated the SILAC and mass spectrometry procedures used in the BioID experiment. Thus, 40_{MYC}/30_{GSP} and the 427 parental cell lines were grown in SDM-79-SILAC media. The 427 parental cell line was grown in SDM-79-SILAC containing the light labels arginine (R₀) and lysine (K₀). 40_{MYC}/30_{GSP} cells were grown independently in both SDM-79-SILAC containing the medium labels arginine (R₆) and lysine (K₄) and SDM-79-SILAC

containing heavy labels arginine (R₁₀) and lysine (K₈). All cultures were allowed to undergo at least 7 divisions in the SILAC media containing light, medium or heavy labels prior to harvesting. The three cultures were harvested separately and lysed using only filtered ice-cold water in the presence of the cocktail of inhibitors (Section 2.34). This lysis approach was used rather than the IP Lysis/Wash buffer because the latter contains EDTA and NP-40, which can interfere with downstream processes when preparing samples for mass spectrometry. Complete lysis of the cells was checked visually using a light microscope before freezing the intact lysate at -80°C.

Again, anti-myc antibody was immobilised to the 'AminoLink Plus Coupling Resin' although this time in triplicate, since the light-, medium- and heavy-labelled cell lines needed to be kept separate from each other during the co-immunoprecipitation steps. The three lysed samples were defrosted on ice and then the-still-separate lysates incubated with the 'Control Agarose Resin' at 4°C for one hour. This pre-clearance step removed any non-specific proteins binding to the resin. Lysates were then incubated with the myc-immobilised resin in separate columns at 4°C for 2 hours. The flow-through from the three columns was removed, and each column washed 3 times with IP Lysis/Wash buffer and the resin from each column transferred to separate LoBind Eppendorfs. The resin was then boiled for 5 minutes at 95°C in Laemlii buffer in order to break any covalent bonds with the immobilised myc antibody. The protein samples in Laemlii buffer were then subject to SDS-PAGE electrophoresis, stained with Instant Blue and washed with MilliQ water. Lanes from the gel were then chopped up, mixing the light-, medium- and heavy-labelled samples and preparing the combined sample for digestion with trypsin as described in Section 2.35.2. The digested peptide samples were lyophilised and sent for analysis by liquid chromatography tandem mass spectrometry. The data obtained was processed using MaxQuant (Cox and Mann, 2008) and peptides identified with the *T. brucei brucei* 927 protein database (Version 25, available from TriTrypDB) and 'common known contaminants'. The output from MaxQuant was analysed using the Perseus framework and the data summarised in Figures 5.11-5.13.

From the mass spectrometry results, both the 'medium-labelled' 40_{MYC}/30_{GSP} cell line versus the 'light-labelled' 427 parental cell line and the 'heavy-labelled' 40_{MYC}/30_{GSP} cell line versus the 'light labelled' 427 parental cell line returned almost identical results. Two proteins of high intensity with respect to peptide number detected in the mass spectrometer and enriched more than 4-fold were identified: *TbHP40* and *TbHP30*. No other 'medium'- or 'heavy'-labelled proteins were enriched greater than 1.5-fold. Thus, all other proteins identified in the sample were categorised as background noise (Figures 5.11 and 5.12). This

result is even clearer in Figure 5.13 where the 'medium-labelled' 40_{MYC}/30_{GSP} against 'heavy-labelled' 40_{MYC}/30_{GSP} lysates are compared since these are intrinsic repeats built into the same analysis. Both *TbHP40* and *TbHP30* lie to the far top right corner of the graph, and are the only two proteins that are greatly enriched. Confidence in the identification of *TbHP40* and *TbHP30* is supported by the data shown in Table 5.1 (*TbHP40*) and Table 5.2 (*TbHP30*) which indicate the peptides identified by LC-MS/MS. For *TbHP40* 42 peptides, 22 of which were unique, were matched to the amino acid sequence for *TbHP40* and covered 69% of the predicted amino acid sequence (Table 5.1 B, identified peptides shown in red). For *TbHP30* 14 peptides, 11 of which were unique, were matched to the amino acid sequence for *TbHP30*, covering 35% of the predicted amino acid sequence (Table 5.2 B).

5.3 Summary of attempts to identify interacting partners of *TbHP40* and *TbHP30*

Using the co-immunoprecipitation approach combined with SILAC and mass spectrometry provided the first evidence to indicate that *TbHP40* and *TbHP30* interact with each other *in vivo*. It is interesting that *TbHP30* (my original prey protein when attempting to identify only *TbHP30::GAFSINPAM* by immunoprecipitation) was the only protein apart from *TbHP40* (the bait protein) that was greatly enriched by SILAC and 'pulled down' using the co-immunoprecipitation steps. It suggests that *TbHP40* and *TbHP30* are not likely to interact as part of a larger complex with other proteins within a cell, and probably form a small complex with one-another.

It can be reasoned that the problems encountered with the 'classical' co-immunoprecipitation approach are likely due to proteolytic cleavage, despite the presence of a cocktail of inhibitors. Both myc and GAFSINPAM are small epitope tags, but the myc epitope is repeated three times, whereas GAFSINPAM is just a single epitope. Thus, the loss of the GAFSINPAM epitope signal on Western blots probed with anti-DB9 was realistically due to proteolysis within the GAFSINPAM epitope: work from the 'Benson Lab' suggests that only one amino acid needs to be lost from the GAFSINPAM epitope before the antibody has difficulties recognising its target (FE Benson, personal communication). Thus, the mass spectrometry analysis provided a useful alternative approach to detecting an interaction between *TbHP40* and *TbHP30*. Identification of *TbHP40* and *TbHP30* using LC-MS/MS is of high confidence due to the number of peptides identified for each one. This was a much better and reliable result than that seen in the LC-MS/MS BioID experiments where only 2 peptides relating to *TbHP40* were identified. The peptide coverage of *TbHP40* identified by

the mass spectrometer was also quite extensive, and showed that the N-terminal mitochondrial leader sequence of *TbHP40* must be very short (6 amino acids in length). This is not unprecedented in trypanosomes, and some nuclear encoded mitochondrial proteins are known to have cleavable N-terminal signal peptides of only 7 to 9 amino acids long, this is very short compared to cleavable N-terminal signal peptides found in yeast, which are ~20 amino acids or more in length (Tasker et al., 2001). Other proteins, such as cytochrome oxidase VI, again have very short N-terminal signal peptides but these are not cleaved after import into the *T. brucei* mitochondrion (Tasker et al., 2001). No conclusions can be made regarding the N-terminal targeting sequence for *TbHP30*, because there was no peptide coverage for the first half of the sequence. As discussed previously (Section 3.2), *TbHP30* had been identified in various mitochondrial proteomic studies (Acestor et al., 2009; Fisk et al., 2013; Niemann et al., 2013; Panigrahi et al., 2009; Urbaniak et al., 2013), and again none of the peptides identified in those studies mapped to the first half of the sequence. That might suggest that the start methionine had been incorrectly called, but when C-terminally myc-tagged protein was expressed in procyclic *T. brucei*, protein of the expected size (~43 kDa) from the start methionine predicted in EUPathDB was always detected in Western blot analyses.

The BioID approach may not have been suitable simply because the ~34 kDa C-terminal BirA* tag was too big and caused disruption of the natural function of *TbHP40*. Such a large tag could be causing steric hindrance and preventing *i.e.* *TbHP30* and other potential proteins from interacting, therefore they are not biotinylated and not detected in downstream procedures. The mass spectrometry data obtained did suggest that the BirA* system was functioning correctly in *T. brucei*, since *TbHP40* biotinylated itself and was identified as the only protein to be greatly enriched and abundant in two duplicate experiments. Further evidence supporting the theory that the BirA* tagged-*TbHP40* was detrimental to *T. brucei* may come from the fact that stable transformants were never obtained in the procyclic *T. brucei* cell lines 29-13 and 427. Although when the *TbHP40::BirA** SmOx-P9 cell line was induced using doxycycline, no obvious phenotype was observed and there was no effect on the rate of growth of the cultures. Numerous attempts were made to obtain a bloodstream *T. brucei* cell line expressing *TbHP40::BirA**, but none of these were successful.

Figure 5.1

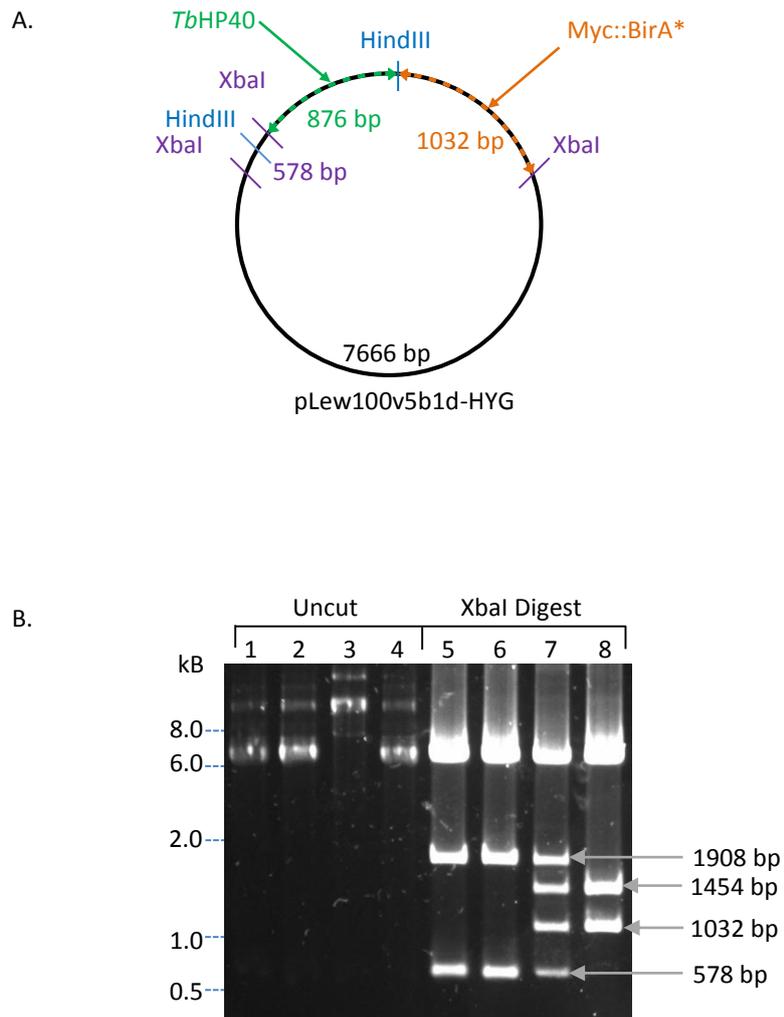


Figure 5.1. Generation of a *TbHP40BirA*::myc* gene fusion with the pLew100v5b1d-HYG plasmid.

A. A cartoon vector map showing the cloning sites present in the pLew100v5b1d-HYG vector which contains the mutant BirA* and an upstream single myc-epitope (highlighted with the orange dashed line). The commercially synthesised *TbHP40* (highlighted with the green dashed line) was digested using HindIII and then sub-cloned into the HindIII-digested vector (positions of the HindIII sites shown in blue). The positions of the XbaI sites are also shown (purple), which were used for subsequent restriction mapping.

B. 1% agarose gel electrophoresis showing the restriction mapping of the *TbHP40BirA*::myc* plasmid. Lanes 1-4: Undigested miniprep DNA. Lanes 5-6: XbaI digested *TbHP40BirA*::myc* plasmids where the *TbHP40* construct was integrated into the myc::BirA* pLew100v5b1d-HYG vector in the correct orientation. Lane 7: Miniprep DNA containing multiple colonies, which show the two possible digest outcomes. Lane 8: XbaI digested *TbHP40BirA*::myc* plasmid that showed the incorrect orientation.

Figure 5.2

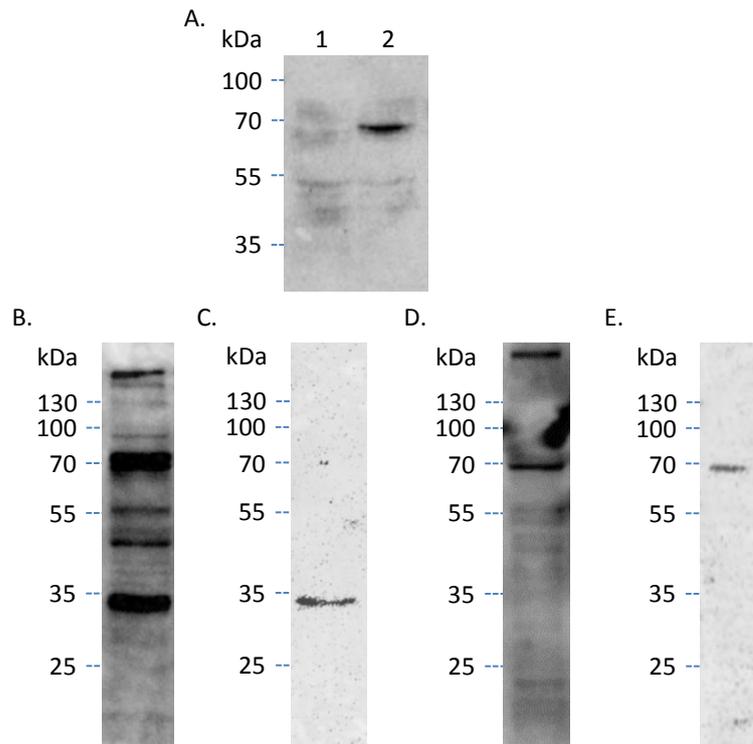


Figure 5.2. Expression of *TbHP40BirA*::myc* in procyclic SmOx-P9 *T. brucei*.

A. Western blot detection of *TbHP40BirA*::myc* expressed in *T. brucei*. Whole cell lysates were prepared in boiling hot Laemlii buffer. Lane 1 corresponds to the lysate from an uninduced population and shows no expression of myc-tagged *TbHP40BirA**. Lane 2 corresponds to the lysate from an induced population, incubated with doxycycline in order to induce the expression of *TbHP40BirA*::myc*. A band which corresponds to the expected size of *TbHP40BirA*::myc* (71 kDa) is detected in this lane. The Western blot was decorated with the anti-myc antibody at a dilution of 1/1000 in blocking buffer A.

B-E. Western blot detection of the expression from *TbHP40BirA*::myc* in the presence of excess external biotin (at a final concentration of 50 μ M).

B: Western blot detection of the soluble fraction from *TbHP40BirA*::myc* lysed cells (extracted using PEME 1% NP-40) which had been induced with doxycycline for 48 hours and incubated with excess biotin for 24 hours prior to cell lysis. The Western blot was decorated with streptavidin at a dilution of 1/10000 in blocking buffer B.

C: Western blot detection of the soluble fraction from *TbHP40BirA*::myc* lysed cells (as described in 5.2 B). The detection of a single band at 35 kDa suggests that *TbHP40BirA*::myc* was a target for proteolysis. The Western blot was decorated with the anti-myc antibody at a dilution of 1/1000 in blocking buffer A.

D: Western blot detection of the soluble fraction from *TbHP40BirA*::myc* lysed cells (extracted using PEME 1% NP-40) which had been induced with doxycycline for 48 hours and incubated with excess biotin for 24 hours prior to cell lysis. Here, lysis was carried out in the presence of an inhibitor cocktail. The lysate was then subjected to Western blot analysis and the Western blot was decorated with streptavidin at a dilution of 1/10000 in blocking buffer B.

E: Western blot detection of the soluble fraction from *TbHP40BirA*::myc* lysed cells (as described in 5.2 D), that were lysed in the presence of an inhibitor cocktail. This shows the inhibitor cocktail provided protection against proteolytic cleavage of *TbHP40BirA*::myc*. The Western blot was decorated with streptavidin at a dilution of 1/10000 in blocking buffer B.

Figure 5.3

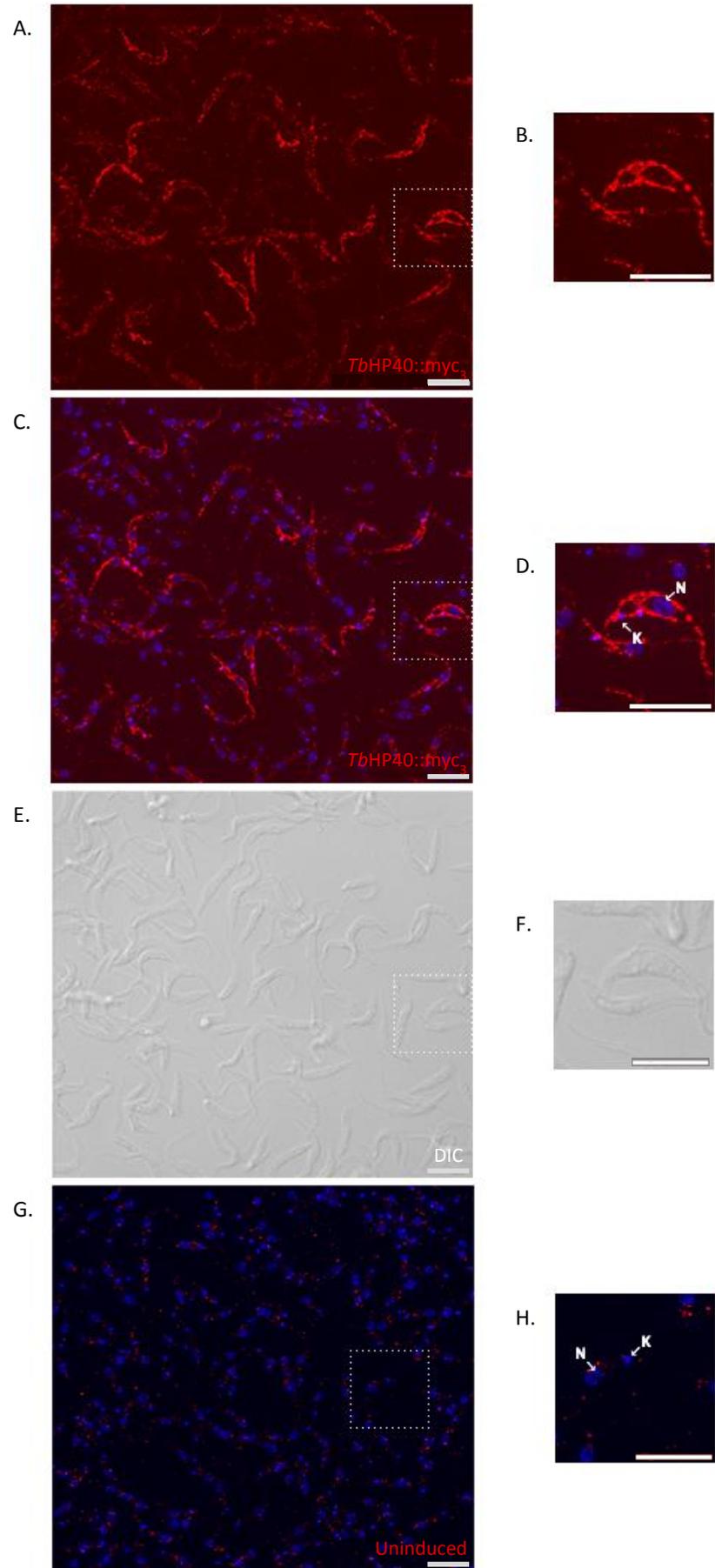


Figure 5.3. Expression of *TbHP40BirA*::myc* in procyclic SmOx-P9 *T. brucei*.

A-F. *TbHP40BirA*::myc* localises to the branched mitochondrion of procyclic *T. brucei*. Cells were incubated with doxycycline 24 hours prior to settling on glass coverslips and fixed with *para*-formaldehyde in preparation for indirect immunofluorescence. A-D: *TbHP40BirA*::myc* fixed permeabilised cells were decorated with anti-myc antibody used at a concentration of 1/1000 (in blocking buffer B), merged with DAPI. E-F: DIC. N, nucleus; K, kinetoplast; scale bars, 10 μ m.

G-H. Smox-P9 *T. brucei* not induced for expression of *TbHP40BirA*::myc*, *i.e.* grown without the presence of doxycycline for comparison against the induced cells were settled onto coverslips for indirect immunofluorescence. Here, fixed and permeabilised cells were decorated with the anti-myc antibody used at a concentration of 1/1000 (in blocking buffer B), merged with DAPI. N, nucleus; K, kinetoplast; scale bars, 10 μ m.

Figure 5.4. Localisation of biotinylated proteins in procyclic SmOx-P9 *T. brucei*, following expression of *TbHP40BirA*::myc* using doxycycline.

A-F. *TbHP40BirA*::myc* cells (post-48 hour incubation with doxycycline and post-24 hour incubation with excess biotin [at a final concentration of 50 μ M]) were settled onto coverslips, fixed, permeabilised and decorated using streptavidin conjugated to TRITC. A-D: Induced *TbHP40BirA*::myc* cells decorated with streptavidin conjugated to TRITC (at a concentration of 1/200 in blocking buffer B), merged with DAPI. E-F: DIC. N, nucleus; K, kinetoplast; scale bars, 10 μ m.

G-H. *T. brucei* SmOx-P9 cells not induced for expression of *TbHP40BirA*::myc*, *i.e.* grown without the presence of doxycycline, but still cultured in the presence of excess biotin, were settled onto coverslips. The slides were then fixed, permeabilised and decorated with streptavidin conjugated to TRITC (at a concentration of 1/200 in blocking buffer B), merged with DAPI. N, nucleus; K, kinetoplast; scale bars, 10 μ m.

Figure 5.4

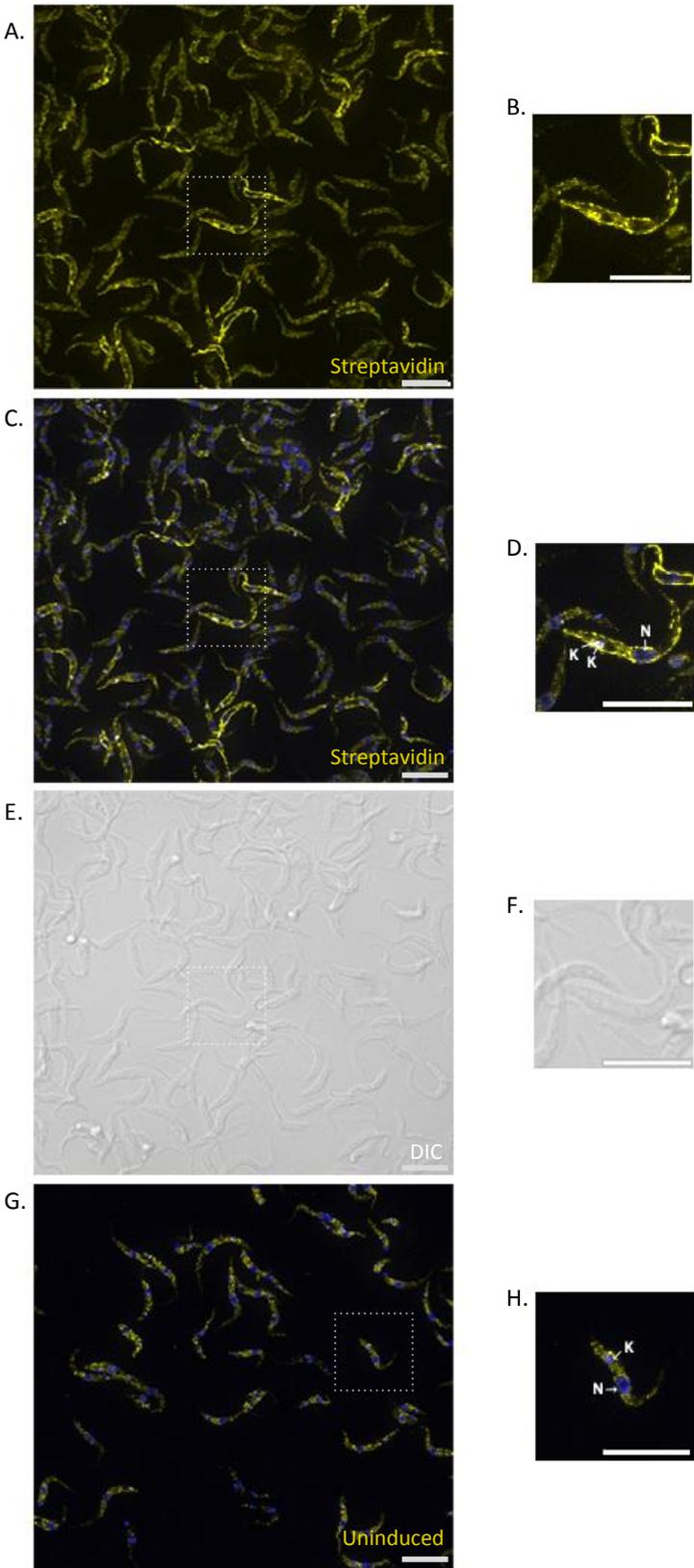


Figure 5.5

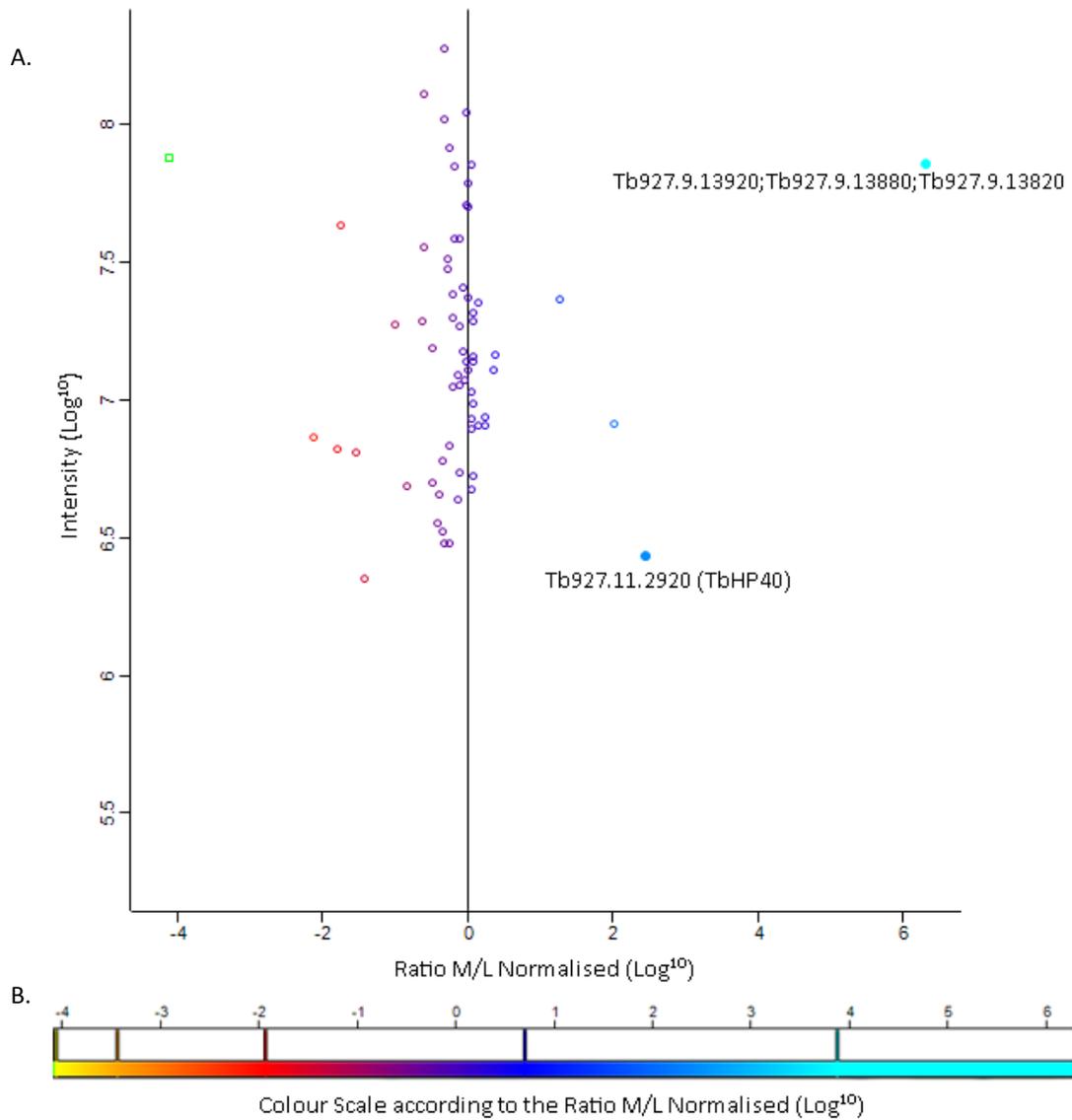


Figure 5.5. The LC-MS/MS scatter plot showing the proteins identified in the SILAC-enriched 'medium'-labelled *TbHP40BirA** cell line ratio versus proteins identified in the 'light'-labelled SmOx-P9 parental cell line plotted with respect to intensity.

The x-axis shows the ratio of 'medium'-labelled proteins versus 'light'-labelled proteins, identified according to the peptides detected by mass spectrometry. The y-axis shows the relative abundance of the proteins. Scales are logged so data is normally distributed, and the ratios normalised because the base peak originally lay to the left of zero. Colours are assigned according to the levels of SILAC enrichment according to the colour scale shown in B.

Figure 5.6

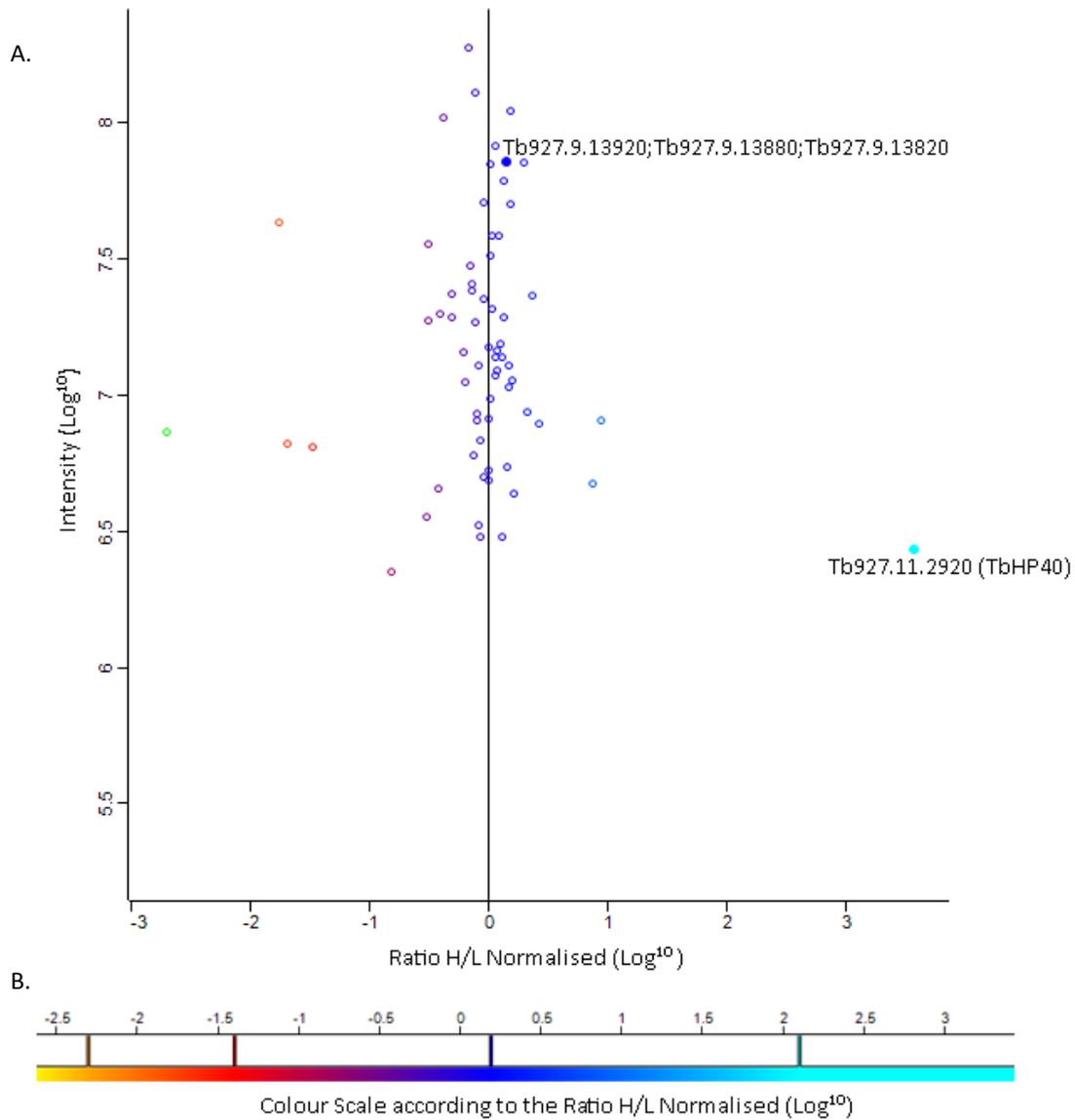


Figure 5.6. The LC-MS/MS scatter plot showing the proteins identified in the SILAC-enriched 'heavy'-labelled *TbHP40BirA** cell line ratio versus proteins identified in the 'light'-labelled SmOx-P9 parental cell line plotted with respect to intensity.

The x-axis shows the ratio of 'heavy'-labelled proteins versus 'light'-labelled proteins, identified according to the peptides detected by mass spectrometry. The y-axis shows the relative abundance of the proteins. Scales are logged so data is normally distributed, and the ratios normalised because the base peak originally lay to the left of zero. Colours are assigned according to the levels of SILAC enrichment according to the colour scale shown in B.

Figure 5.7

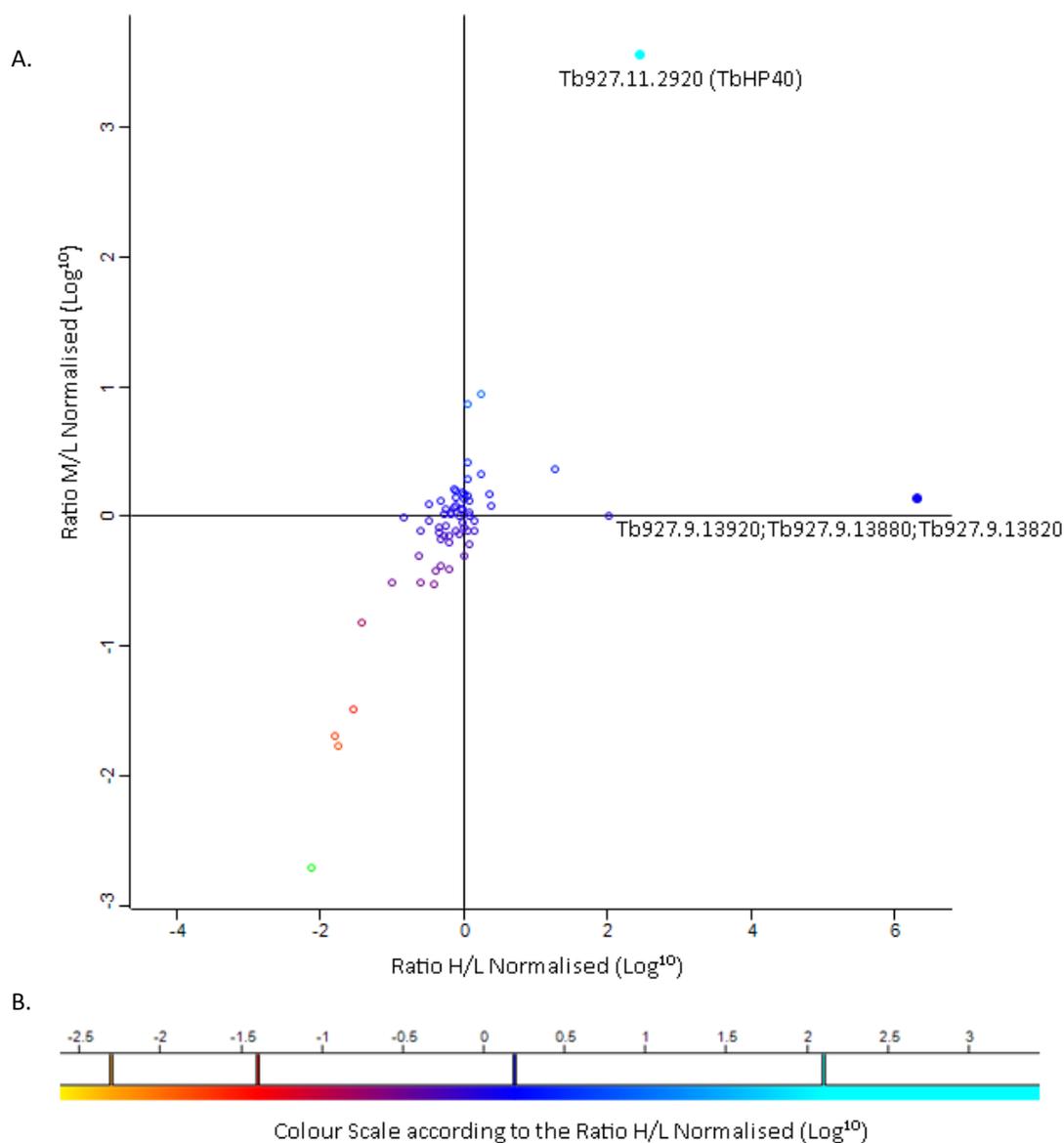


Figure 5.7. TheLC-MS/MS scatter plot comparing proteins identified in the SILAC-enriched 'medium'-labelled *TbHP40BirA** cell line ratio versus 'light'-labelled SmOx-P9 parental cell line against proteins identified in 'heavy'-labelled *TbHP40BirA** cell line ratio versus 'light'-labelled SmOx-P9 parental cell line.

The x-axis shows the ratio of 'heavy'-labelled proteins versus 'light'-labelled proteins, identified according to the peptides detected by mass spectrometry. The y-axis shows the ratio of 'medium'-labelled proteins versus 'light'-labelled proteins, identified according to the peptides detected by mass spectrometry. Scales are logged so data is normally distributed and the ratios normalised because the base peak originally lay to the left of zero. Colours are assigned according to the levels of SILAC enrichment according to the colour scale shown in B.

Figure 5.8

A.

Protein ID	Peptide Sequence	Charge	Mass (Da)	Modifications	Modified sequence
<i>Tb</i> 927.11.2920 (<i>Tb</i> HP40)	DNVQLFR	2	890.491	Unmodified	_DNVQLFR_
<i>Tb</i> 927.11.2920 (<i>Tb</i> HP40)	EQEIGNFQPLDK	2	1416.689	Unmodified	_EQEIGNFQPLDK_

B.

***Tb*927.11.2920 (*Tb*HP40)**

MSRCGRKLDMSSGYWLP RRLLLVRHGSEANVDRALYSKVPDWKIPLTARGREQAFECGRRLRKIIKNEKL
YVYSPYTRTRQTLTEVRKSLPSQVQGEREDERLR**EQEIGNFQPLDK**MDEMWAERSEFGRSYRFPDGES
SVDVGDVRSKFFDSLFRERVELNYLSARKQMITGSSNDVGPASFTVPD DDDHNVVISHGLLIRL FVGRWYS
APMEVFETMKNPPNCGIVVLERREAGRLVMTDTSKLLFGSDP LLEMMKFDGK**DNVQLFR**HLFAEGGYSF
SAGEGTDR

C.

Protein ID	Peptide Sequence	Charge	Mass (Da)	Modifications	Modified sequence
<i>Tb</i> 927.9.13920 <i>Tb</i> 927.9.13880 <i>Tb</i> 927.9.13820	FFADKPDEATLS PEMK	3	1824.86	Unmodified	_FFADKPDEATLSP EMK_
<i>Tb</i> 927.9.13920 <i>Tb</i> 927.9.13880 <i>Tb</i> 927.9.13820	FFADKPDEATLS PEMK	3	1824.86	Unmodified	_FFADKPDEATLSP EMK_

D.

***Tb*927.9.13920 (KMP-11)**

MATTYEEFAAKLDRLDAEFAKKMEEQNKR**FFADKPDEATLSP**EMKEHYEKFEKMIQEHTDKFNKKMRE
HSEHFKA KFAELLEQQKNAQFPGK

***Tb*927.9.13880 (Unspecified Product)**

MATTYEEFAAKLDRLDAEFAKKMEEQNKR**FFADKPDEATLSP**EMKEHYEKFEKMIQEHTDKFNKKMRE
HSEHFKA KFAELLEQQKNAQFPGK

***Tb*927.9.13820 (KMP-11)**

MATTYEEFAAKLDRLDAEFAKKMEEQNKR**FFADKPDEATLSP**EMKEHYEKFEKMIQEHTDKFNKKMRE
HSEHFKA KFAELLEQQKNAQFPGK

Figure 5.8. Identification of *Tb*HP40 from mass spectrometry results

A. Details of the two peptides identified by mass spectrometry that are associated with *Tb*HP40 (*Tb*927.11.2920).

B. The full length amino acid sequence of *Tb*HP40. The positions of the peptides identified by mass spectrometry are highlighted in red.

C. Details of the anomalous enriched peptide identified by mass spectrometry in the M/L normalised data (Figure 5.5). This peptide was associated with three possible candidate protein IDs.

D. The full length amino acid sequences of the three candidate protein IDs described in Figure 5.8 C. The location of the peptide identified by mass spectrometry is highlighted in purple.

Figure 5.9

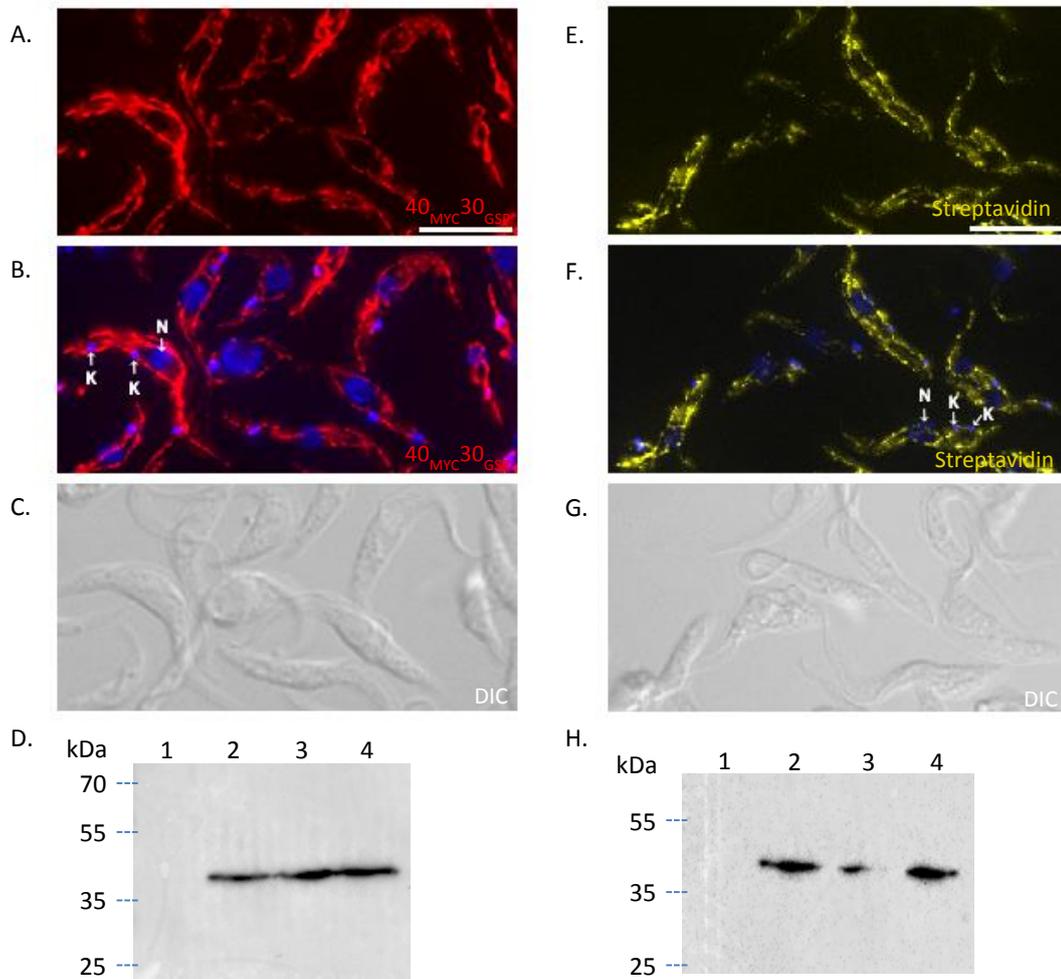


Figure 5.9. Detection of the constitutive co-expression of *TbHP40::myc₃* from pDex477 and *TbHP30::GAFSINPAM* from pDex377 in procyclic 427 *T. brucei* (referred to as 40_{MYC}30_{GSP}).

A-C. The co-expression of 40_{MYC}30_{GSP}, both of which localise to the mitochondrion of *T. brucei*, imaged by immunofluorescence. A-B: 40_{MYC}30_{GSP} decorated with anti-myc antibody at a concentration of 1/1000 in blocking buffer B; B is merged with DAPI; N, nucleus; K, kinetoplast; C: DIC.

D. Western blot detection of 40_{MYC}30_{GSP} in *T. brucei*. Lane 1: 427 *T. brucei* whole cell protein sample (negative control); Lane 2: 40_{MYC}30_{GSP} whole cell protein sample; Lane 3: 40_{MYC}30_{GSP} cells lysed using IP Lysis/Wash buffer, insoluble pellet fraction; Lane 4: 40_{MYC}30_{GSP} cells lysed IP Lysis/Wash buffer, soluble supernatant fraction protein sample. Lanes 2-4 show detection of a protein which corresponds with the expected size of *TbHP40::myc₃* (~39 kDa). The blot was decorated with the anti-myc antibody at a dilution of 1/1000 in blocking buffer A.

E-G. The co-expression of 40_{MYC}30_{GSP}, both of which localise to the mitochondrion of *T. brucei*, imaged by immunofluorescence. E-F: 40_{MYC}30_{GSP} decorated with anti-DB9 antibody at 1/200 in blocking buffer B; F: merged with DAPI; N, nucleus; K, kinetoplast; G: DIC.

H. Western blot detection of 40_{MYC}30_{GSP}. Lane 1: 427 *T. brucei* whole cell protein sample (negative control); Lane 2: 40_{MYC}30_{GSP} whole cell protein sample; Lane 3: 40_{MYC}30_{GSP} cells lysed using IP Lysis/Wash buffer, insoluble pellet fraction; Lane 4: 40_{MYC}30_{GSP} cells lysed IP Lysis/Wash buffer, soluble supernatant fraction protein sample. Lanes 2-4 show detection of a protein which corresponds to the expected size of *TbHP30::GAFSINPAM* (~44 kDa). The blot was decorated with the anti-DB9 antibody at a dilution of 1/100 in blocking buffer A.

Figure 5.10

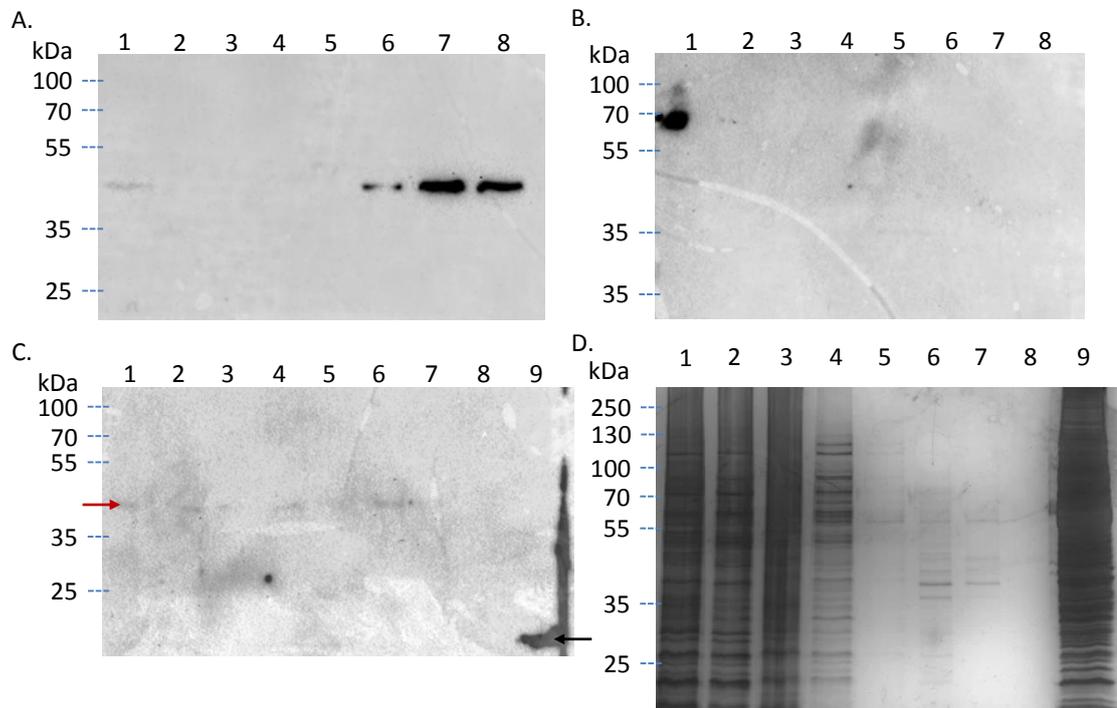


Figure 5.10. Detection of $40_{\text{MYC}}30_{\text{GSP}}$ co-expressed in procyclic *T. brucei* using co-immunoprecipitation techniques.

A. Western blot detection of the co-immunoprecipitation of $40_{\text{MYC}}30_{\text{GSP}}$. Lane 1: Flow-through from myc-immobilised resin; Lane 2-4: resin washes; Lane 5: Conditional resin wash; Lane 6-8: resin elutions. All protein samples were loaded in 'non-reducing' Lane Marker Sample Buffer. Lanes 6-8 show the detection of a protein which corresponds with the expected size of *TbHP40::myc*₃. The blot was decorated with anti-myc antibody at a dilution of 1/1000 in blocking buffer A.

B. Western blot detection of the co-immunoprecipitation of $40_{\text{MYC}}30_{\text{GSP}}$. Lane order: Same as 5.10 A. However, there is no detection of the *TbHP30::GAFSINPAM* signal, which is lost after $40_{\text{MYC}}30_{\text{GSP}}$ is incubated with the myc-immobilised resin. All protein samples were loaded in 'non-reducing' Lane Marker Sample Buffer. The blot was decorated with anti-DB9 antibody at a dilution of 1/100 in blocking buffer A.

C. A Western blot examining the stability of *TbHP30::GAFSINPAM* when incubated with the co-immunoprecipitation IP Lysis/Wash buffer in the presence of a cocktail of protease inhibitors. Lane 1: Time (post-incubation on the myc-immobilised resin) = 0 hours; Lane 2: Time = 1 hour; Lane 3: Time = 2 hours; Lane 4: Time = 3 hours; Lane 5: Time = 4 hours; Lane 6: Time = 5 hours; Lane 7: Time = 16 hours; Lane 8: Blank; Lane 9: recombinant TDP43 (positive control) indicated with a black arrow. The Western blot detected faint bands which corresponded to the expected size of *TbHP30::GAFSINPAM* up until 5 hours (Lane 6), positive control still detected. The blot was decorated with anti-DB9 antibody at a dilution of 1/100 in blocking buffer A.

D. Silver stain detection of proteins from a repeated co-immunoprecipitation experiment in the presence of the protease inhibitor cocktail. Lane 1: $40_{\text{MYC}}30_{\text{GSP}}$ lysate (pre-resin incubation); Lane 2: $40_{\text{MYC}}30_{\text{GSP}}$ lysate 1 hour post-incubation on resin; Lane 3: resin flow-through after 4 hour incubation with $40_{\text{MYC}}30_{\text{GSP}}$ lysate; Lane 4: resin wash one; Lane 5: resin wash three; Lanes 6-8: resin elutions; Lane 9: $40_{\text{MYC}}30_{\text{GSP}}$ whole cell protein sample in Laemlii buffer. All protein samples except Lane 9 were loaded in 'non-reducing' Lane Marker Sample Buffer.

Figure 5.11

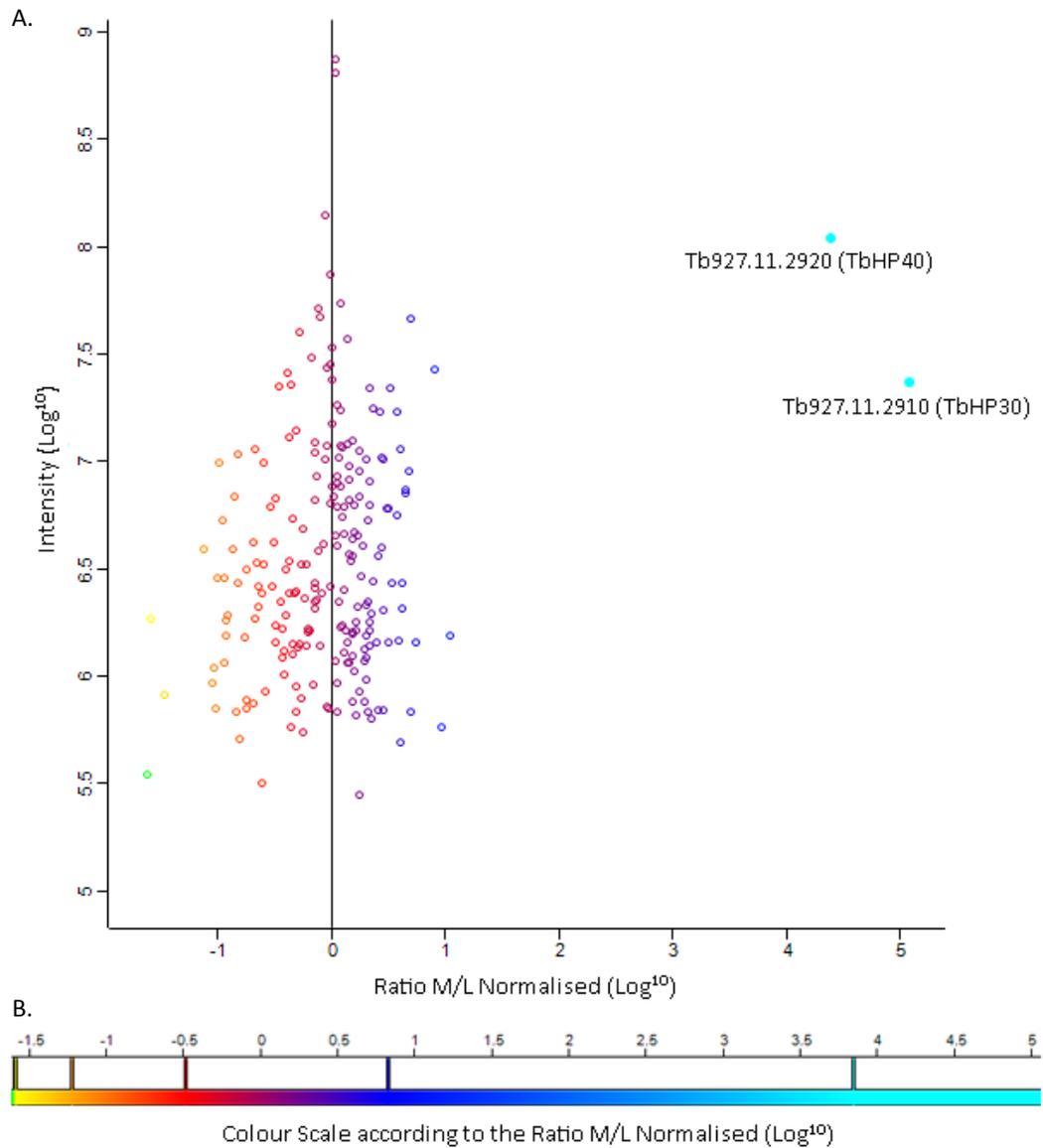


Figure 5.11. The LC-MS/MS scatter plot showing the proteins identified in the SILAC-enriched 'medium'-labelled $40_{\text{MYC}}30_{\text{GSP}}$ cell line ratio versus proteins identified in the 'light'-labelled 427 parental cell line plotted with respect to intensity.

The x-axis shows the ratio of 'medium'-labelled proteins versus 'light'-labelled proteins, identified according to the peptides detected by mass spectrometry. The y-axis shows the relative abundance of the proteins. Scales are logged so data is normally distributed and the ratios normalised because the base peak originally lay to the left of zero. Colours are assigned according to the levels of SILAC enrichment according to the colour scale shown in B.

Figure 5.12

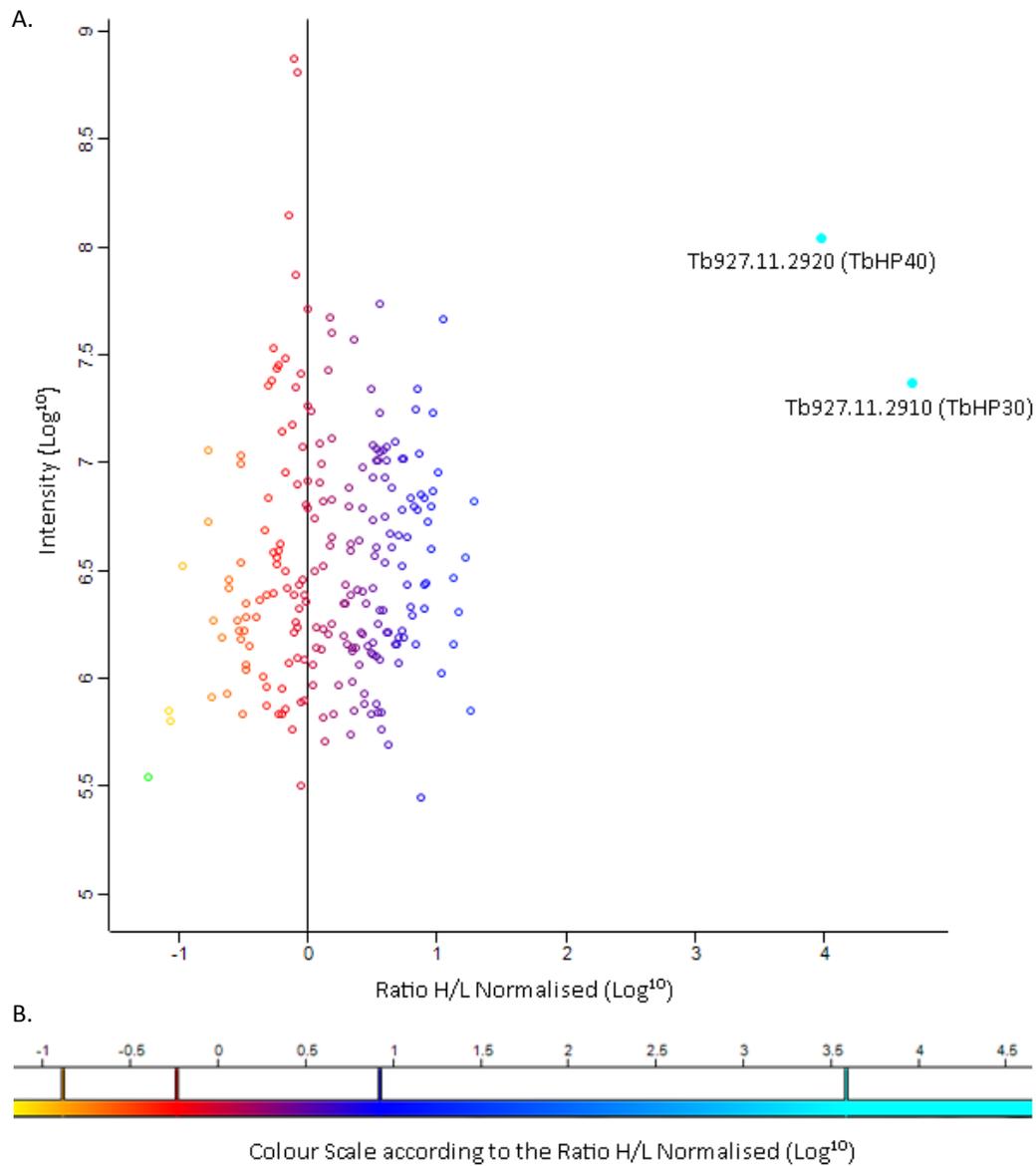


Figure 5.12. The LC-MS/MS scatter plot showing the proteins identified in the SILAC-enriched 'heavy'-labelled 40_{MYC}30_{GSP} cell line ratio versus proteins identified in the 'light'-labelled 427 parental cell line plotted with respect to intensity.

The x-axis shows the ratio of 'heavy'-labelled proteins versus 'light'-labelled proteins, identified according to the peptides detected by mass spectrometry. The y-axis shows the relative abundance of proteins. Scales are logged so data is normally distributed and the ratios normalised because the base peak originally lay to the left of zero. Colours are assigned according to the levels of SILAC enrichment according to the colour scale shown in B.

Figure 5.13

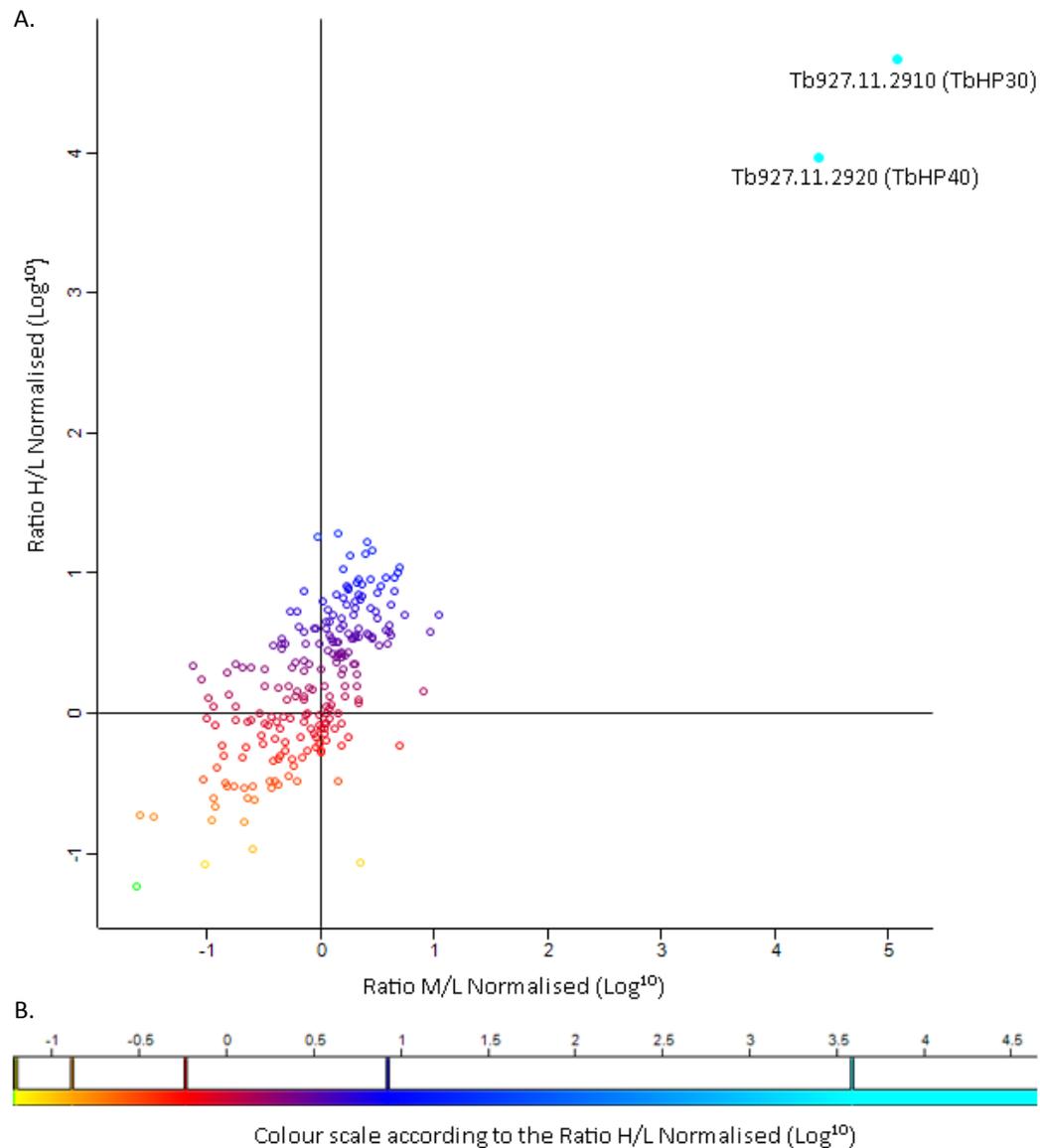


Figure 5.13 The LC-MS/MS scatter plot comparing proteins identified in the SILAC-enriched 'medium'-labelled 40_{MYC}30_{GSP} cell line ratio versus 'light'-labelled 427 parental cell line against proteins identified in 'heavy'-labelled 40_{MYC}30_{GSP} cell line ratio versus 'light'-labelled 427 parental cell line.

The x-axis shows the ratio of 'heavy'-labelled proteins versus 'light'-labelled proteins, identified according to the peptides detected by mass spectrometry. The y-axis shows the ratio of 'medium'-labelled proteins versus 'light'-labelled proteins, identified according to the peptides detected by mass spectrometry. Scales are logged so data is normally distributed and the ratios normalised because the base peak originally lay to the left of zero. Colours are assigned according to the levels of SILAC enrichment according to the colour scale shown in B.

Table 5.1

Table 5.1 Peptide data of the co-immunoprecipitation SILAC-enriched 40_{MYC}30_{GSP} experiment obtained by LC-MS/MS.

A. Table showing the peptides relating to *TbHP40* (*Tb927.11.2920*).

B. The full length amino acid sequence of *TbHP40*, the location of the peptides identified by LC-MS/MS are highlighted in red.

A.

Protein ID	Sequence	Charge	Mass (Da)	Modifications	Modified sequence
<i>Tb927.11.2920</i>	DNVQLFR	2	890.461	Unmodified	<u>DNVQLFR</u>
<i>Tb927.11.2920</i>	EQEIGNFQPLDK	2	1416.689	Unmodified	<u>EQEIGNFQPLDK</u>
<i>Tb927.11.2920</i>	EQEIGNFQPLDK	2	1416.689	Unmodified	<u>EQEIGNFQPLDK</u>
<i>Tb927.11.2920</i>	EQEIGNFQPLDK MDEMWAER	3	2497.089	2 Oxidation (M)	<u>EQEIGNFQPLDKM(ox))DEM(ox)WAER</u>
<i>Tb927.11.2920</i>	ERVELNYLSAR	3	1348.71	Unmodified	<u>ERVELNYLSAR</u>
<i>Tb927.11.2920</i>	ERVELNYLSAR	2	1348.71	Unmodified	<u>ERVELNYLSAR</u>
<i>Tb927.11.2920</i>	ERVELNYLSAR	2	1348.71	Unmodified	<u>ERVELNYLSAR</u>
<i>Tb927.11.2920</i>	ERVELNYLSAR	3	1348.71	Unmodified	<u>ERVELNYLSAR</u>
<i>Tb927.11.2920</i>	ERVELNYLSAR	2	1330.699	Glu->pyro-Glu	<u>(gl)ERVELNYLSAR</u>
<i>Tb927.11.2920</i>	FDGKDNVQLFR	3	1337.673	Unmodified	<u>FDGKDNVQLFR</u>
<i>Tb927.11.2920</i>	FDGKDNVQLFR	3	1337.673	Unmodified	<u>FDGKDNVQLFR</u>
<i>Tb927.11.2920</i>	FDGKDNVQLFR	2	1337.673	Unmodified	<u>FDGKDNVQLFR</u>
<i>Tb927.11.2920</i>	FDGKDNVQLFR	3	1337.673	Unmodified	<u>FDGKDNVQLFR</u>
<i>Tb927.11.2920</i>	FFDSLFR	2	930.46	Unmodified	<u>FFDSLFR</u>
<i>Tb927.11.2920</i>	FPDGESSVDVGD R	2	1378.6	Unmodified	<u>FPDGESSVDVGDR</u>
<i>Tb927.11.2920</i>	FPDGESSVDVGD RVSK	3	1692.795	Unmodified	<u>FPDGESSVDVGD RVSK</u>
<i>Tb927.11.2920</i>	FPDGESSVDVGD RVSK	3	1692.795	Unmodified	<u>FPDGESSVDVGD RVSK</u>
<i>Tb927.11.2920</i>	GREQAFECGR	2	1208.536	Unmodified	<u>GREQAFECGR</u>

Table 5.1. Continued...

Tb927.11.2920	GREQAFECGR	2	1208.536	Unmodified	_GREQAFECGR_
Tb927.11.2920	KIIKNEK	2	871.5491	Unmodified	KIIKNEK
Tb927.11.2920	KLDMSSGYWLP R	3	1467.718	Oxidation (M)	_KLDM(ox)SSGYWLPR_
Tb927.11.2920	KLDMSSGYWLP R	3	1467.718	Oxidation (M)	_KLDM(ox)SSGYWLPR_
A. Tb927.11.2920	KQMITGSSNDV GPASFTVPDDDD HNVVVIISHGLLIR	4	3862.916	Oxidation (M)	_KQM(ox)ITGSSNDVGP SFTVPDDDDHNVVISHGL LIR
Tb927.11.2920	LFGSDPLEMCK	2	1411.673	2 Oxidation (M)	_LFGSDPLEM(ox)M(ox) K_
Tb927.11.2920	LFGSDPLEMCK	2	1411.673	2 Oxidation (M)	_LFGSDPLEM(ox)M(ox) K_
Tb927.11.2920	LREQEIGNFQPL DK	3	1685.874	Unmodified	_LREQEIGNFQPLDK_
Tb927.11.2920	LREQEIGNFQPL DK	3	1685.874	Unmodified	_LREQEIGNFQPLDK_
Tb927.11.2920	LVMTDTSKK	3	1037.543	Oxidation (M)	_LVM(ox)TDTSKK_
Tb927.11.2920	LYVYSPYTR	2	1323.65	Unmodified	_LYVYSPYTR_
Tb927.11.2920	NPPNCGIVVLER	2	1366.703	Unmodified	_NPPNCGIVVLER_
Tb927.11.2920	QTLTEVRK	2	973.5556	Unmodified	_QTLTEVRK_
Tb927.11.2920	QTLTEVRK	2	973.5556	Unmodified	_QTLTEVRK_
Tb927.11.2920	SLLPSQVQGER	2	1212.646	Unmodified	_SLLPSQVQGER_
Tb927.11.2920	SLLPSQVQGERE DER	2	1741.859	Unmodified	_SLLPSQVQGEREDER_
Tb927.11.2920	SYRFPDGESSV DVGDR	3	1947.86	Unmodified	SYRFPDGESSVDVGDR
Tb927.11.2920	SYRFPDGESSV DVGDR	3	1947.86	Unmodified	SYRFPDGESSVDVGDR
Tb927.11.2920	SYRFPDGESSV DVGDR	2	1947.86	Unmodified	SYRFPDGESSVDVGDR
Tb927.11.2920	SYRFPDGESSV DVGDR	2	1947.86	Unmodified	SYRFPDGESSVDVGDR
Tb927.11.2920	SYRFPDGESSV DVGDRVSK	3	2262.055	Unmodified	_SYRFPDGESSVDVGDR VSK_
Tb927.11.2920	SYRFPDGESSV DVGDRVSK	3	2262.055	Unmodified	_SYRFPDGESSVDVGDR VSK_
Tb927.11.2920	VELNYSAR	2	1063.566	Unmodified	_VELNYSAR_

B. **Tb927.11.2920 (TbHP40)**

MSRCGR**KLDMSSGYWLP**RRLLLVRHGESEANVDRALYSKVPDWKIPLTARG**REQAFECGR**RLRK**KIIKNE**
KLYVYSPYTRTR**QTLTEVRK****SLLPSQVQGEREDER**L**REQEIGNFQPLDK**M**DEMWAER**SE**FGRSYRFP**
DGESSVDVGDRVSKFF**DSLFRERVELNYSAR**K**QMITGSSNDVGPASFTVPDDDDHNVVVIISHGLLIR**L**F**
VGRWYSAPMEVFETMK**NPPNCGIVVLER**RE**AGRLVMTDTSKK**L**F****GSDPLEMCK****FDGKDNVQLFRH**
LFAEGGYSFSAGEGTDR

Table 5.2

Table 5.2 Peptide data of the co-immunoprecipitation SILAC-enriched 40_{MYC}30_{GSP} experiment obtained by LC-MS/MS

A. Table showing the peptides relating to *TbHP30* (*Tb927.11.2910*).

B. The full length amino acid sequence of *TbHP30*, the location of the peptides identified by LC-MS/MS are highlighted in red.

A.

Protein ID	Sequence	Charge	Mass (Da)	Modifications	Modified sequence
<i>Tb927.11.2910</i>	ERMDFPMDTNV VILTHGQTIR	4	2504.215	2 Oxidation (M)	_ERM(ox)DFPM(ox)DTNNV ILTHGQTIR_
<i>Tb927.11.2910</i>	FPHGESGADVCD R	3	1445.599	Unmodified	_FPHGESGADVCDR_
<i>Tb927.11.2910</i>	FPHGESGADVCD RVTSFLDAFQR	4	2610.192	Unmodified	_FPHGESGADVCDRVTSFLD AFQR_
<i>Tb927.11.2910</i>	IPIVPEGEEESYD AGR	2	1759.826	Unmodified	_IPIVPEGEEESYDAGR_
<i>Tb927.11.2910</i>	IPIVPEGEEESYD AGRR	3	1915.928	Unmodified	_IPIVPEGEEESYDAGR_
<i>Tb927.11.2910</i>	LAQLIGDEPVYYY LSPYVR	3	2258.162	Unmodified	_LAQLIGDEPVYYYLSPYVR_
<i>Tb927.11.2910</i>	LAQLIGDEPVYYY LSPYVR	2	2258.162	Unmodified	_LAQLIGDEPVYYYLSPYVR_
<i>Tb927.11.2910</i>	LAQLIGDEPVYYY LSPYVR	2	2258.162	Unmodified	_LAQLIGDEPVYYYLSPYVR_
<i>Tb927.11.2910</i>	MDFPMDTNVVI LTHGQTIR	3	2219.071	2 Oxidation (M)	_M(ox)DFPM(ox)DTNNVIL THGQTIR_
<i>Tb927.11.2910</i>	SEHKMEGESIVG VR	3	1572.757	Oxidation (M)	_SEHKM(ox)EGESIVGVR_
<i>Tb927.11.2910</i>	SEHKMEGESIVG VR	3	1572.757	Oxidation (M)	_SEHKM(ox)EGESIVGVR_
<i>Tb927.11.2910</i>	SLANVNVCTHVT TPDWR	3	1968.948	Unmodified	_SLANVNVCTHVTPDWR_
<i>Tb927.11.2910</i>	VTSFLDAFQR	2	1182.603	Unmodified	_VTSFLDAFQR_
<i>Tb927.11.2910</i>	WFNLTVDTYHS MVSPPTGSISTLT R	3	2825.369	Oxidation (M)	_WFNLTVDTYHSM(ox)VSP PTGSISTLTR_

B.

***Tb927.11.2910* (*TbHP30*)**

MAVVRRCSHHRICKQWDRIVSCGSLKAPLCDVMHFTSSTWPPHLHPQPTTQTVRPLEVSKETKALVEEPAVAV
 NLCETAANAATDPASSMFVSPSPTSAYVTHTESQGRMSITIRSGDRRQMLKFPTQPVEPVKRIIFLRNGR**SLAN**
VNVCTHVTPDWRIPIVPEGEEESYDAGRRLAQLIGDEPVYYYLSPYVRCRQSFKHVLRGYDAYR**SEHKMEGESIV**
GVREEDVRLRDGDIGRYKSGELLHHLAEREKYGKFYR**FPHGESGADVCDRVTSFLDAFQR**ERMDFPMDTNVVI
LTHGQTIRMFVKR**WFNLTVDTYHSMVSPPTGSISTLTR**MHHRSSFRLLDACIESMRLPPLSNKYNGYKYNKQV
 LGSMSTGAPFM

Chapter 6

Heterologous expression in *Trypanosoma brucei* of candidate mitochondrial proteins from *Naegleria gruberi*

Classically *N. gruberi* was thought of as an obligate aerobe that relied on mitochondrial oxidative phosphorylation for its survival. However, sequencing and annotation of the *N. gruberi* nuclear genome predicted an unexpected capacity for anaerobic metabolism (Fritz-Laylin et al., 2010b; Ginger et al., 2010). Within the list of predicted enzymes for anaerobic metabolism were FeFe-hydrogenase, three proteins (HydE, HydF and HydG) which are commonly associated with FeFe-hydrogenase maturation, a nitrite reductase, and other enzymes associated with substrate level phosphorylation such as acetyl CoA synthetase (ADP forming) (Figure 6.1). FeFe-hydrogenase is a crucial enzyme that is essential for the production of hydrogen during anaerobic fermentation and is generally inactivated when in the presence of oxygen (see Section 1.4). Presumably, the FeFe-hydrogenase and its other enzymes classically linked to anaerobic metabolism permits *N. gruberi* survival in anaerobic conditions.

In my hands, placement of *N. gruberi* trophozoites under anaerobic conditions in the presence and absence of nitrate and nitrite source and treatment with metabolic inhibitors (rotenone, antimycin A, oligomycin A, sodium azide or salicylhydroxamic acid at concentrations ranging from 0.5-500 μ M) resulted in encystment. In contrast, application of the mitochondrial poison rotenone (a complex I inhibitor) at 5 μ M killed the cells (data not shown). Thus, there appears to be no potential for anaerobic growth and division by axenically cultured *N. gruberi*. However, when placed under anaerobic conditions a small amount of H₂ production has been observed (David Lloyd, University of Cardiff, personal communications; Tsaousis et al., (2014)).

Some of the enzymes associated with anaerobic metabolism in *N. gruberi* were predicted to be mitochondrial on a basis of N-terminal leader sequences that were predicted to act as mitochondrial targeting sequences (Fritz-Laylin et al., 2010b). Others were predicted to be mitochondrial on a basis of their likely substrate specificity (e.g. acetyl CoA synthetase [ADP forming]). Yet, until now experimental evidence for mitochondrial localisation was lacking. Indeed, a recent study performed on a basis of heterologous expression in yeast of partial *N. gruberi* FeFe-hydrogenase sequences suggested that this enzyme was actually cytosolic

(Tsaousis et al., 2014). In this chapter I first provide a bioinformatic overview of FeFe-hydrogenase occurrence and protein architecture among eukaryotes, and then describe the results from a series of experiments designed to probe the localisation of the enzymes listed in Table 6.1.

6.1 Bioinformatic analysis of a range of eukaryotic FeFe-hydrogenases

All FeFe-hydrogenases identified here (Figure 6.2) contain a large iron hydrogenase domain, *ca.* 300 amino acids, located towards the C-terminus. Many proteins also have a '4Fe-4S ferredoxin' domain (along with at least one copy of its associated binding domain) and a complex I 'G-subunit'. *Blastocystis hominis*, a stramenopile capable of inhabiting the gastrointestinal tract of humans and many other animals, contains an FeFe-hydrogenase that localises to the hydrogenosomes in a cell (Stechmann et al., 2008). The *B. hominis* FeFe-hydrogenase has an additional flavoprotein-like domain found downstream of the iron hydrogenase domain that is not seen in the FeFe-hydrogenase of *N. gruberi*, or indeed any of the other eukaryotes mentioned in this section.

Chlamydomonas reinhardtii is a flagellated green alga found in soil and freshwater that encounters anoxic conditions and is capable of a rapid switch to anaerobic metabolism (Gfeller and Gibbs, 1984). Nutrient limitations can trigger *C. reinhardtii* to produce H₂ under light and dark conditions (Melis and Happe, 2001; Melis et al., 2000). This is due to the presence of FeFe-hydrogenases that localise to the chloroplast (Atteia et al., 2013; Forestier et al., 2003; Happe and Kaminski, 2002). *Entamoeba histolytica* is an obligate anaerobic flagellate that possesses a cytosolic FeFe-hydrogenase, although so far hydrogen production can only be detected when the FeFe-hydrogenase is overexpressed in transformed *E. histolytica* cells (Nixon et al., 2003). Both *C. reinhardtii* and *E. histolytica* are unrelated yet have short, relatively simple FeFe-hydrogenases compared to that of *N. gruberi*. *Giardia intestinalis* and *Trichomonas vaginalis* are also anaerobic flagellates both of which, along with *E. histolytica*, are human parasites. *T. vaginalis* does not possess a classical mitochondrion, and instead has membrane-bound hydrogenosomes that rely on fermentation using H⁺ as a final electron acceptor to make molecular hydrogen (see Section 1.4) (Muller et al., 2012; Putz et al., 2006). *G. intestinalis*, like *E. histolytica* have the greatly reduced forms of mitochondria known as mitosomes (Müller et al., 2012). The *G. intestinalis* FeFe-hydrogenase, like *E. histolytica*, is also cytosolic and both function independently of the mitosome (Nixon et al., 2003). The *G. intestinalis* FeFe-hydrogenase has low hydrogen

production rates (approximately 10-fold lower than *T. vaginalis*) (Ellis et al., 1992; Lloyd et al., 2002) Both *T. vaginalis* and *G. intestinalis* again have relatively simple FeFe-hydrogenases that look almost identical to one another, differing in length by only 6 residues and both are fairly similar to that found in *C. reinhardtii*. However, *T. vaginalis* also possesses an additional bigger 64 kDa FeFe-hydrogenase that contains a ferredoxin domain, which is not present in the shorter isoform. In the case of *T. vaginalis*, the presence of multiple FeFe-hydrogenases may reflect the fact it (or its ancestor) is an obligate anaerobe that has adapted to a specific environment, whereas *C. reinhardtii* is an autotroph that is also capable of surviving in the prolonged absence of oxygen (Gfeller and Gibbs, 1984; Happe and Kaminski, 2002).

Chlorella variabilis is another chlorophyte green alga that classically photo-respires, but it too possesses FeFe-hydrogenase and can generate molecular hydrogen (Meuser et al., 2011). The *C. variabilis* FeFe-hydrogenase appears almost identical in domain architecture to that of *Naegleria gruberi*, and only differs in length by 33 amino acids. *Sawyeria marylandensis* belongs to the Heterolobosea, as does *N. gruberi*. It is a microaerophile that contains hydrogenosomes and grows with trace amounts of oxygen in its environment (Barbera et al., 2010). Again the *S. marylandensis* FeFe-hydrogenase strongly resembles that of *N. gruberi* and *C. variabilis*. Another heterolobosean amoeboflagellate is *Psalteriomonas lanterna* that contains an FeFe-hydrogenase which localises to the hydrogenosomes (de Graaf et al., 2009). The domain architecture of the *P. lanterna* FeFe-hydrogenase appears to be slightly less complex than that of *N. gruberi*, as it lacks an apparent ferredoxin domain towards the N-terminus. *Neocallimastix frontalis* is an anaerobic chytrid that contains hydrogenosomes and an FeFe-hydrogenase (Davidson et al., 2002; van der Giezen et al., 1997). The *N. frontalis* FeFe-hydrogenase again looks very similar in architecture to that of *Naegleria gruberi*, *C. variabilis* and *S. marylandensis*. The unrelated ciliate *Nyctotherus ovalis* is an anaerobic symbiont that grows in the hindgut of cockroaches. *N. ovalis* also has hydrogenosomes (Boxma et al., 2005a) and it is perhaps not surprising to see that its FeFe-hydrogenase is longer and more complex than the others analysed, as this is its primary method of energy generation (Boxma et al., 2005a). *N. ovalis* still retains features in the first half of the enzyme which look very similar to the FeFe-hydrogenases of *Naegleria gruberi*, *C. variabilis*, and *S. marylandensis*, but then has additional domains related to Complex I subunits in the C-terminal half of the protein (Boxma et al., 2007).

All of the hydrogenases found in *Chlamydomonas reinhardtii*, *Entamoeba histolytica*, *Giardia intestinalis* and *Trichomonas vaginalis* are shorter and less complex in domain architecture than the FeFe-hydrogenase found in *Naegleria gruberi*. The domain architecture in *N. gruberi*

most closely resembles that found in the green alga *Chlorella variabilis* since they both have extended N-termini compared to many of the other FeFe-hydrogenases illustrated in Figure 6.2. The *N. gruberi* FeFe-hydrogenase also greatly resembles those found in the close relative *S. marylandensis* and the unrelated fungi *Neocallimastix frontalis*. The FeFe-hydrogenases found in *Nyctotherus ovalis* and *B. hominis* appear more complex in terms of their domain architecture than that of *Naegleria gruberi* since both have additional C-terminal domains that are not present in *N. gruberi*. The *B. hominis* sequence is only 8 residues longer than *N. gruberi* but the *Nyctotherus ovalis* sequence is significantly longer by 446 residues.

The different architectures seen across a range eukaryotic hydrogenases perhaps argue in favour of multiple independent acquisitions of this enzyme. The phylogenetic relationships between all of the FeFe-hydrogenases described above was not pursued, because of previous reports that concluded that FeFe-hydrogenase phylogenies have poor resolution capabilities (Embley, 2006; Ginger et al., 2010; Hug et al., 2010). However, these earlier publications did not compare the domain architecture of the full length FeFe-hydrogenase proteins.

6.2 Antibody-based detection of *N. gruberi* FeFe-hydrogenases

Upon arrival to Lancaster, polyclonal FeFe-hydrogenase anti-peptide antibodies had been raised against short peptide sequences found in the *N. gruberi* FeFe-hydrogenase (*NgFeHyd*). The relative positions of the peptides used as antigens are shown in red on the cartoon schematic of the *NgFeHyd* (Figure 6.3 B). Preliminary Western blot analysis of amoebae and differentiating flagellates revealed several proteins decorated by the anti-peptide antibody. The *NgFeHyd* would be expected to be seen at approximately 80 kDa; although no bands are detected in this region, a series of distinct bands were observed (Figure 6.3 A). Two proteins of higher molecular weight than the 80kDa size marker are seen in Figure 6.3 A. One of these bands is approximately 100 kDa in size and if *NgFeHyd* migrates aberrantly on SDS-PAGE conceivably the antibody recognises its target (this is expected since trophozoites have been observed to produce H₂, therefore its anticipated that *NgFeHyd* protein would be present in whole cell protein samples - see Tsaousis et al., (2014)). Proteins of lower molecular weight identified by the affinity purified anti-sera may correspond to cross-reacting proteins or may simply be due to proteolysis of the sample.

To explore further whether any of the protein bands detected in Western blotting using the anti-FeFe-hydrogenase antibodies correspond to the *NgFeHyd* or a proteolytic cleavage product, anti-sera was next tested against recombinant protein. Here, a gene sequence coding for the first 273 amino acids of FeFe-hydrogenase and flanked by NdeI and XhoI restriction sites was commercially synthesised (known as FeHyd^{Δ274-752}). This commercially synthesised gene was excised from the standard cloning vector pEX-A (in which it was supplied) by digestion with NdeI and XhoI and cloned into NdeI-XhoI digested pET-28a (Novagen) (for vector map see Appendix II Figure B.2) to give FeHyd^{Δ274-752} with an N-terminal hexa-histidine tag. This was transformed into XL-1 blue competent *E.coli*. Following successful transformation of the *E. coli*, colonies were observed on the 'ligation' plate and six colonies were picked to be taken forwards for overnight growth and the isolation of plasmid DNA using a GeneJET™ Plasmid Miniprep kit (ThermoFisher Scientific). Plasmid DNA was screened by digestion with the restriction enzymes XbaI and XhoI. Following digestion, presence of a vector-based backbone at size ~5300 bp and an insert at size ~800 bp corresponding to FeHyd^{Δ274-752} was taken to be indicative of a successful clone. From the image shown in Figure 6.4 A it was clear that two plasmid mini-preps (colonies 1 and 5) contained insert. Colony 6 yielded plasmid of uncertain origin and was not pursued further, colony 5 (marked with an asterisk on figure 6.4 A) was taken forward for DNA sequencing, then subsequently used in further downstream protein expression experiments.

The FeHyd^{Δ274-752} pET-28a colony 5 was re-transformed into Rosetta™2(DE3)pLysS competent cells (Novagen) in preparation for small scale expression and protein solubility experiments. Small scale induction showed IPTG inducible expression of a protein of ~35 kDa in size; this was found to be insoluble protein rather than soluble (Figure 6.5 A-B). Detection with the anti-His antibody confirmed that the induced protein did possess a hexa-histidine tag (Figure 6.5 C). The induced protein was also detected using the *NgFeHyd* anti-peptide antibody (Figure 6.6 D), indicating that the antibodies recognised the peptide antigen originally synthesised for antibody production, and therefore that in theory the antibodies detect the *N. gruberi* FeFe-hydrogenase.

Since the *NgFeHyd* anti-peptide antibody could theoretically recognise its target, the decision was made to attempt to raise another polyclonal antibody, FeHyd^{Δ274-752}, by carrying out large scale induction followed by purification of the recombinant FeHyd^{Δ274-752} protein to provide antigen for antibody production in chicken hosts. Chickens were chosen so that the

derived anti-sera derived could potentially be used in conjunction with the NgFeHyd anti-peptide antibody (sourced from rabbit hosts) for immunofluorescence experiments.

Large scale induction of the recombinant FeHyd^{Δ274-752} in Rosetta™2 (DE3) pLysS competent cells (Novagen) was utilised to provide enough protein for antibody production. The recombinant protein was purified using the ÄKTA-prime insoluble protein protocol, using metal-ion chromatography and SDS-PAGE downstream purification (Figure 6.6 A). The purified protein was sent to Covalab for polyclonal antibody production using chicken hosts. FeHyd^{Δ274-752} was also purified using the benchtop pH shift insoluble protein protocol (Figure 6.6 D-E), which eliminated the need for downstream purification using SDS-PAGE gels and could be used in ELISA testing. Unfortunately, the Covalab ELISA tests of both chickens immunised with FeHyd^{Δ274-752} showed no reactivity (data not shown) and there was no detection of recombinant FeHyd^{Δ274-752} using the raw sera from either chicken under the conditions outlined in Section 2.19 (data not shown).

6.3 Heterologous expression of *N. gruberi* enzymes in *T. brucei*

Since attempts to raise a polyclonal antibody against FeHyd^{Δ274-752} were unsuccessful experiments to look at FeFe-hydrogenase localisation in *N. gruberi* were not pursued further, but an alternative strategy to identify where the *N. gruberi* FeFe-hydrogenase (and subsequently its associated maturases) might localise was used. To do this, *T. brucei* was used for heterologous expression experiments. Here the rationale was that if any of the *N. gruberi* proteins localise to the *T. brucei* mitochondrion, then despite the several or many hundred million years since *T. brucei* and *N. gruberi* shared their last common ancestor (Hampl et al., 2009; Simpson et al., 2006a), this would be good evidence for the mitochondrial localisation of FeFe-hydrogenase and its associated maturases in *N. gruberi*. For instance, mitochondrial targeting sequences from diatoms (sister group to the alveolates) can successfully target yeast mitochondria (Danne and Waller, 2011) and the human homologue of the conserved mitochondrial protein frataxin can target the mitochondrion of *T. brucei* (Long et al., 2008a).

I also looked at the expression profiles of various anaerobic enzymes under the assumption that the expression of RNA translates to the production of protein. Microarray analysis of *N. gruberi* steady state RNA as it differentiates into its flagellate form (Fritz-Laylin and Cande, 2010) indicates that there is transcription of the FeFe-hydrogenase gene and its three

associated maturases, HydE, HydF and HydG (Figure 6.7 B-E); which is expected because trophozoites produce H₂ (Tsaousis et al., 2014). There is also some possibility that the expression of FeFe-hydrogenase, HydE and HydG may increase during differentiation (Figure 6.7 B-C, Figure 6.7 E).

For FeFe-hydrogenase localisation, focus was on the importance of the N-terminal leader sequence. Codons for the first 60 N-terminal amino acids of FeFe-hydrogenase were fused upstream of coding sequence for a HA-epitope tag and GFP (Figure 6.7 A) and cloned into the *T. brucei* pDex377 expression vector by a previous post-graduate student (J. Kelly) (henceforth this plasmid is known as FeHyd^{Δ61-752}:HA:GFP). Previous preliminary experiments regarding the localisation of the tagged proteins had proved inconclusive.

For the three FeFe-hydrogenase associated maturases, (since none of the encoding maturation proteins genes were predicted to contain introns) a slightly different approach was used. Full length coding sequences for the gene models HydE, HydF and HydG were amplified by PCR (figure 6.8 A), using *N. gruberi* NEG-M genomic DNA as a template. The resultant PCR products for HydE, HydF and HydG were purified using the GeneJET™ PCR Purification Kit and then ligated into the pGEM®-T Easy (Promega) vector. The ligated HydE, HydF and HydG plasmids in the pGEM®-T Easy (Promega) vector were transformed into XL-1 blue competent *E. coli* and colonies observed on the three different 'ligation' plates. Colonies from the HydE, HydF and HydG 'ligation' plates were selected for overnight growth in LB broth and the plasmid DNA extracted from the *E. coli* using the GeneJET™ Plasmid Miniprep Kit (Fermentas). Miniprep DNA was screened by digestion with the HindIII and XhoI restriction enzymes. Restriction vector maps shown in Figure 6.8 A (HydE), 6.8 B (HydF) and 6.8 C (HydG).

Plasmid DNA for HydE isolated from the T Easy vector by restriction digest using HindIII and XhoI was then cloned into a variant of the pDex377 expression plasmid (Wickstead et al., 2003) (for vector map see Appendix II Figure B.3 A) which contained a pre-existing triple-myc tag. When this pDex377 variant was cut with HindIII and XhoI then ligated with HindIII and XhoI digested insert, this would result in HydE C-terminally fused to 3 tandemly placed myc epitopes.

Plasmid DNA for HydF and HydG had to be cloned into the pDex377-myc₃ using an alternative strategy due to HydF and HydG coding sequences each containing internal HindIII sites. The HydF and HydG open reading frames cloned into T Easy were initially digested away from the T Easy vector using NotI. NotI ends were then 'blunted' using the Klenow

fragment of DNA polymerase I. The 'blunt ended' HydF and HydG inserts were then digested with XhoI. The pDex377-myc₃ was initially digested with HindIII, 'blunted' with the Klenow fragment, digested with XhoI and then ligated with 'blunt'-XhoI ended HydF and HydG inserts.

Ligated plasmids were transformed into XL-1 blue competent *E. coli* and miniprep plasmid DNA prepared in the usual fashion using the restriction enzymes BglII and XhoI to screen for insert (BglII was used for screening because the pDex377-myc₃ has a BglII site 144 bases upstream of the HindIII site which was destroyed during the cloning process; the HydF coding sequence also has an internal BglII site which helped differentiate it easily from HydG during the screening process). Plasmids containing HydF and HydG inserts were sent for DNA sequencing and before transfection into procyclic *T. brucei*.

The pDex377-derived vectors containing FeHyd^{Δ61-752}:HA:GFP, HydE::myc₃, HydF::myc₃ and HydG::myc₃ were all linearised by digestion with NotI and then transfected into procyclic 427 *T. brucei*. These constructs are targeted to the 177-base pair repeats of the minichromosomes in *T. brucei*, and in a 427 background constitutive expression of myc-tagged or GFP proteins occurs (Wickstead et al., 2002, 2004). Evidence for the expression of these constructs was sought by fluorescence microscopy and Western blotting.

For cell lines expressing FeHyd^{Δ61-752}:HA:GFP, HydE::myc₃, HydF::myc₃ and HydG::myc₃ whole cell lysates were separated by SDS-PAGE and blotted onto Amersham Hybond-P membranes. Membranes were probed with anti-HA (Figure 6.9 A) or anti-myc (Figure 6.9 B) monoclonal antibodies. The bands detected on the anti-myc Western blots corresponded with the sizes expected for full length proteins for HydE (predicted size of 59.6 kDa), HydF (predicted size of 61 kDa) and HydG (predicted size of 63.9 kDa). A band detected on the anti-HA Western blot corresponds with the expected size of the FeHyd^{Δ61-752}:HA:GFP (34.3 kDa).

Indirect immunofluorescence microscopy was used to look at the localisation of HydE::myc₃, HydF::myc₃ and HydG::myc₃ using anti-myc monoclonal antibody and the anti-HA monoclonal antibody to look at the localisation of FeHyd^{Δ61-752}:HA:GFP. Fluorescence was performed in a background where live cells had been incubated with MitoTracker® Red CMXRos (Invitrogen) prior to cell fixation. DeltaVision images of all four populations show clear mitochondrial localisation of FeHyd^{Δ61-752}:HA:GFP, HydE::myc₃, HydF::myc₃ and HydG::myc₃, all of which were confirmed by MitoTracker co-localisation experiments (Figure 6.10).

In subsequent experiments I went on to assess the mitochondrial candidature of *N. gruberi* nitrite reductase (NirK) and acetyl CoA synthetase (ADP forming). As with FeFe-hydrogenase and its associated maturases, microarray analysis suggests that there may be increased steady state mRNA expression levels of NirK during *N. gruberi* differentiation from the amoeboid trophozoites to the flagellate form, whereas the candidate acetyl CoA synthetase (ADP forming) appears to be more abundant at the transcript level in trophozoites (Figure 6.11 A-B).

The full length coding sequence of the *N. gruberi* nitrite reductase enzyme (henceforth known as NirK:HA) fused in-frame to produce a C-terminal 10 amino acid HA epitope-tagged protein (Figure 6.11 C) had been cloned previously into the pDex377 by post-graduate student (J. Kelly). Although the expression plasmid had been produced, no convincing localisation experiments had again been conducted.

For the acetyl CoA synthetase (ADP forming), just as with FeFe-hydrogenase, the focus was on whether the N-terminal sequence was sufficient to direct import of GFP into a *T. brucei* mitochondrion. A synthetic construct of the first 100 N-terminal amino acids of the *N. gruberi* acetyl CoA synthase (ADP forming) (henceforth known as ACS^{Δ101-1100}:HA:GFP) with a C-terminal HA-epitope tag and a downstream GFP tag (Figure 6.11 D) was ordered (Eurofins). The ACS^{Δ101-1100}:HA:GFP construct was then also cloned into the pDex377 expression vector.

Both the NirK:HA and ACS^{Δ101-1100} constructs were transfected into procyclic 427 *T. brucei* for constitutive expression, and evidence of expression was detected by fluorescence microscopy and Western blotting.

Western blot analysis of NirK:HA expression (Figure 6.9 C) using the anti-HA antibody shows expression of a protein that corresponds with the expected size of NirK:HA (44.5 kDa). The ACS^{Δ101-1100} Western blot (Figure 6.9 D) when probed with anti-HA antibody detects a protein that is slightly bigger than the expected size of the ACS^{Δ101-1100}::HA::GFP fusion (39 kDa). The indirect immunofluorescence slides using the anti-HA monoclonal antibody revealed that both NirK:HA and ACS^{Δ101-1100}::HA::GFP show clear mitochondrial localisation when co-localisation with MitoTracker was performed (Figures 6.12 A-D, 6.12 E-H, respectively).

6.4 Summary of the characterisation attempts of *N. gruberi* novel mitochondrial proteins

In this chapter I have provided evidence for the mitochondrial localisation of a variety of candidate *N. gruberi* enzymes that are classically associated with anaerobic metabolism in other organisms. Predictions were initially made regarding the mitochondrial localisation of FeFe-hydrogenase, its three associated maturases HydE, HydF and Hyd G, the nitrite reductase and the acetyl CoA synthetase (ADP forming) and experiments reported here tested those predictions. In my view, based on other studies showing the conservation of mitochondrial protein targeting across species (Danne and Waller, 2011; Long et al., 2008a), my data proves a strong indication that the *N. gruberi* proteins tested would also localise to mitochondria in *N. gruberi*. Collectively the data suggests that the mitochondria of *N. gruberi* are extremely metabolically flexible.

In contrast to the results presented here, a previous study (Tsaousis et al., 2014) found evidence which suggested the *N. gruberi* FeFe-hydrogenase and the HydE maturase both localised to the cytosol rather than the mitochondria. In attempts to try and resolve why there is such conflicting evidence, the 5' region of the FeFe-hydrogenase (1251 bp; referred to from hereon as FeHyd^{Δ418-752}) used by Tsaousis et al., (2014) was fused in-frame upstream of a C-terminal HA and GFP tag, then transfected into procyclic *T. brucei* in the normal manner. Despite several independent attempts, stable transformants were never obtained. Since expression of the FeHyd^{Δ418-752} mutant was never observed in *T. brucei*, the mitochondrial localisation results yielded from the yeast expression system used by Tsaousis et al., (2014) could not be compared. However, the C-terminal deletion in the FeHyd^{Δ418-752} construct occurs halfway through the iron hydrogenase domain (IPR009016). This could affect how the recombinant protein folds since incorrect disulphide bond formations may occur, and some may be missing entirely because of the incomplete iron hydrogenase domain that is expressed. This can affect the tertiary structure of the expressed protein and potentially lead to protein mis-folding (Rosano and Ceccarelli, 2014), which in turn could have an effect on the stability of the heterologously expressed protein. Consequentially, mis-folding could result in a masking effect on the N-terminal mitochondrial targeting sequence, resulting in the protein being unable to reach the correct target. This may contribute to a partial explanation for why mitochondrial localisations for FeFe- hydrogenase and HydE were not seen in yeast by Tsaousis et al., (2014). Additionally, since yeast is much more evolutionary distant to *N. gruberi* than *T. brucei* is, it may be that the yeast system cannot recognise the *N. gruberi* mitochondrial targeting sequences.

In my experiments, I only expressed the first 60 N-terminal amino acids in order to see whether the predicted N-terminal import sequence was sufficient to target the *T. brucei* mitochondrion, and therefore import the remainder of the protein. Tsaousis et al., (2014) also argued that the N-terminal region of the *N. gruberi* FeFe-hydrogenase is unusually long when aligned with other eukaryotic FeFe-hydrogenase sequences, but when the domain architecture is analysed carefully (Figure 6.2), the N-terminus of the *N. gruberi* FeFe-hydrogenase is in similar length to those found in *C. reinhardtii* and *P. lanterna*. The evolutionary history of eukaryotic FeFe-hydrogenases is obscure and sequences align poorly with one-another (Embley, 2006; Ginger et al., 2010; Hug et al., 2010). An in depth study of the phylogeny of FeFe-hydrogenases (Hug et al., 2010) found that most of the deep structure of the resultant phylogenetic trees are poorly resolved and no convincing evidence for eukaryotic monophyly could be found. Many eukaryotic FeFe-hydrogenases were interspersed with homologous eubacterial sequences in two of the three clades because of at least two separate origins of eukaryotic FeFe-hydrogenases from within the eubacteria (Hug et al., 2010). With regards to the FeFe-hydrogenase associated maturases, Tsaousis et al., (2014) also only expressed the first 690 bp of HydE, whereas in my experiments full length sequences of HydE, HydF and HydG were expressed in procyclic *T. brucei*.

Some of the immunofluorescence data presented by Tsaousis et al., (2014) used heterologous antibodies derived from *T. vaginalis* proteins (Bui and Johnson, 1996) to probe the localisation of *N. gruberi* FeFe-hydrogenase and HydE in trophozoites. The authors reported that the *N. gruberi* FeFe-hydrogenase and HydE are cytosolic and exhibit no co-localisation with the Mitotracker signal. However, the representative trophozoite cells shown by Tsaousis et al., (2014) have a more complex fluorescence pattern detected by the *T. vaginalis* anti-sera than just cytoplasmic localisation alone. There is a possibility that two signals are present in the cell, in both the cytoplasm and the mitochondria since there is what appears to be partial punctate co-localisation with the Mitotracker signal on the merged images shown in Figure 3.4 D & H. from Tsaousis et al., (2014). Using heterologous antibodies in immunofluorescence localisation experiments has been shown previously to yield dubious and unreliable results *e.g.* (Regoes et al., 2005). Here, in immunofluorescence microscopy experiments Regoes et al. (2005) used heterologous anti-Cpn60 antibodies against fixed *Giardia* cells and suggested localisation of a *Giardia* Cpn60 protein that proved incorrect when *Giardia*-specific anti-Cpn60 antibodies were made (Regoes et al., 2005). Therefore one might be cautious over the cellular localisation data presented by Tsaousis et al., (2014) that relied upon a heterologous antibody for detection.

Figure 6.1

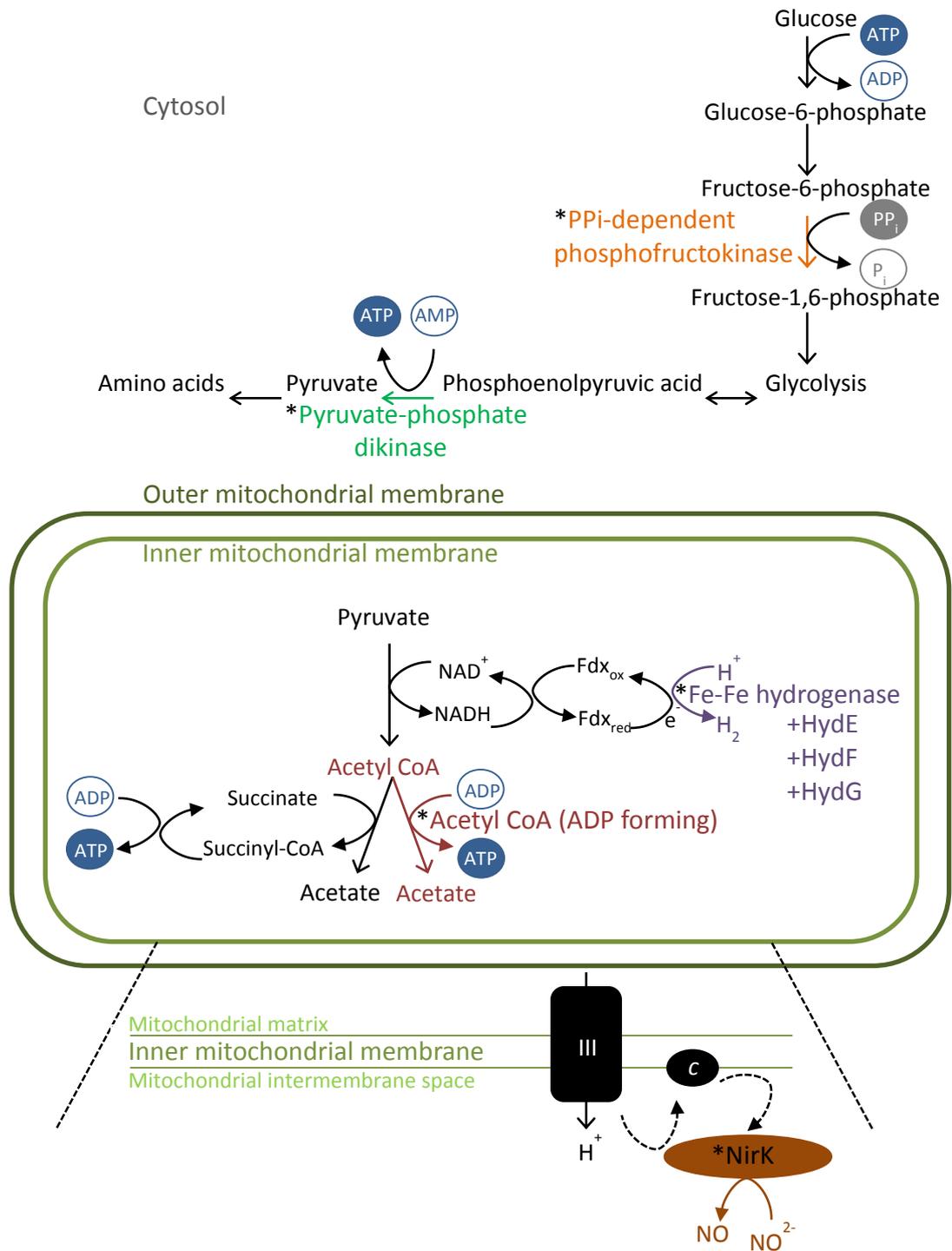


Figure 6.1. A simplified cartoon schematic highlighting the potential anaerobic metabolic capabilities of *N. gruberi*.

Enzymes linked to anaerobic metabolism are highlighted with a black asterisk. Abbreviations: Fdx_{ox}, ferredoxin in its oxidised state; Fdx_{red}, ferredoxin in its reduced state; III, complex III; c, cytochrome C; NirK, nitrite reductase.

Table 6.1

Table 6.1. Accession numbers and JGI protein IDs of the candidate mitochondrial proteins from *N. gruberi* analysed in this thesis.

Genes encoding full length proteins were expressed in procyclic *T. brucei* or alternatively N-terminal coding sequences (FeFe-hydrogenase, acetyl CoA synthetase [ADP forming]) were fused to a GFP gene for expression in procyclic *T. brucei*.

Protein Name	JGI Protein ID	GenBank accession number
FeFe-hydrogenase	NAEGDRAFT_80699	XP_002674266
HydE	NAEGDRAFT_81640	XP_002671091
HydF	NAEGDRAFT_65416	XP_002679321
HydG	NAEGDRAFT_81639	XP_002671139
NirK	NAEGDRAFT_50708	XP_002674759
Acetyl CoA synthetase (ADP forming)	NAEGDRAFT_82174	XP_002669321

Figure 6.2

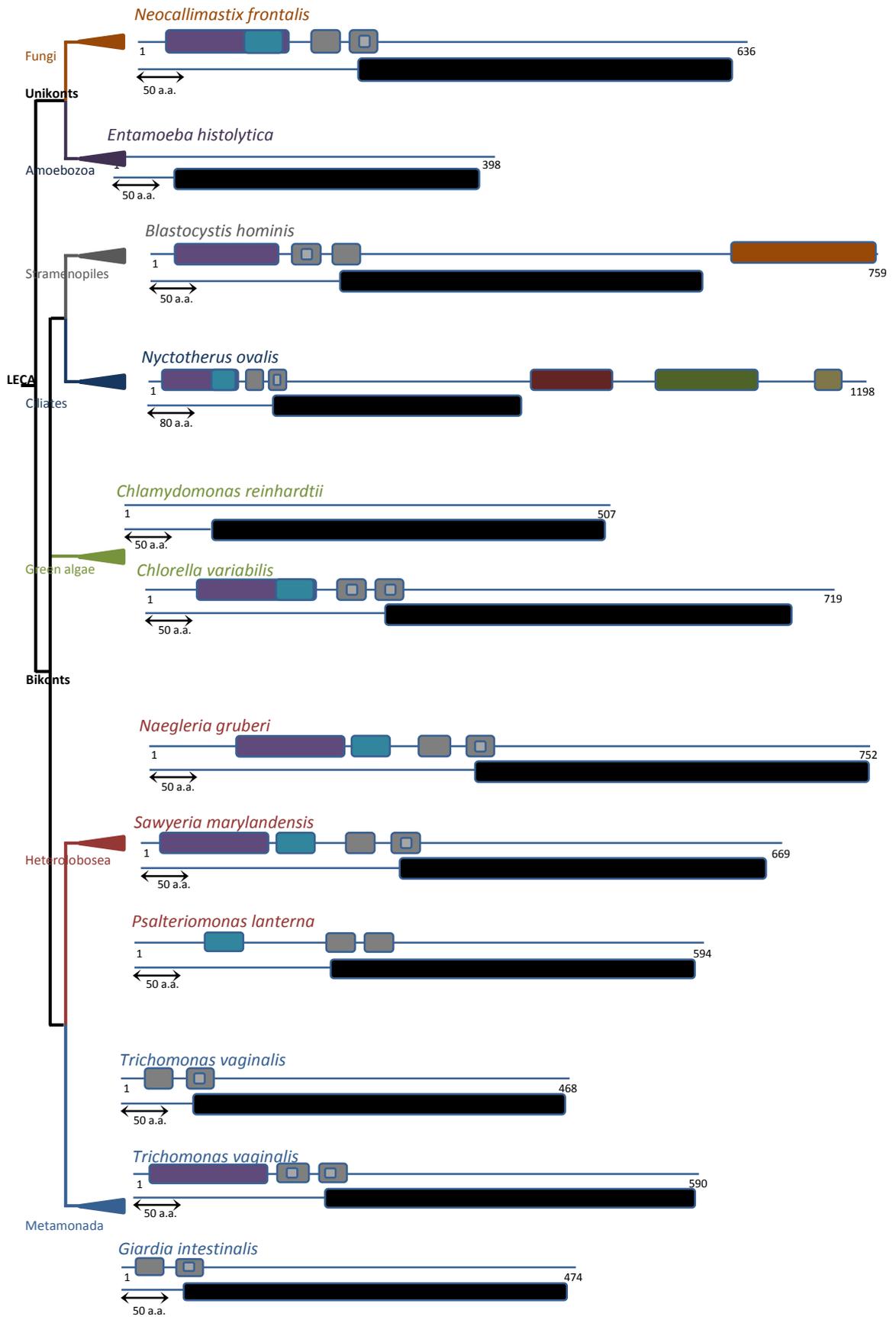


Figure 6.2. Cartoon schematic diagrams showing the domain architecture of a range of eukaryotic FeFe-hydrogenases.

FeFe-hydrogenase domains were detected using InterPro (Mitchell et al., 2015) and SMART (Letunic et al., 2014; Schultz et al., 1998).

No FeFe-hydrogenases have, to my knowledge, been identified yet in red algae or the apicomplexan *Cryptosporidium* sp.

Note that some '4Fe-4S' ferredoxin domains do not always contain two associated 4Fe-4S motifs, which classically mediate the transfer of electrons in metabolic reactions (pale grey squares located within the dark grey squares)

Abbreviations: a.a., amino acids; LECA, last eukaryotic common ancestor

See below for key - InterPro reference number in brackets:

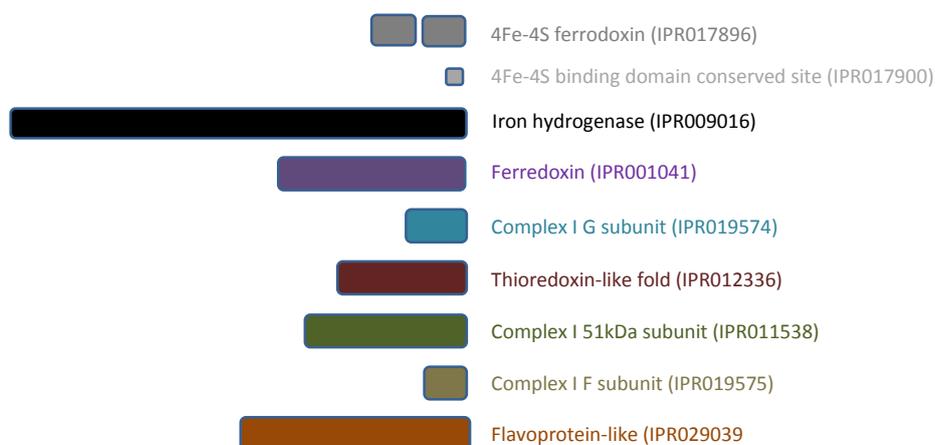


Figure 6.3

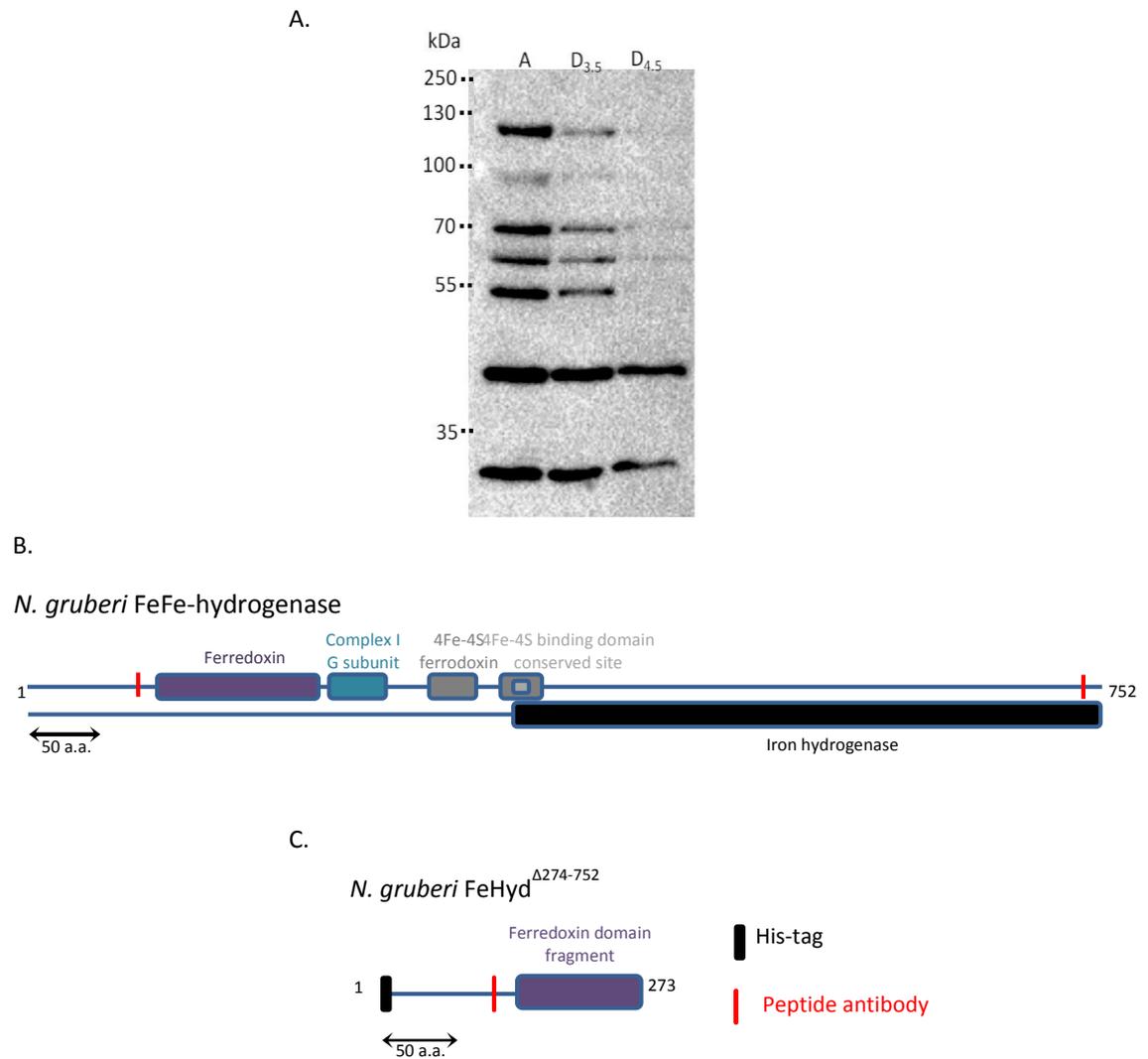


Figure 6.3. Initial characterisation of the *N. gruberi* FeFe-hydrogenase using the anti-peptide antibodies.

A. Protein samples taken during *N. gruberi* differentiation were separated by SDS-PAGE using a 10% gel, blotted onto an Amersham Hybond-P PVDF membrane, and decorated using the anti-peptide antibody at a dilution of 1/250 in blocking buffer A. The lane designated A is the protein sample taken whilst *N. gruberi* was still in its amoebal form, before the initiation of differentiation; lane D_{3.5} is the protein sample taken 3.5 hours into differentiation; lane D_{4.5} is the protein sample taken 4.5 hours into differentiation.

B. A cartoon schematic of the *N. gruberi* FeFe-hydrogenase. Note the two red bars towards the beginning and the end of the blue sequence line, which indicate the relative positions of the peptides that the anti-peptide anti-hydrogenase antibodies are targeted towards.

C. A cartoon schematic of the FeHyd^{Δ274-752} used firstly to test further the usefulness of the anti-peptide anti-hydrogenase antibodies and then expressed as recombinant protein for antibody production (termed anti-FeHyd^{Δ274-752}) in chicken hosts. Highlighted is the N-terminal His-tag (black bar); a peptide sequence used as antigen for the aforementioned anti-peptide antibody (red bar) and the fragment of the ferredoxin domain, present in the *N. gruberi* FeFe-hydrogenase. a.a., amino acids.

Figure 6.4

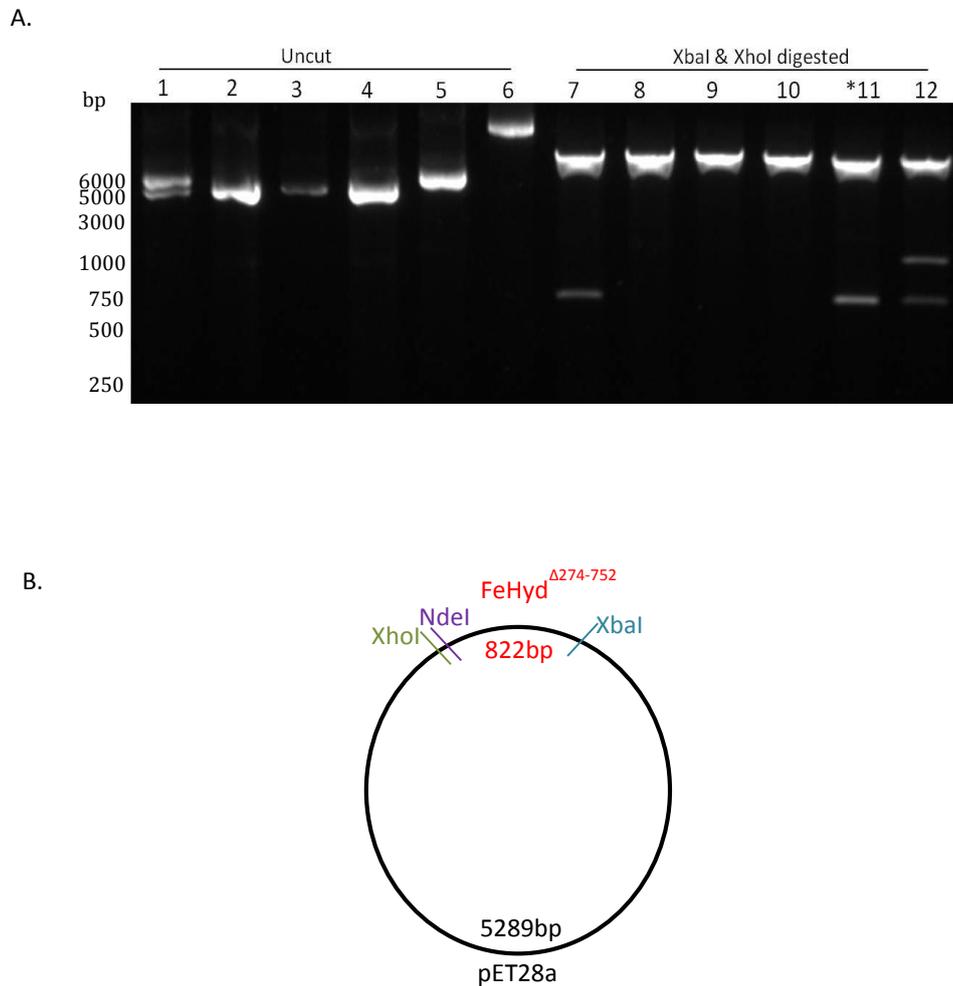


Figure 6.4. Restriction mapping for successful cloning of the coding sequence for FeHyd^{Δ274-752} into the expression vector pET28a.

A. 1% agarose gel electrophoresis of XbaI and XhoI digested plasmid minipreps isolated from colonies transformed with a ligation mixture that contained NdeI-XhoI cut FeHyd^{Δ274-752} coding sequence and NdeI-XhoI cut pET28a vector. Resultant plasmids were then screened by digesting with XbaI and XhoI. For plasmids containing the FeHyd^{Δ274-752} insert the double digest released an insert approximately 800 bp in size, and a vector backbone of approximately 5 kb. Following sequencing, the colony marked with an asterisk was taken forwards for transformation into Rosetta™ 2(DE3)pLysS competent *E. coli* cells (Novagen). Abbreviations: bp, base pairs.

B. A cartoon restriction map for FeHyd^{Δ274-752} plasmid. Highlighted are the cloning sites and the expected sizes of the resultant DNA fragments after digestion with the XbaI and XhoI restriction enzymes.

Figure 6.5

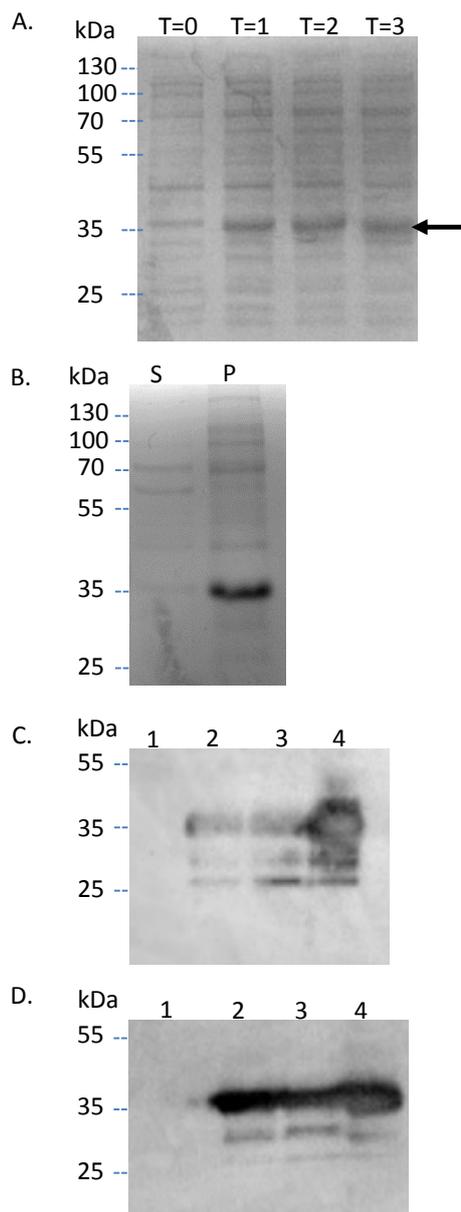


Figure 6.5. Expression and detection of the recombinant FeHyd^{Δ274-752} expressed in *Escherichia coli*.

A. The 10% SDS-PAGE gel of the small scale induction of FeHyd^{Δ274-752} in *E. coli* with IPTG over a 3 hour period; T=0 hour sample was taken before the addition of IPTG. A protein approximately 35 kDa in size not present at T=0 can be seen to be gradually increasing in expression levels over time.

B. The 10% SDS-PAGE gel for the solubility test. The same 35 kDa protein from 6.5 A can be seen distinctly in the pellet (P) sample, indicating that FeHyd^{Δ274-752} is an insoluble protein, presumably expressed in inclusion bodies. S, supernatant.

C. Western blot detection of FeHyd^{Δ274-752} samples taken during small scale induction with IPTG. The Western blot was decorated with the anti-His antibody at a dilution of 1/5000 in blocking buffer A. Lane 1: Sample before induction; Lane 2: Sample taken 1 hour post induction; Lane 3: Sample taken 2 hours post induction; Lane 4: Sample taken 3 hours post induction.

D. Western blot detection of FeHyd^{Δ274-752} samples taken during small scale induction with IPTG. The Western blot was decorated with the NgFeHyd anti-peptide antibody at a dilution of 1/250 in blocking buffer A. Lane 1: Sample before induction; Lane 2: Sample taken 1 hour post induction; Lane 3: Sample taken 2 hours post induction; Lane 4: Sample taken 3 hours post

Figure 6.6

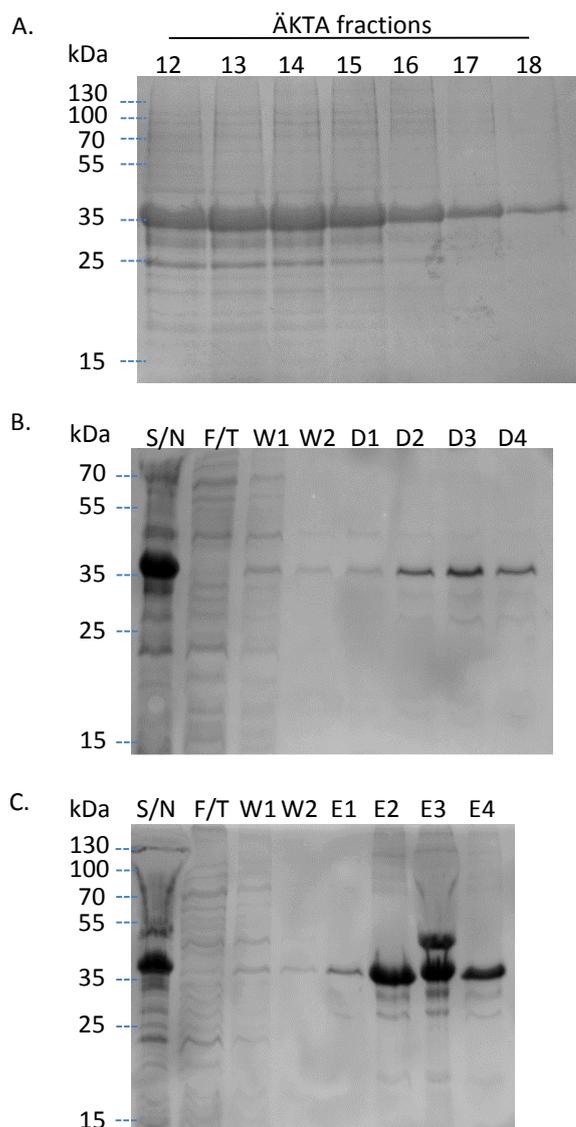


Figure 6.6. Purification of recombinant FeHyd^{Δ274-752} in *Escherichia coli* in preparation for polyclonal antibody production.

A. The 10% SDS-PAGE gel showing the post-purification analysis of ÄKTA-Prime Plus fractions of the elution of FeHyd^{Δ274-752}. Fractions 11-16 were pooled and subjected to further purification.

B. The 8% SDS-PAGE gel showing purification under denaturing conditions of FeHyd^{Δ274-752} using urea and pH changes for elution. This method eliminated the need for SDS-PAGE gels for further purification and the eluted FeHyd^{Δ274-752} is in an ELISA-compatible buffer. Abbreviations: S/N, supernatant; F/T, flow through; W1, wash 1; W2, Wash 2; D1-4, elution buffer D pH 5.9.

C. The 8% SDS-PAGE gel showing the purification under denaturing conditions of FeHyd^{Δ274-752} using urea and pH changes for elution. Abbreviations: S/N, supernatant; F/T, flow through; W1, wash 1; W2, Wash 2; E1-4, elution buffer E pH 4.5. Elution of FeHyd^{Δ274-752} is seen predominantly in the E2 and E3 lanes.

Figure 6.7

A. *N. gruberi* FeHyd^{Δ61-752}:HA:GFP

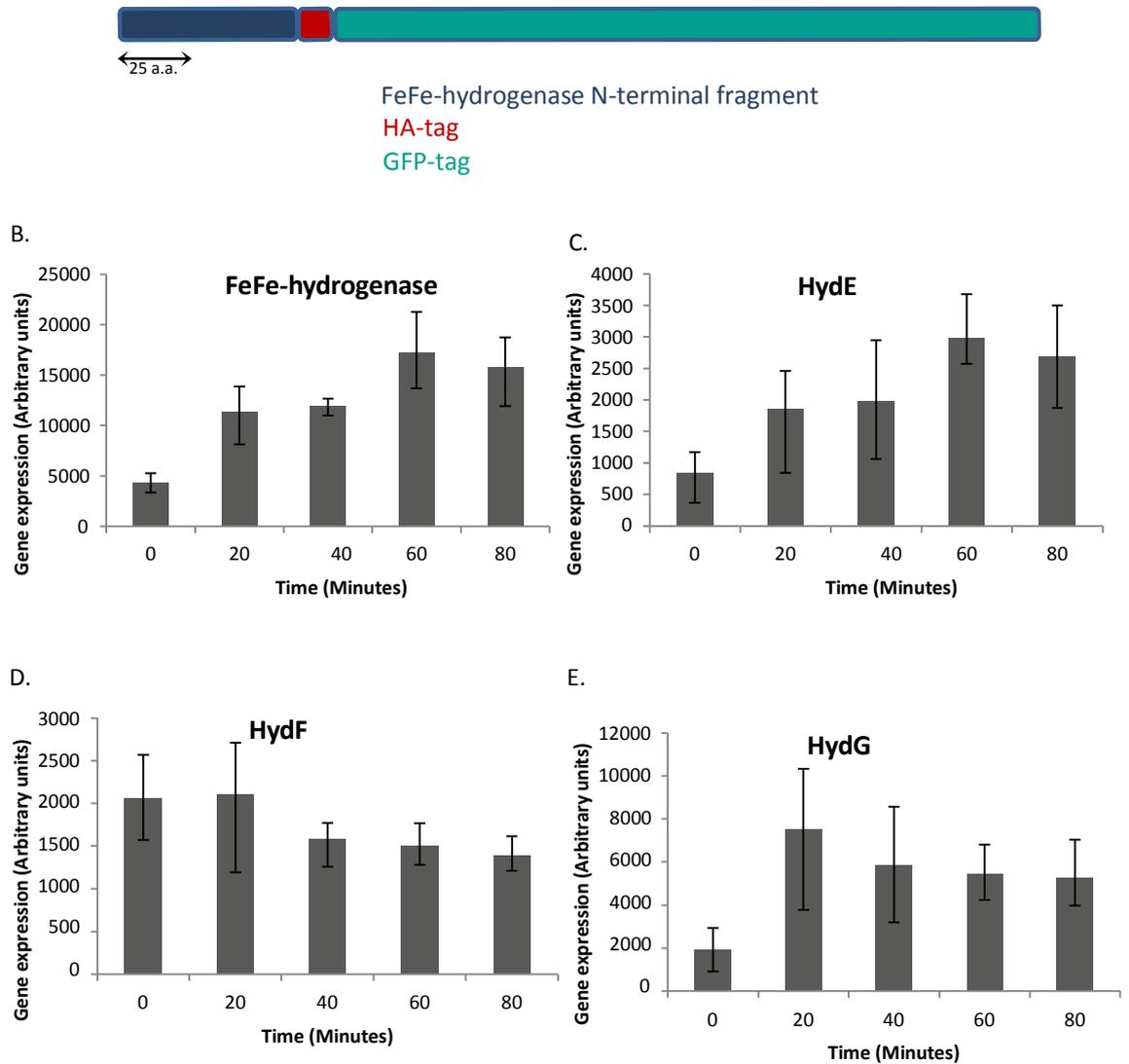


Figure 6.7. Expression of the *N. gruberi* FeFe-hydrogenase and the associated maturases

A. Cartoon of the FeHyd^{Δ61-752}:HA:GFP protein expressed in procyclic *T. brucei*. The putative leader sequence of the FeFe-hydrogenase was fused to a HA-tag followed by a GFP-tag at the C-terminus. B-E. Microarray data analysis suggests the upregulation of mRNA expression for FeFe-hydrogenase and the three associated maturases, HydE, HydF and HydG, as *N. gruberi* differentiates from amoebae to the flagellates. X-axis, Time in minutes; Y-axis, Arbitrary units of expression.

Figure 6.8

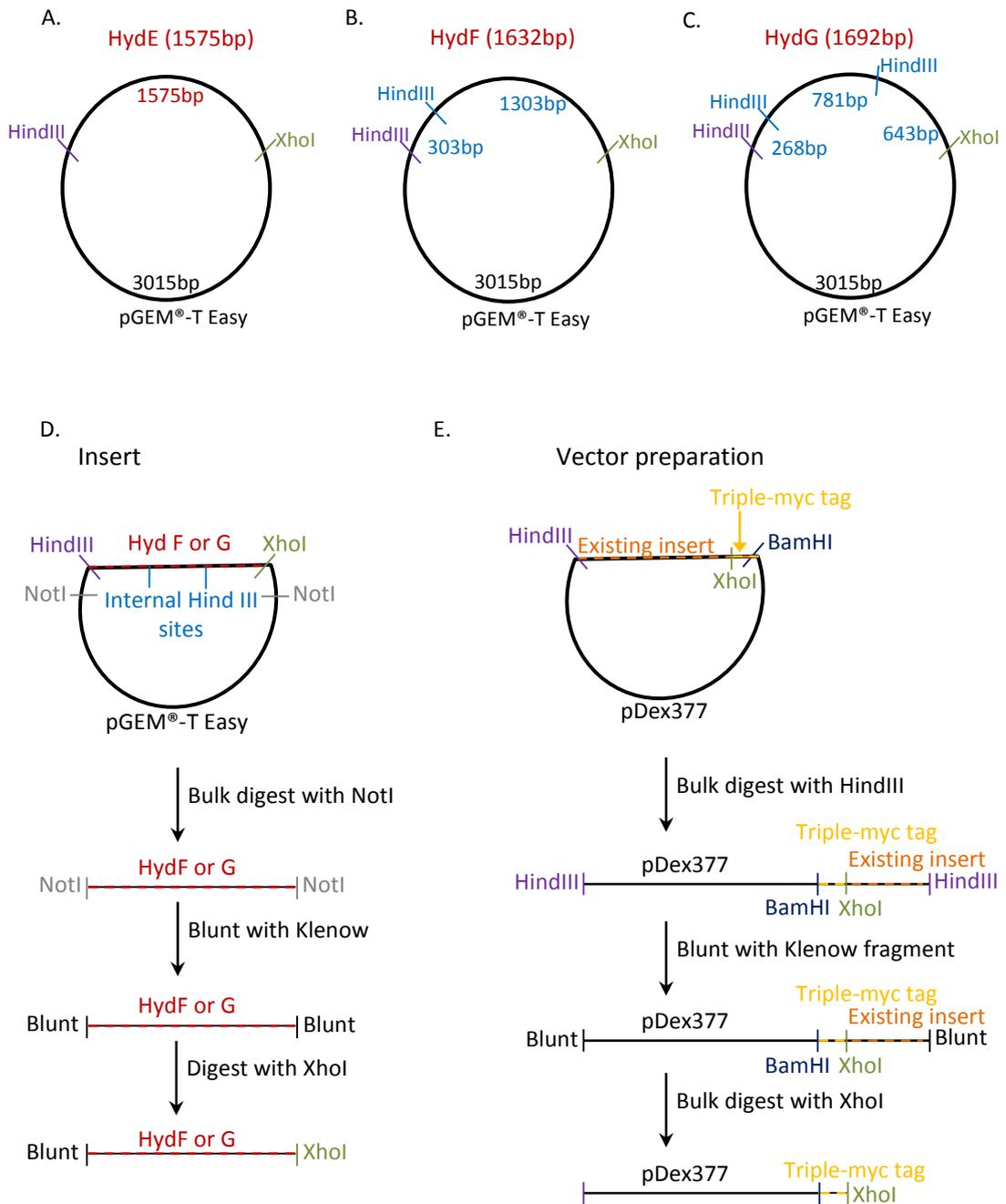


Figure 6.8. Cartoon restriction maps for HydE, HydF and HydG coding sequences with T Easy vectors, and the strategies for cloning genes into the pDex377-derived expression vector in preparation for expression in *T. brucei*.

A-C. Enzyme restriction maps for HydE, HydF and HydG respectively, highlighting the internal HindIII sites in HydF and HydG and fragment sizes expected from restriction mapping.

D-E. Flow diagrams highlighting the use of the Klenow enzyme for blunt end cloning of HydF and HydG into the pDex377 expression vector.

Restriction enzyme labels are colour co-ordinated with positions on the diagrams. bp, base pairs. Cartoons are not drawn to scale.

Figure 6.9

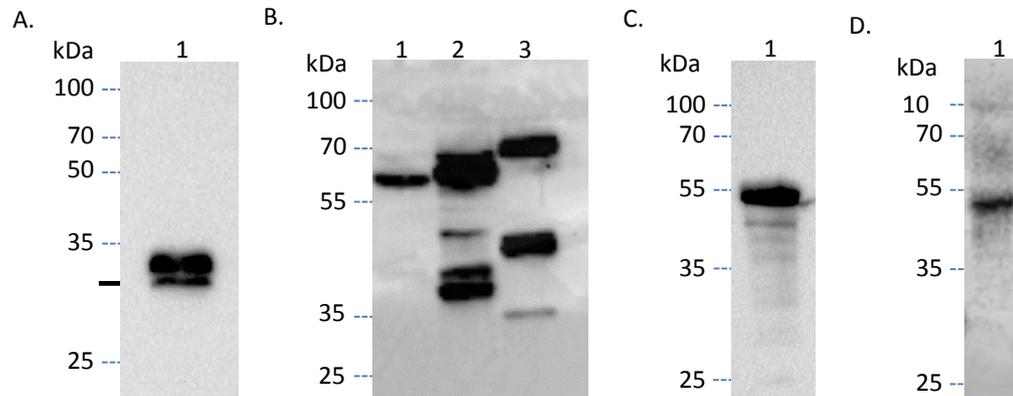


Figure 6.9. Expression of *N. gruberi* enzymes in procyclic 427 *T. brucei*. All protein samples were whole cell lysates resuspended in boiling hot Laemlii buffer in absence of protease inhibitors.

A. Western blot detection of FeHyd^{Δ61-752}:HA:GFP in *T. brucei*. Lane 1 shows expression of two bands ~33 kDa and ~31 kDa respectively, the predicted size of full length protein being 34 kDa. The smaller band could be due to processing of the predicted mitochondrial transit peptide following import of the GFP into the mitochondrion. The blot was decorated with the anti-HA antibody at a dilution of 1/1000 in blocking buffer A.

B. Western blot detection of FeFe-hydrogenase associated maturases, expressed in *T. brucei*. The Western blot was decorated with the anti-myc antibody at a dilution of 1/1000 in blocking buffer A. Lane 1 shows HydE::myc₃ with a expression detected at ~60 kDa which corresponds with the predicted size of HydE (59.6 kDa). Lane 2 shows HydF::myc₃ with expression detected at ~64 kDa, which corresponds to the predicted size of HydF (61 kDa). The array of other smaller proteins is possibly due to proteolytic degradation. Lane 3 shows HydG::myc₃ with a expression detected at ~69 kDa, which is close to the predicted size of HydG (63.9 kDa), again the other smaller proteins detected are likely to be products of proteolysis.

C. Western blot detection of NirK:HA, expressed in *T. brucei*. The Western blot was decorated with the anti-HA antibody at a dilution of 1/1000 in blocking buffer A. Lane 1: NirK:HA construct with expression detected at ~50 kDa, the predicted size of NirK:HA being 44.5 kDa. The other fainter bands in this lane are likely to be products of proteolysis.

D. Western blot analysis of ACS^{Δ101-1100}:HA:GFP, expressed in procyclic *T. brucei*. The Western blot was decorated with the anti-HA antibody at a dilution of 1/1000 in blocking buffer A. Lane 2 shows ACS^{Δ101-1100}:HA:GFP with a band at ~49 kDa, the predicted size of ACS:HA:GFP is 39 kDa.

Figure 6.10

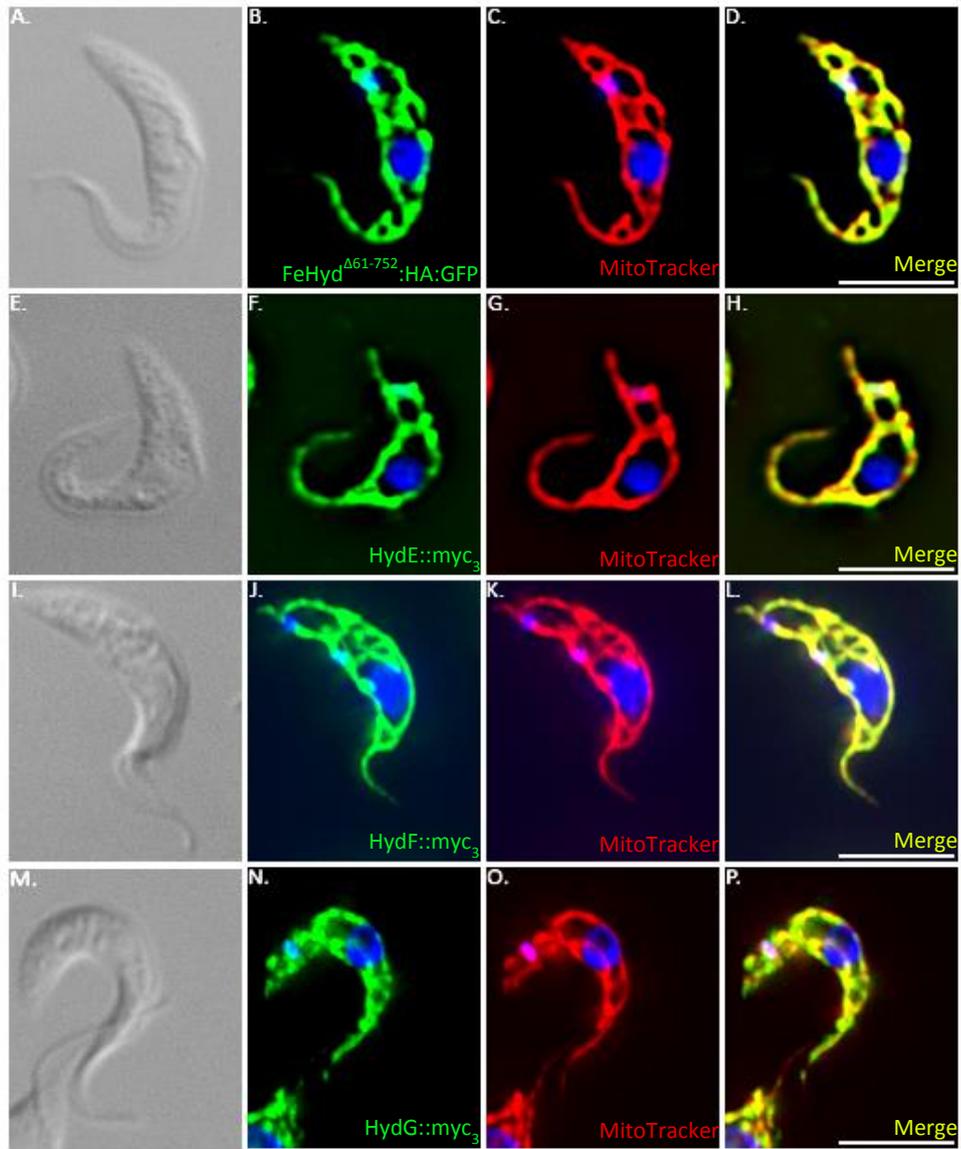


Figure 6.10. Expression and localisation of *N. gruberi* FeHyd^{Δ61-752}:HA:GFP and associated maturases HydE, HydF and HydG in procyclic 427 *T. brucei*.

A-D. Localisation of FeHyd^{Δ61-752}:HA:GFP to the mitochondrion of *T. brucei*. Indirect immunofluorescence used to image the HA-tagged FeHyd^{Δ61-752}:HA:GFP. A: DIC; B: FeHyd^{Δ61-752}:HA:GFP slides decorated with the anti-HA antibody at a dilution of 1/1000 in blocking buffer B, merged with DAPI; C: MitoTracker merged with DAPI; D: Merged.

E-H. Localisation of HydE::myc₃ to the mitochondrion of *T. brucei*. Indirect immunofluorescence used to image the myc-tagged HydE. A: DIC; B: HydE slides decorated with anti-myc at a dilution of 1/1000 in blocking buffer B, merged with DAPI; C: MitoTracker merged with DAPI; D: Merged.

I-L. Localisation of HydF::myc₃ to the mitochondrion of *T. brucei*. Indirect immunofluorescence used to image the myc-tagged HydF. A: DIC; B: HydF slides decorated with anti-myc, merged with DAPI; C: MitoTracker merged with DAPI; D: Merged.

M-P. Localisation of HydG::myc₃ to the mitochondrion of *T. brucei*. Indirect immunofluorescence used to image the myc-tagged HydG. A: DIC; B: HydG slides probed with anti-myc, merged with DAPI; C: MitoTracker merged with DAPI; D: Merged.

Figure 6.11

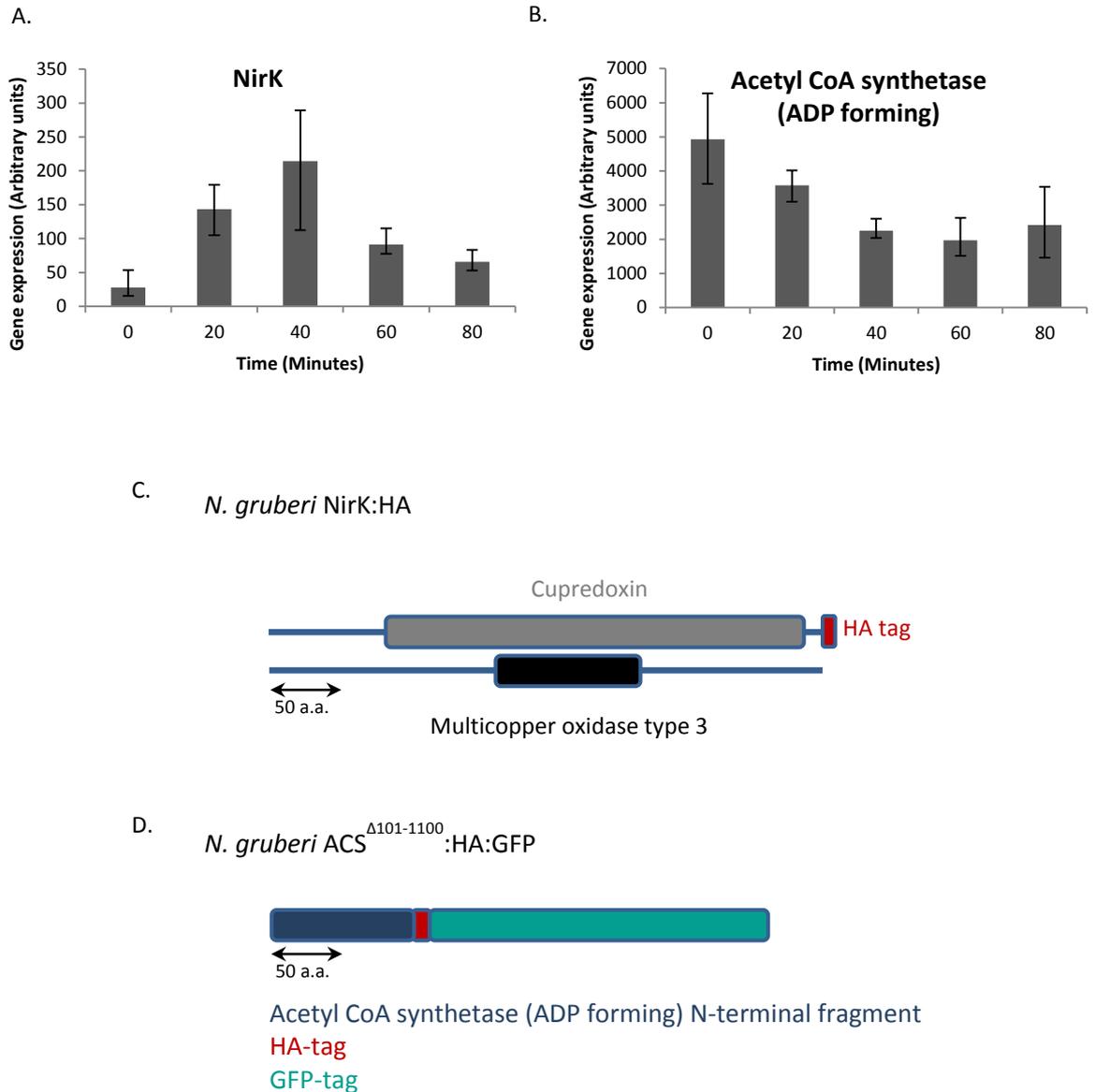


Figure 6.11. Expression of *N. gruberi* nitrite reductase (NirK) and acetyl CoA synthase (ACS^{Δ101-1100}:HA:GFP).

A-B. Microarray data analysis shows upregulation of mRNA expression for NirK and acetyl CoA synthetase (ADP forming) as *N. gruberi* differentiates from amoebae to flagellages. X-axis, Time in minutes; Y-axis, Arbitrary units of expression.

C. Cartoon of the NirK protein expressed in procyclic 427 *T. brucei*. The domain architecture of the full length NirK sequence is shown, along with the HA tag fused at the C-terminus.

D. Cartoon of the ACS^{Δ101-1100}:HA:GFP protein expressed in procyclic 427 *T. brucei*. The first 100 amino acids of the N-terminal sequence were fused to a C-terminal HA-tag followed by a GFP tag.

Figure 6.12

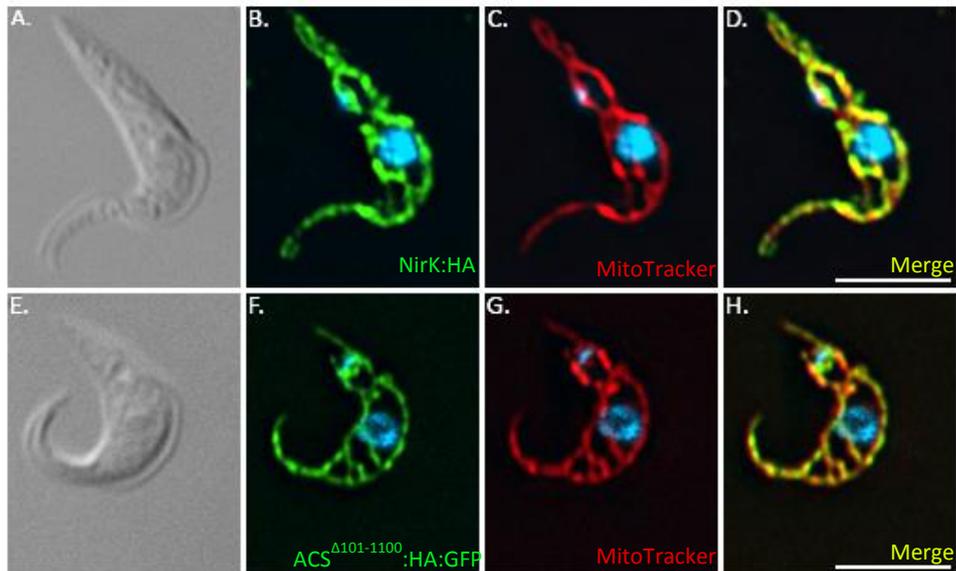


Figure 6.12. Expression and localisation of *N. gruberi* NirK:HA and ACS^{Δ101-1100}:HA:GFP in procyclic 427 *T. brucei*.

A-D. Localisation of NirK:HA to the mitochondrion of *T. brucei*. Indirect immunofluorescence was used to image the HA-tagged NirK:HA. A: DIC; B: NirK:HA slides decorated with the anti-HA antibody at a dilution of 1/1000 in blocking buffer B, merged with DAPI; C: MitoTracker merged with DAPI; D: Merged.

E-H. Localisation of ACS^{Δ101-1100}:HA:GFP to the mitochondrion of *T. brucei*. Indirect immunofluorescence was used to image the HA-tagged ACS^{Δ101-1100}:HA:GFP. A: DIC; B: ACS^{Δ101-1100}:HA:GFP slides decorated with the anti-HA antibody at a dilution of 1/1000 in blocking buffer B, merged with DAPI; C: MitoTracker merged with DAPI; D: Merged.

Chapter 7

General Discussion

Often, classical biochemical and molecular biological techniques utilised in the 20th century focused around identifying and developing assays for purification of enzymes from cells and molecular approaches including degenerate PCR and genomic DNA or cDNA library screening to move from protein sequence to gene identification and cloning (Guengerich et al., 2010). However, with the development and advancements in next-generation sequencing techniques, complete (or near complete) genomes can be rapidly generated for a selected organism, and at relatively low-costs compared to the original Sanger sequencing methods. Subsequently, the sequencing of genomes from hundreds of eukaryotes and thousands of prokaryotes has revealed a plethora of novel gene products for each organism that are of unknown function (Lynch et al., 2014). The challenge then faced for a lot of these uncharacterised gene products is to attempt to uncover the function of a protein *in vivo*, often when an *in vitro* assay of function is not available (Hanson et al., 2010).

The paralogous gene pair *TbHP40* and *TbHP30* are considered to be members of the histidine phosphatase super-family, based on the analysis of the amino acid sequence and the presence of a PGAM domain. Despite *TbHP30* containing a degenerate crucial active site residue, there are still enough sequence similarities to consider it a histidine phosphatase. However, due to the massive diversities amongst known histidine phosphatase family members, in terms of function and substrate-specificity (Rigden, 2008), associations with this family this does not provide much further indication as to the function of *TbHP40* and *TbHP30* within *T. brucei*.

In the data presented here, both *TbHP40* and *TbHP30* are shown to localise to the mitochondria in procyclic *T. brucei*. This provides the first indication that they may have an enzymatic function and be associated with a function or pathway within mitochondria. *TbHP30* is also thought to be the degenerative paralogue of *TbHP40*, and it is not unheard of in trypanosome biology for a catalytically dead paralogue to assume a regulatory role in controlling the active copy of the enzyme. In examples reported thus far, the catalytically-inactive paralogue is required to form a complex with the catalytically active copy, in order to produce a fully-active enzyme (Nguyen et al., 2013; Velez et al., 2013; Willert et al., 2007). The functional S-adenosylmethionine decarboxylase (*AdoMetDC*) gene in *T. brucei* codes for an essential enzyme involved in the polyamine biosynthesis pathway, the genome also

encodes a second AdoMetDC gene that is a paralogue named prozyme (Willert et al., 2007). Both AdoMetDC and prozyme were found to be expressed in procyclic and bloodstream form *T. brucei* by Northern blot analysis. The analysis of AdoMetDC showed it was stimulated by putrescine, but with low efficiency. However, upon the addition of prozyme, AdoMetDC activity increased 1200-fold and the complex formed between the two could be co-purified using a nickel agarose and anion-exchange chromatography. The two subunits could then be detected as a single peak on a Western blot at a 1:1 molar ratio (Willert et al., 2007). In humans, AdoMetDC functions as a homodimer which is regulated by putrescine, but the trypanosomatids have evolved this novel mechanism for regulation of AdoMetDC. By relying on the formation of a heterodimer of AdoMetDC with a catalytically inactive subunit prozyme, which arose due to a tandem duplication of the ancestral gene, activity of AdoMetDC can be regulated by expression levels of prozyme (Velez et al., 2013; Willert et al., 2007).

A further step in the polyamine biosynthesis pathway is also regulated using the same mechanism (Nguyen et al., 2013). The gene duplication of deoxyhypusine synthase (DHS) has resulted in a catalytically impaired copy and a catalytically inactive paralogue. The paralogue is required to oligomerise with the subunits capable of catalytic activity in order to form a fully functioning heterotetrameric enzyme. Both DHS copies were shown to be essential for *T. brucei* growth and infectivity in mice and could be co-purified as a stable heterotetramer consisting of a 2:2 ratio of active:inactive paralogous gene products (Nguyen et al., 2013).

Based on these novel examples of enzyme regulation in *T. brucei*, it is therefore not unreasonable to hypothesise that the degenerate *TbHP30* may have a regulatory role in controlling the function of *TbHP40*, in a similar mechanism to AdoMetDC and DHS. The homology surface modelling of *TbHP40* revealed that the active site of *TbHP40* is located within a relatively deep pocket. This could be indicative of the potential substrates that *TbHP40* may be capable of acting upon; relatively small molecules are more likely to access the recessed active site, rather than large protein substrates which would encounter much more steric hindrance. Unfortunately, since no significant sequence matches within PDB could be found for *TbHP30*, it was impossible to produce a reliable homology model for this protein. Dimerization of histidine phosphatase subunits in order to form active enzymes is not uncommon. For example, the TIGAR histidine phosphatase from *Danio rerio* (freshwater zebrafish), upon which *TbHP40* was partially modelled, functions as a homodimer *in vivo* (Li and Jögl, 2009). The structure of zebrafish TIGAR was solved using x-ray crystallography and has the classic histidine phosphatase fold consisting of a central six-stranded mixed β -strand

core that is flanked either side by two α -helices (Li and Jögl, 2009). The β 5 strand of one TIGAR monomer interacts with the β 5 strand of another to form a symmetrical homodimer (Li and Jögl, 2009). However, due to *TbHP30* being unsuitable for homology modelling using the MOE software, it was subsequently unfeasible to attempt homology modelling of both *TbHP40* and *TbHP30* as a heterodimer. Due to this, the homology model presented in this thesis should be reviewed with a certain amount of caution. If *TbHP40* and *TbHP30* do interact as a heterodimer within cells, the binding of the two subunits could cause substantial conformational changes to the configuration of the active site region. It was perhaps just luck that *TbHP40* was identified as the sequence to study from all of the identified kinetoplast orthologues. Attempts to model orthologues from a range of other organisms, including *Leishmania major*, *Naegleria gruberi* and *Arabidopsis thaliana*, were either unsuitable for homology modelling (as was the case with *TbHP30*) or produced unviable models, *i.e.* when surface modelling the active site, it was found that the 'RHG' triad was buried deep within the molecule, with no obvious access to potential substrates from the surface.

Searching for proteins that potentially interact with *TbHP40* provided the first evidence that *TbHP40* and *TbHP30* do interact with one-another within procyclic *T. brucei*. It was surprising that only *TbHP40* itself was biotinylated and identified in the mass spectrometry data from the BioID experiments, but this could be due to the presence of the large BirA* tag at the C-terminus of ectopically-expressed *TbHP40*. At 35 kDa in size, the addition of the BirA* tag could be causing steric hindrance problems, especially if a *TbHP30* subunit is required to form a heterodimer with *TbHP40*. Interestingly, the data from the co-immunoprecipitation experiments, when analysed by mass spectrometry, identified both *TbHP40* and *TbHP30* as interacting partners, but no other candidate proteins. Both *TbHP40* and *TbHP30* peptides were recovered from the mass spectrometry data with good sequence coverage from the N- and C- terminals for *TbHP40*, however, no sequence coverage to the N-terminus of *TbHP30* was found. This is consistent with the data from the proteomics studies of both bloodstream and procyclic *T. brucei*, all of which also recovered *TbHP30* peptides (Fisk et al., 2013; Niemann et al., 2013; Panigrahi et al., 2009; Urbaniak et al., 2012; Urbaniak et al., 2013). None of these studies found any peptide evidence that corresponded with the long N-terminus of *TbHP30*. *TbHP30* is predicted to have (a presumably cleaved) N-terminal mitochondrial import leader sequence. Yet in Western analyses, the protein detected is always close to the expected size that would correspond to the full length *TbHP30* with the addition of a C-terminal tag. In regards to *TbHP40*, only one proteomic dataset, the one

undertaken by Niemann et al., (2013), reported peptides that corresponded to *TbHP40* in procyclic *T. brucei* only. *TbHP30* protein may be more abundant in the cell than *TbHP40*, in both the bloodstream and procyclic stages of the lifecycle. This in turn may suggest that *TbHP40* is the limiting factor, and the expression levels of *TbHP40* regulates how much 'active' protein is available within a cell ('active' in the sense that *TbHP40* and *TbHP30* subunits combine to form a functional heterodimer). On the other hand, it may just be that *TbHP30* is a better substrate for trypsin digestion during the sample preparation, or that *TbHP30*-derived tryptic peptides ionise more readily in the mass spectrometer and thus, the protein is therefore easier to detect than *TbHP40*.

The patchy distribution of candidate orthologues and homologues to *TbHP40* and *TbHP30* is unusual and intriguing. *TbHP30* is exclusively found within the kinetoplastids, for which sequence data is available. Its presence in the free-living *Bodo saltans* indicates that this gene duplication arose before many lineages adapted to parasitism, but the retention of both copies in all of the parasitic lineages analysed suggests that both have a function within a cell, and this could be the first indication that they may be essential or at least serve an important function, if the generalisation that parasites streamline their metabolism is to be believed. The heterolobosean *Naegleria gruberi* was identified to contain two candidate orthologues to *TbHP40*, even though one of the predicted proteins identified lacked a start codon (gene models predicted for *Naegleria* are not always perfect (Fritz-Laylin et al., 2010)). It still had the 'RHG' catalytic triad. As previously mentioned, these orthologues were not suitable for homology modelling, but the published microarray dataset is available for *N. gruberi*, as it differentiates from the amoebal form to the flagellate form (Fritz-Laylin and Cande, 2010). Further analysis of the microarray dataset suggests both of these genes may be upregulated during differentiation, particularly the gene for which a start methionine is not predicted for the protein product (Figure 7.1).

The attempts to knock-out *TbHP40* and *TbHP30* provide further indication that both genes are essential in procyclic *T. brucei*, since it proved so difficult to remove the *TbHP40/TbHP30* locus on the second allele. Even after introducing ectopically-expressed copies of both *TbHP40* and *TbHP30* on the minichromosomes of *T. brucei*, it still proved too difficult to produce a null mutant. However, both of these ectopically-expressed copies contained C-terminal triple-myc tags, which like the BirA* tag, could in some way interfere with the way the *TbHP40* and *TbHP30* protein folds or potentially block interactions with other potential partners *in vivo*, masking the true functions of the protein. In order to proceed with these experiments if time had permitted, the next step would be to use site-directed mutagenesis

to remove the C-terminal tags from the ectopically-expressed *TbHP40* and *TbHP30*, and then attempt the second round knock-outs again with just native proteins being expressed. These potential problems with the myc tags may also provide an explanation for why no additional candidate interacting partners were identified in the co-immunoprecipitation experiments.

If one is to consider how next to move forward in determining the function of *TbHP40* and *TbHP30*, it is maybe worth reflecting on examples of proteins where function was initially elucidated in the absence of a functional assay. Members of the cytochrome P450 superfamily utilise haem when oxidising a substrate, and at least 58 cytochrome P450 enzymes have been identified in humans (Stark and Guengerich, 2007). Some cytochrome P450 enzymes can be very promiscuous with respect to the substrates that they will act upon and collectively, the range of physiological substrates that can be subjected to catalysis by the cytochrome P450 family is very broad. The function of the cytochrome P450 enzymes in fungi such as *Claviceps purpurea* and bacteria such as *Streptomyces coelicolor* were elucidated firstly by identifying sequence homology with the mammalian cytochrome P450 enzymes and further characterised using gene knockouts and heterologous expression experiments (Haarmann et al., 2006; Zhao and Waterman, 2007), emphasising perhaps the relevance of the techniques I employed to try and work out the function of *TbHP40* and *TbHP30*, albeit that their function still remains unclear.

In other work, phylogenetic profiling showing the strict co-occurrence of frataxin with two other proteins indicated a role for frataxin, previously an enigmatic protein, in mitochondrial Fe-S cluster assembly (Huynen et al., 2001). In many organisms, including trypanosomes, frataxin is an essential component of the Fe-S cluster assembly machinery (Long et al., 2008b). Recently, proteome microarray analyses have led to the identification of YcgC, an *Escherichia coli* protein, as a completely novel class of protein deacetylase. No clue to the function of this protein was evident from its amino acid sequence (Tu et al., 2015). Similarly, phylogenetic profiling was key to the identification of enzymes responsible for the synthesis of isopentenyl diphosphate via an alternative pathway to the classic mevalonate pathway found in many organisms, and which in humans is most classically associated with cholesterol biosynthesis (statins are powerful inhibitors of the rate-limiting step of the human mevalonate pathway) (Dellas et al., 2013). Thus, a range of techniques can be blended to work out the function of proteins first identified as part of genome sequencing projects. Given (i) the striking differences in the metabolic pathways active in bloodstream versus procyclic *T. brucei* mitochondria – relatively few of the classical mitochondrial pathways are believed to be active in bloodstream mitochondria (Bringaud et

al., 2006) – and (ii) the indication that *TbHP40* and *TbHP30* are expressed in bloodstream and procyclic mitochondria, if *TbHP40* and *TbHP30* take an intermediary metabolite as substrate then it may be possible to look through the inventory of metabolic pathways predicted to be active in the mitochondrion of bloodstream *T. brucei* and ask if any pathway is missing the identification of enzymes catalysing reactions known to be required for that pathway. If any such missing reaction is likely to be catalysed by a mutase or phosphatase, then *TbHP40* and *TbHP30* would be prime candidates for the missing enzyme(s). Alternatively, if *TbHP40* and/or *TbHP30* dephosphorylate proteins, and thus act within a signalling network, then it is worth looking through (a) phosphoproteomic data published by Urbaniak, Zilberstein and others (Urbaniak et al., 2013; Zilberstein, 2015) to determine the extent to which known mitochondrial proteins are subject to phosphorylation and (b) published mitochondrial proteomes to assess the abundance and diversity of protein kinases and phosphatases that would form the basis of mitochondrial signalling networks. In *Toxoplasma gondii*, bioinformatics-led assembly of a secreted repertoire of protein kinases has been used to provide insight into the host-parasite interaction (Peixoto et al., 2010).

The final chapter of my results provided experimental support for the mitochondrial localisation of FeFe-hydrogenase and other enzymes of anaerobic metabolism predicted from the *N. gruberi* genome. One might therefore suggest an additional class of mitochondrion to those proposed by Muller et al., (2012): the class VI mitochondrion capable of oxidative phosphorylation, anaerobic hydrogen production, and other anaerobic forms of ATP production (e.g. NO_2^- respiration). However, perhaps the more pertinent question is whether the ubiquitous, if still little studied, protist *N. gruberi* is highly unusual with regard to its metabolic flexibility or if it merely represents a tip of an iceberg with respect to metabolic flexibility amongst free-living eukaryotic heterotrophs? Currently, amongst unicellular eukaryotes, genome sequencing efforts have been skewed towards organisms of biotechnological relevance or pathogens. Amongst the algae, genome annotations of *Chlamydomonas reinhardtii* and *Cyanophora paradoxa* have revealed extensive potential for anaerobic metabolism (Mus et al., 2007; Price et al., 2012). There are also a wide variety of endosymbiotic associations between protists that involve elaborate metabolic dependencies (Nowack and Melkonian, 2010) and examples of where protists such as the ciliate *Histiobalantium natans* sequester chloroplasts to sustain an oxidative metabolism within an anoxic environment (Esteban et al., 2009). Thus, there is every reason to believe that as the genome sequencing of an ever-broadening range of free-living eukaryotes is completed, more metabolically-flexible protists like *Naegleria* will be

discovered. Many eukaryotes remain unculturable and thus our appreciation of biological diversity in the protists is maybe incomplete. Recently, *Subulatomonas tetraspora* was identified as a member of a previously unrecognised major group of eukaryotes (Katz et al., 2011), *Tsukubamonas globosa* was identified as a divergent member of the Discoba (Kamikawa et al., 2014; Yabuki et al., 2011), and the rappemonads were identified as an unculturable diverse group of plastid-bearing algae (Kim et al., 2011). These papers perhaps illustrate how much is yet to be learnt with regard to eukaryotic diversity. It will also be important to learn how *N. gruberi* makes use of its anaerobic metabolic potential, although this might not be straight forward as the genome sequenced strain of *Naegleria* has been laboratory entrained for over 50 years and currently no method for genetic transformation of *Naegleria* is available.

Figure 7.1

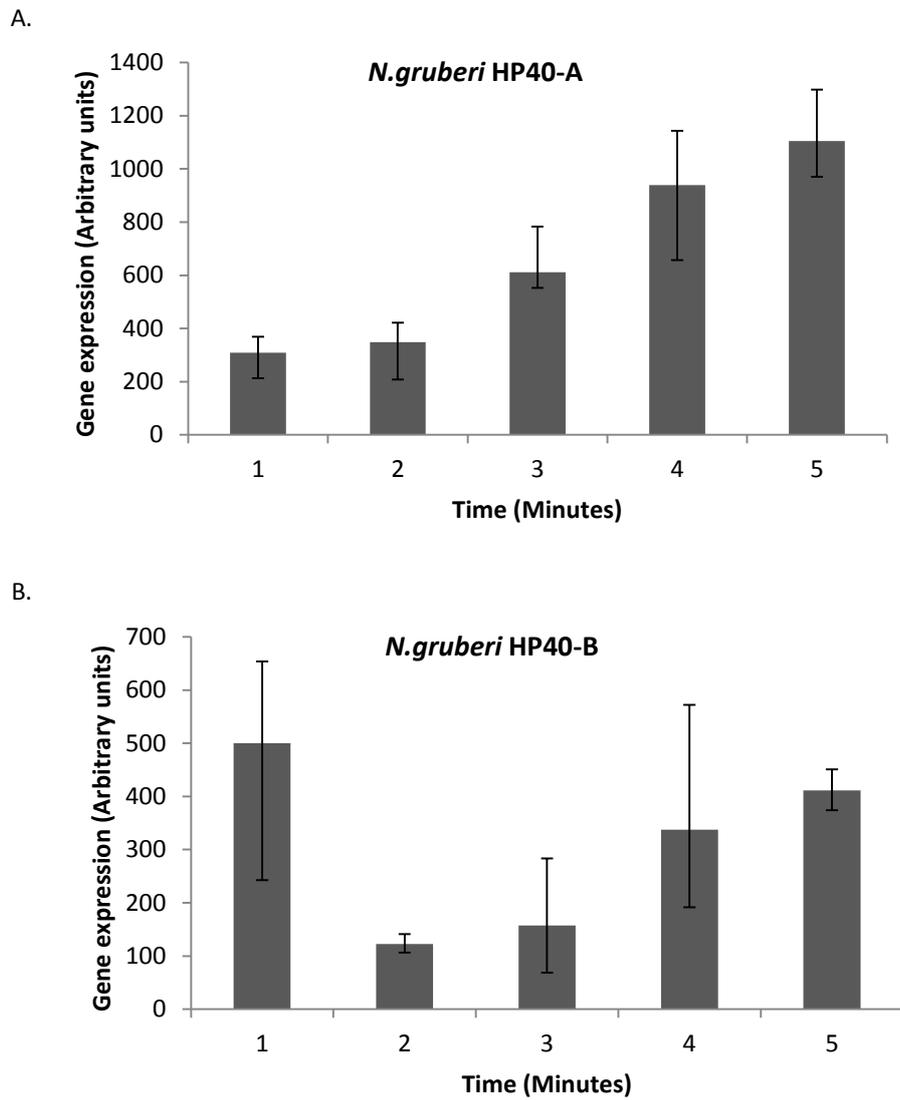


Figure 7.1. mRNA abundance of the candidate *Naegleria gruberi* HP40 orthologues. A-B. Microarray analysis suggests upregulation of the two identified *Tb*HP40 candidate orthologues, referred to as '*N. gruberi* HP40-A' (the sequence that lacks a start codon) and '*N. gruberi* HP40-B' in amino acid sequence alignments in Chapter 3, as *N. gruberi* differentiates from the amoeba to the flagellate. X-axis, Time in minutes; Y-axis, Arbitrary units of expression.

Chapter 8

Appendices

8.1 Appendix I - abbreviations from Figure 1.12

AA, amino acid; AOB, amino oxobutyrate; 1,3BPGA, 1,3-bisphosphoglycerate; C, cytochrome c; Cit, citrate; CoASH, coenzyme A; DHAP, dihydroxyacetone phosphate; F-6-P, fructose 6-phosphate; FBP, fructose 1,6-bisphosphate; G-3-P, glyceraldehyde 3-phosphate; G-6-P, glucose 6-phosphate; GLU, glutamate; Gly-3-P, glycerol 3-phosphate; IsoCit, isocitrate; 2Ket, 2-ketoglutarate; OA, 2-oxoacid; Oxac, oxaloacetate; PEP, phosphoenolpyruvate; 3-PGA, 3-phosphoglycerate; Pi, inorganic phosphate; P_i, inorganic pyrophosphate; SAG, glutamate - semialdehyde; SucCoA, succinyl-CoA; T[SH]₂, reduced form of trypanothione; UQ, ubiquinone pool. Enzymes are: 1, hexokinase; 2, glucose-6-phosphate isomerase; 3, phosphofructokinase; 4, aldolase; 5, triose-phosphate isomerase; 6, glycerol-3-phosphate dehydrogenase; 7, glycerol kinase; 8, glyceraldehyde-3-phosphate dehydrogenase; 9, glycosomal phosphoglycerate kinase; 10, cytosolic phosphoglycerate kinase; 11, phosphoglycerate mutase; 12, enolase; 13, pyruvate kinase; 14, phosphoenolpyruvate carboxykinase; 15, pyruvate phosphate dikinase; 16, glycosomal malate dehydrogenase; 17, cytosolic (and glycosomal) fumarase (F_{Hc}); 18, glycosomal NADH-dependent fumarate reductase; 19, mitochondrial fumarase (F_{Hm}); 20, mitochondrial NADH-dependent fumarate reductase; 21, glycosomal adenylate kinase; 22, malic enzyme; 23, unknown enzyme; 24, alanine aminotransferase; 25, pyruvate dehydrogenase complex; 26, acetate:succinate

8.2 Appendix II – restriction maps for plasmids used in this study

Figure B.1

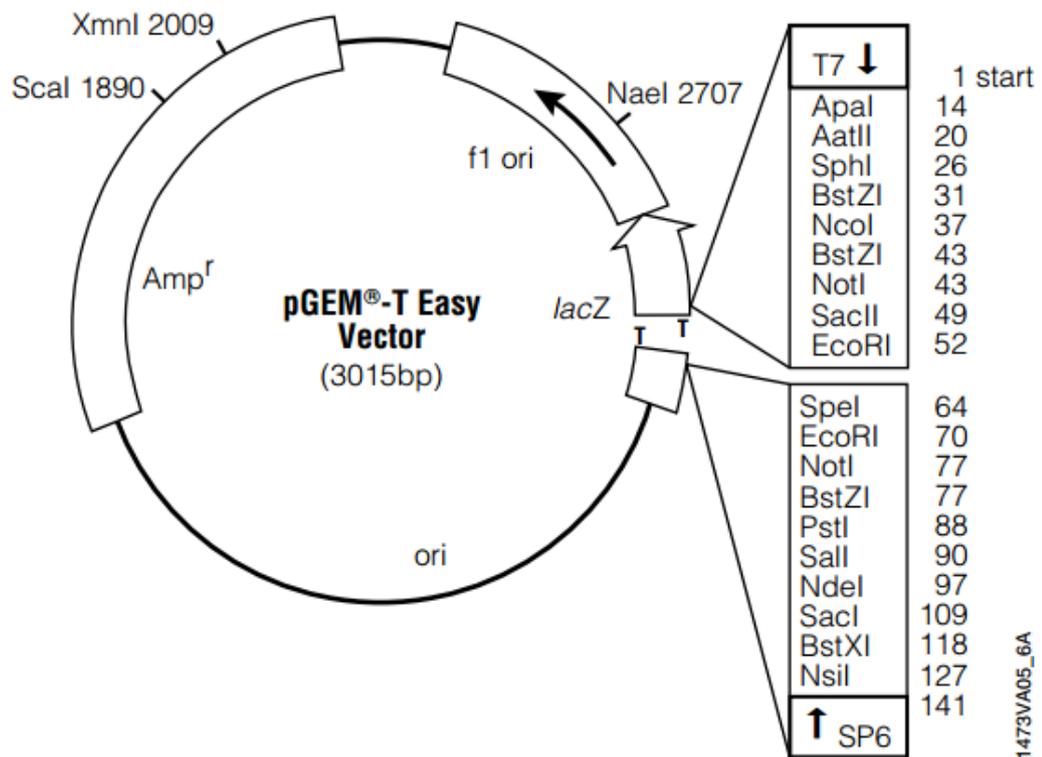


Figure B.1. The pGEM®-T Easy Vector map used for routine cloning. Reproduced from the Promega Technical Manual - pGEM®-T and pGEM®-T Easy Vector Systems.

Figure B.2

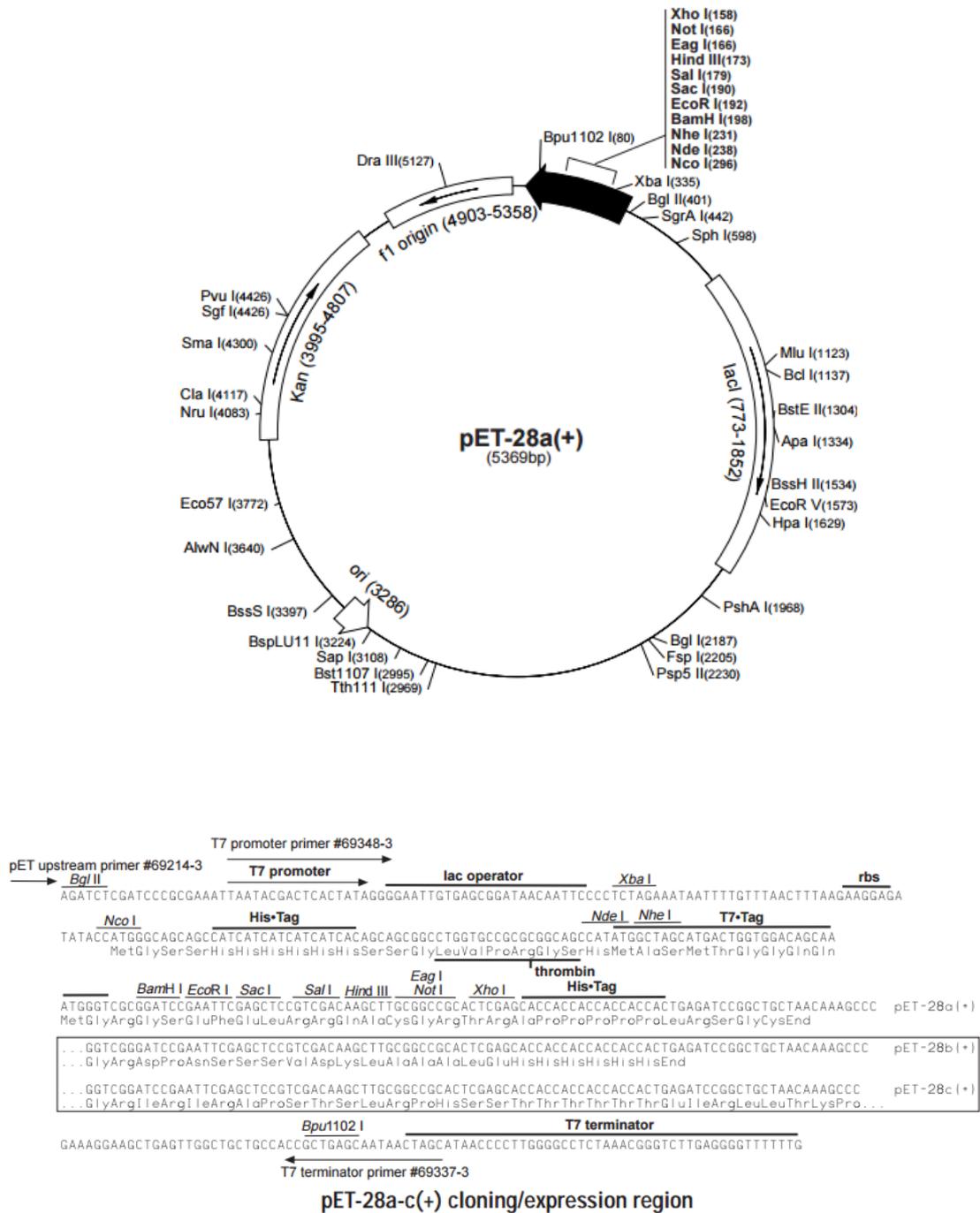


Figure B.2. The pET-28a vector map and a detailed map of the cloning/expression region, highlighting the presence of the His•Tag. Reproduced from Novagen pET-28a-c(+) Vectors data sheet.

Figure B.3

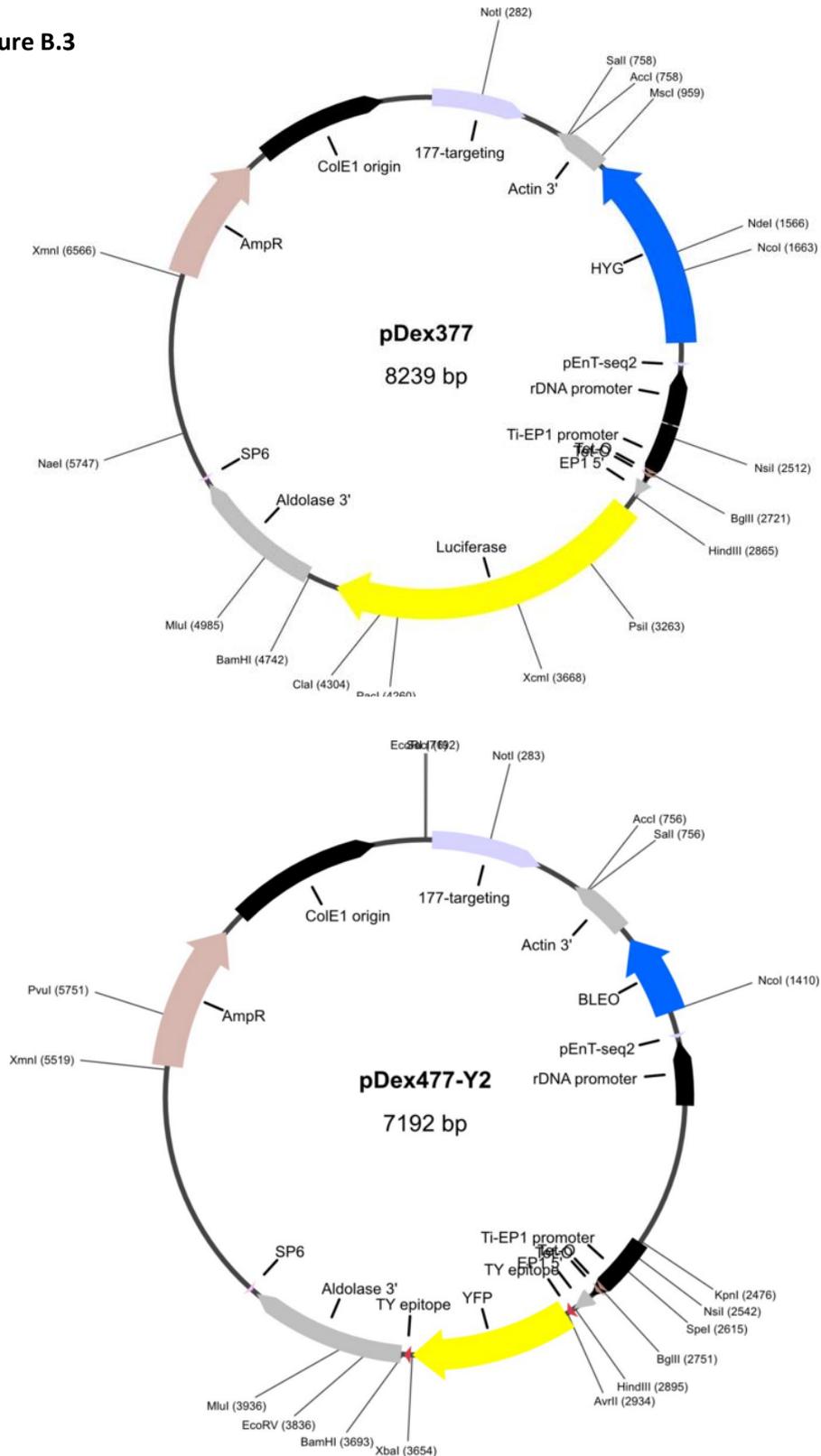


Figure B.3. The pDex-based vector maps. Reproduced from the Wickstead lab. Website: <http://www.wicksteadlab.co.uk/vectors.shtml>

A. The pDex 377 vector map, as described by Kelly et al. (2007).

B. The pDex477-Y2 vector map, as by Kelly et al. (2007). This vector was kindly provided by Bill Wickstead (University of Nottingham).

Chapter 9

References

- Acestor, N., A. K. Panigrahi, Y. Ogata, A. Anupama, and K. D. Stuart, 2009, Protein composition of *Trypanosoma brucei* mitochondrial membranes: Proteomics, v. 9, p. 5497-508.
- Adl, S. M., A. G. Simpson, C. E. Lane, J. Lukes, D. Bass, S. S. Bowser, M. W. Brown, F. Burki, M. Dunthorn, V. Hampl, A. Heiss, M. Hoppenrath, E. Lara, L. Le Gall, D. H. Lynn, H. McManus, E. A. Mitchell, S. E. Mozley-Stanridge, L. W. Parfrey, J. Pawlowski, S. Rueckert, R. S. Shadwick, C. L. Schoch, A. Smirnov, and F. W. Spiegel, 2012, The revised classification of eukaryotes: J Eukaryot Microbiol, v. 59, p. 429-93.
- Akhmanova, A., F. G. Voncken, K. M. Hosea, H. Harhangi, J. T. Keltjens, H. J. op den Camp, G. D. Vogels, and J. H. Hackstein, 1999, A hydrogenosome with pyruvate formate-lyase: anaerobic chytrid fungi use an alternative route for pyruvate catabolism: Mol Microbiol, v. 32, p. 1103-14.
- Aslett, M., C. Aurrecochea, M. Berriman, J. Brestelli, B. P. Brunk, M. Carrington, D. P. Depledge, S. Fischer, B. Gajria, X. Gao, M. J. Gardner, A. Gingle, G. Grant, O. S. Harb, M. Heiges, C. Hertz-Fowler, R. Houston, F. Innamorato, J. Iodice, J. C. Kissinger, E. Kraemer, W. Li, F. J. Logan, J. A. Miller, S. Mitra, P. J. Myler, V. Nayak, C. Pennington, I. Phan, D. F. Pinney, G. Ramasamy, M. B. Rogers, D. S. Roos, C. Ross, D. Sivam, D. F. Smith, G. Srinivasamoorthy, C. J. Stoeckert, Jr., S. Subramanian, R. Thibodeau, A. Tivey, C. Treatman, G. Velarde, and H. Wang, 2010, TriTrypDB: a functional genomic resource for the Trypanosomatidae: Nucleic Acids Res, v. 38, p. D457-62.
- Atteia, A., R. van Lis, A. G. Tielens, and W. F. Martin, 2013, Anaerobic energy metabolism in unicellular photosynthetic eukaryotes: Biochim Biophys Acta, v. 1827, p. 210-23.
- Averner, M., and C. Fulton, 1966, Carbon dioxide: signal for excystment of *Naegleria gruberi*: J Gen Microbiol, v. 42, p. 245-55.
- Baldauf, S. L., 2003, The deep roots of eukaryotes: Science, v. 300, p. 1703-6.
- Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle, 2000, A kingdom-level phylogeny of eukaryotes based on combined protein data: Science, v. 290, p. 972-7.
- Barbera, M. J., I. Ruiz-Trillo, J. Y. Tufts, A. Bery, J. D. Silberman, and A. J. Roger, 2010, *Sawyeria marylandensis* (Heterolobosea) has a hydrogenosome with novel metabolic properties: Eukaryot Cell, v. 9, p. 1913-24.
- Baumgartner, M., S. Eberhardt, J. F. De Jonckheere, and K. O. Stetter, 2009, *Tetramitus thermacidophilus* n. sp., an amoeboflagellate from acidic hot springs: J Eukaryot Microbiol, v. 56, p. 201-6.
- Beagley, C. T., R. Okimoto, and D. R. Wolstenholme, 1998, The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code: Genetics, v. 148, p. 1091-108.
- Bensaad, K., A. Tsuruta, M. A. Selak, M. N. Vidal, K. Nakano, R. Bartrons, E. Gottlieb, and K. H. Vousden, 2006, TIGAR, a p53-inducible regulator of glycolysis and apoptosis: Cell, v. 126, p. 107-20.
- Berriman, M., E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, D. C. Bartholomeu, N. J. Lennard, E. Caler, N. E. Hamlin, B. Haas, U. Bohme, L. Hannick, M. A. Aslett, J. Shallom, L. Marcello, L. Hou, B. Wickstead, U. C. Alsmark, C. Arrowsmith, R. J. Atkin, A. J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T. J. Chillingworth, C. Churcher, L. N. Clark, C. H. Corton, A. Cronin, R. M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M. C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B. R. Harris, H. Hauser, J. Hostetler, A. Ivens, K. Jagels, D. Johnson, J. Johnson, K. Jones, A. X.

- Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. Macleod, P. J. Mooney, S. Moule, D. M. Martin, G. W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C. S. Peacock, J. Peterson, M. A. Quail, E. Rabinowitsch, M. A. Rajandream, C. Reitter, S. L. Salzberg, M. Sanders, S. Schobel, S. Sharp, M. Simmonds, A. J. Simpson, L. Tallon, C. M. Turner, A. Tait, A. R. Tivey, S. Van Aken, D. Walker, D. Wanless, S. Wang, B. White, O. White, S. Whitehead, J. Woodward, J. Wortman, M. D. Adams, T. M. Embley, K. Gull, E. Ullu, J. D. Barry, A. H. Fairlamb, F. Opperdoes, B. G. Barrell, J. E. Donelson, N. Hall, C. M. Fraser, et al., 2005, The genome of the African trypanosome *Trypanosoma brucei*: Science, v. 309, p. 416-22.
- Boxma, B., R. M. de Graaf, G. W. van der Staay, T. A. van Alen, G. Ricard, T. Gabaldon, A. H. van Hoek, S. Y. Moon-van der Staay, W. J. Koopman, J. J. van Hellemond, A. G. Tielens, T. Friedrich, M. Veenhuis, M. A. Huynen, and J. H. Hackstein, 2005a, An anaerobic mitochondrion that produces hydrogen: Nature, v. 434, p. 74-9.
- Boxma, B., R. M. d. Graaf, G. W. M. v. d. Staay, T. A. v. Alen, G. Ricard, T. Gabaldón, A. H. A. M. v. Hoek, S. Y. M.-v. d. Staay, W. J. H. Koopman, J. J. v. Hellemond, A. G. M. Tielens, T. Friedrich, M. Veenhuis, M. A. Huynen, and J. H. P. Hackstein, 2005b, An anaerobic mitochondrion that produces hydrogen: Nature, v. 434, p. 74-79.
- Boxma, B., G. Ricard, A. H. van Hoek, E. Severing, S. Y. Moon-van der Staay, G. W. van der Staay, T. A. van Alen, R. M. de Graaf, G. Cremers, M. Kwantes, N. R. McEwan, C. J. Newbold, J. P. Jouany, T. Michalowski, P. Pristas, M. A. Huynen, and J. H. Hackstein, 2007, The [FeFe] hydrogenase of *Nyctotherus ovalis* has a chimeric origin: BMC Evol Biol, v. 7, p. 230.
- Brayton, K. A., A. O. Lau, D. R. Herndon, L. Hannick, L. S. Kappmeyer, S. J. Berens, S. L. Bidwell, W. C. Brown, J. Crabtree, D. Fadrosch, T. Feldblum, H. A. Forberger, B. J. Haas, J. M. Howell, H. Khouri, H. Koo, D. J. Mann, J. Norimine, I. T. Paulsen, D. Radune, Q. Ren, R. K. Smith, Jr., C. E. Suarez, O. White, J. R. Wortman, D. P. Knowles, Jr., T. F. McElwain, and V. M. Nene, 2007, Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa, PLoS Pathog, v. 3: United States, p. 1401-13.
- Bringaud, F., L. Riviere, and V. Coustou, 2006, Energy metabolism of trypanosomatids: adaptation to available carbon sources: Mol Biochem Parasitol, v. 149, p. 1-9.
- Brocks, J., and R. Summons, 2003, 8.03 – Sedimentary Hydrocarbons, Biomarkers for Early Life, p. 63–115.
- Brown, M. W., J. D. Silberman, and F. W. Spiegel, 2012, A contemporary evaluation of the acrasids (Acrasidae, Heterolobosea, Excavata): Eur J Protistol, v. 48, p. 103-23.
- Brown, R. W. B., 2011, Post-Genomic Investigations of Metabolism in the African Trypanosomes, Lancaster University, 388 p.
- Brun, R., and Schonenberger, 1979, Cultivation and in vitro cloning or procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. Short communication: Acta Trop, v. 36, p. 289-92.
- Bui, E. T., and P. J. Johnson, 1996, Identification and characterization of [Fe]-hydrogenases in the hydrogenosome of *Trichomonas vaginalis*: Mol Biochem Parasitol, v. 76, p. 305-10.
- Burger, G., L. Forget, Y. Zhu, M. W. Gray, and B. F. Lang, 2003, Unique mitochondrial genome architecture in unicellular relatives of animals, Proc Natl Acad Sci U S A, v. 100: United States, p. 892-7.
- Burger, G., M. W. Gray, L. Forget, and B. F. Lang, 2013, Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists: Genome Biol Evol, v. 5, p. 418-38.
- Burki, F., 2014, The eukaryotic tree of life from a global phylogenomic perspective: Cold Spring Harb Perspect Biol, v. 6, p. a016147.

- Burri, L., B. A. Williams, D. Bursac, T. Lithgow, and P. J. Keeling, 2006, Microsporidian mitochondria retain elements of the general mitochondrial targeting system: *Proc Natl Acad Sci U S A*, v. 103, p. 15916-20.
- Butterfield, N., and A. Knoll, Swett, K, 1984, Paleobiology of the Upper Proterozoic Svanbergfjellet Formation, Spitsbergen: *Lethaia*, v. 27.
- Butterfield, N. J., 2000, *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes: *Paleobiology*, v. 26, p. 386-404.
- Butterfield, N. J., 2005, Probable Proterozoic fungi.
- Butterfield, N. J., 2007, Macroevolution and macroecology through deep time: *Palaeontology*, v. 50, p. 41-55.
- Butterfield, N. J., 2015, Early evolution of the Eukaryota: *Palaeontology*, v. 58.
- Callahan, H. A., R. W. Litaker, and E. J. Noga, 2002, Molecular taxonomy of the suborder Bodonina (Order Kinetoplastida), including the important fish parasite, *Ichthyobodo necator*: *J Eukaryot Microbiol*, v. 49, p. 119-28.
- Calvo, S. E., and V. K. Mootha, 2010, The mitochondrial proteome and human disease: *Annu Rev Genomics Hum Genet*, v. 11, p. 25-44.
- Camara, J. M., and T. B. Rauchfuss, 2012, Combining acid-base, redox and substrate binding functionalities to give a complete model for the [FeFe]-hydrogenase: *Nat Chem*, v. 4, p. 26-30.
- Canfield, D. E., 1998, A new model for Proterozoic ocean chemistry: *Nature*, v. 396, p. 450-453.
- Carlton, J. M., R. P. Hirt, J. C. Silva, A. L. Delcher, M. Schatz, Q. Zhao, J. R. Wortman, S. L. Bidwell, U. C. Alsmark, S. Besteiro, T. Sicheritz-Ponten, C. J. Noel, J. B. Dacks, P. G. Foster, C. Simillion, Y. Van de Peer, D. Miranda-Saavedra, G. J. Barton, G. D. Westrop, S. Muller, D. Dessi, P. L. Fiori, Q. Ren, I. Paulsen, H. Zhang, F. D. Bastida-Corcuera, A. Simoes-Barbosa, M. T. Brown, R. D. Hayes, M. Mukherjee, C. Y. Okumura, R. Schneider, A. J. Smith, S. Vanacova, M. Villalvazo, B. J. Haas, M. Peretea, T. V. Feldblyum, T. R. Utterback, C. L. Shu, K. Osoegawa, P. J. de Jong, I. Hrdy, L. Horvathova, Z. Zubacova, P. Dolezal, S. B. Malik, J. M. Logsdon, Jr., K. Henze, A. Gupta, C. C. Wang, R. L. Dunne, J. A. Upcroft, P. Upcroft, O. White, S. L. Salzberg, P. Tang, C. H. Chiu, Y. S. Lee, T. M. Embley, G. H. Coombs, J. C. Mottram, J. Tachezy, C. M. Fraser-Liggett, and P. J. Johnson, 2007, Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*: *Science*, v. 315, p. 207-12.
- Carpenter, K. J., and P. J. Keeling, 2007, Morphology and phylogenetic position of *Eucomonympha imla* (Parabasalia: Hypermastigida): *J Eukaryot Microbiol*, v. 54, p. 325-32.
- Cavalier-Smith, T., 2002, The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa: *Int J Syst Evol Microbiol*, v. 52, p. 297-354.
- Cavalier-Smith, T., 2010, Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree: *Biol Lett*, v. 6, p. 342-5.
- Cavalier-Smith, T., 2013, Early evolution of eukaryote feeding modes, cell structural diversity, and classification of the protozoan phyla Loukozoa, Sulcozoa, and Choanozoa: *Eur J Protistol*, v. 49, p. 115-78.
- Cavalier-Smith, T., and E. E. Chao, 2010, Phylogeny and evolution of apusomonadida (protozoa: apusozoa): new genera and species: *Protist*, v. 161, p. 549-76.
- Cervantes-Sandoval, I., J. Serrano-Luna Jde, E. Garcia-Latorre, V. Tsutsumi, and M. Shibayama, 2008, Characterization of brain inflammation during primary amoebic meningoencephalitis: *Parasitol Int*, v. 57, p. 307-13.
- Chen, A. L., E. W. Kim, J. Y. Toh, A. A. Vashisht, A. Q. Rashoff, C. Van, A. S. Huang, A. S. Moon, H. N. Bell, L. A. Bentolila, J. A. Wohlschlegel, and P. J. Bradley, 2015, Novel

- components of the *Toxoplasma* inner membrane complex revealed by BioID: MBio, v. 6, p. e02357-14.
- Chien, P., and L. M. Gierasch, 2014, Challenges and dreams: physics of weak interactions essential to life: Mol Biol Cell, v. 25, p. 3474-7.
- Claros, M. G., and P. Vincens, 1996, Computational method to predict mitochondrially imported proteins and their targeting sequences: Eur J Biochem, v. 241, p. 779-86.
- Cohen, P. A., J. W. Schopf, N. J. Butterfield, A. B. Kudryavtsev, and F. A. Macdonald, 2011, Phosphate biomineralization in mid-Neoproterozoic protists.
- Connolly, M. L., 1983, Solvent-accessible surfaces of proteins and nucleic acids: Science, v. 221, p. 709-13.
- Cox, J., and M. Mann, 2008, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification: Nat Biotechnol, v. 26, p. 1367-72.
- Cox, J., N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann, 2011, Andromeda: a peptide search engine integrated into the MaxQuant environment: J Proteome Res, v. 10, p. 1794-805.
- Dacks, J. B., J. D. Silberman, A. G. Simpson, S. Moriya, T. Kudo, M. Ohkuma, and R. J. Redfield, 2001, Oxymonads are closely related to the excavate taxon Trimastix: Mol Biol Evol, v. 18, p. 1034-44.
- Danne, J. C., and R. F. Waller, 2011, Analysis of dinoflagellate mitochondrial protein sorting signals indicates a highly stable protein targeting system across eukaryotic diversity: J Mol Biol, v. 408, p. 643-53.
- Danovaro, R., A. Dell'Anno, A. Pusceddu, C. Gambi, I. Heiner, and R. M. Kristensen, 2010, The first metazoa living in permanently anoxic conditions, BMC Biol, v. 8: England, p. 30.
- Davidson, E. A., M. van der Giezen, D. S. Horner, T. M. Embley, and C. J. Howe, 2002, An [Fe] hydrogenase from the anaerobic hydrogenosome-containing fungus *Neocallimastix frontalis* L2: Gene, v. 296, p. 45-52.
- de Graaf, R. M., I. Duarte, T. A. van Alen, J. W. Kuiper, K. Schotanus, J. Rosenberg, M. A. Huynen, and J. H. Hackstein, 2009, The hydrogenosomes of *Psalteriomonas lanterna*: BMC Evol Biol, v. 9, p. 287.
- De Jonckheere, J., and H. van de Voorde, 1976, Differences in destruction of cysts of pathogenic and nonpathogenic *Naegleria* and *Acanthamoeba* by chlorine: Appl Environ Microbiol, v. 31, p. 294-7.
- De Jonckheere, J. F., 2004, Molecular definition and the ubiquity of species in the genus *Naegleria*: Protist, v. 155, p. 89-103.
- De Jonckheere, J. F., 2006, Isolation and molecular identification of free-living amoebae of the genus *Naegleria* from Arctic and sub-Antarctic regions: Eur J Protistol, v. 42, p. 115-23.
- De Jonckheere, J. F., 2011, Origin and evolution of the worldwide distributed pathogenic amoeboflagellate *Naegleria fowleri*: Infect Genet Evol, v. 11, p. 1520-8.
- Dellas, N., S. T. Thomas, G. Manning, and J. P. Noel, 2013, Discovery of a metabolic alternative to the classical mevalonate pathway: Elife, v. 2, p. e00672.
- Dingle, A. D., and C. Fulton, 1966, Development of the flagellar apparatus of *Naegleria*: J Cell Biol, v. 31, p. 43-54.
- Dolezal, P., O. Smid, P. Rada, Z. Zubacova, D. Bursac, R. Sutak, J. Nebesarova, T. Lithgow, and J. Tachezy, 2005, *Giardia* mitosomes and trichomonad hydrogenosomes share a common mode of protein targeting: Proc Natl Acad Sci U S A, v. 102, p. 10924-9.
- Edgar, R., M. Domrachev, and A. E. Lash, 2002, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository: Nucleic Acids Res, v. 30, p. 207-10.
- Edgar, R. C., 2004a, MUSCLE: a multiple sequence alignment method with reduced time and space complexity: BMC Bioinformatics, v. 5, p. 113.

- Edgar, R. C., 2004b, MUSCLE: multiple sequence alignment with high accuracy and high throughput: *Nucleic Acids Res*, v. 32, p. 1792-7.
- El Albani, A., S. Bengtson, D. E. Canfield, A. Riboulleau, C. Rollion Bard, R. Macchiarelli, L. Ngombi Pemba, E. Hammarlund, A. Meunier, I. Moubiya Mouele, K. Benzerara, S. Bernard, P. Boulvais, M. Chaussidon, C. Cesari, C. Fontaine, E. Chi-Fru, J. M. Garcia Ruiz, F. Gauthier-Lafaye, A. Mazurier, A. C. Pierson-Wickmann, O. Rouxel, A. Trentesaux, M. Vecoli, G. J. Versteegh, L. White, M. Whitehouse, and A. Bekker, 2014, The 2.1 Ga old Francevillian biota: biogenicity, taphonomy and biodiversity: *PLoS One*, v. 9, p. e99438.
- El-Sayed, N. M., P. J. Myler, G. Blandin, M. Berriman, J. Crabtree, G. Aggarwal, E. Caler, H. Renauld, E. A. Worthey, C. Hertz-Fowler, E. Ghedin, C. Peacock, D. C. Bartholomeu, B. J. Haas, A. N. Tran, J. R. Wortman, U. C. Alsmark, S. Angiuoli, A. Anupama, J. Badger, F. Bringaud, E. Cadag, J. M. Carlton, G. C. Cerqueira, T. Creasy, A. L. Delcher, A. Djikeng, T. M. Embley, C. Hauser, A. C. Ivens, S. K. Kummerfeld, J. B. Pereira-Leal, D. Nilsson, J. Peterson, S. L. Salzberg, J. Shallom, J. C. Silva, J. Sundaram, S. Westenberger, O. White, S. E. Melville, J. E. Donelson, B. Andersson, K. D. Stuart, and N. Hall, 2005, Comparative genomics of trypanosomatid parasitic protozoa: *Science*, v. 309, p. 404-9.
- Ellis, J. E., D. Cole, and D. Lloyd, 1992, Influence of oxygen on the fermentative metabolism of metronidazole-sensitive and resistant strains of *Trichomonas vaginalis*: *Mol Biochem Parasitol*, v. 56, p. 79-88.
- Emanuelsson, O., H. Nielsen, S. Brunak, and G. von Heijne, 2000, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence: *J Mol Biol*, v. 300, p. 1005-16.
- Embley, T. M., 2006, Multiple secondary origins of the anaerobic lifestyle in eukaryotes: *Philos Trans R Soc Lond B Biol Sci*, v. 361, p. 1055-67.
- Embley, T. M., and W. Martin, 2006, Eukaryotic evolution, changes and challenges, *Nature*, v. 440: England, p. 623-30.
- Embley, T. M., M. van der Giezen, D. S. Horner, P. L. Dyal, S. Bell, and P. G. Foster, 2003, Hydrogenosomes, mitochondria and early eukaryotic evolution: *IUBMB Life*, v. 55, p. 387-95.
- Eme, L., S. C. Sharpe, M. W. Brown, and A. J. Roger, 2014, On the age of eukaryotes: evaluating evidence from fossils and molecular clocks: *Cold Spring Harb Perspect Biol*, v. 6.
- Esteban, G. F., B. J. Finlay, and K. J. Clarke, 2009, Sequestered organelles sustain aerobic microbial life in anoxic environments: *Environ Microbiol*, v. 11, p. 544-50.
- Evitt, W. R., 1963, A DISCUSSION AND PROPOSALS CONCERNING FOSSIL DINOFLAGELLATES, HYSTRICHOSPHERES, AND ACRITARCHS, II: *Proc Natl Acad Sci U S A*, v. 49, p. 298-302.
- Feagin, J. E., 1992, The 6-kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes: *Mol Biochem Parasitol*, v. 52, p. 145-8.
- Feagin, J. E., M. J. Gardner, D. H. Williamson, and R. J. Wilson, 1991, The putative mitochondrial genome of *Plasmodium falciparum*: *J Protozool*, v. 38, p. 243-5.
- Fechteler, T., U. Dengler, and D. Schomburg, 1995, Prediction of protein three-dimensional structures in insertion and deletion regions: a procedure for searching data bases of representative protein fragments using geometric scoring criteria: *J Mol Biol*, v. 253, p. 114-31.
- Fenchel, T., and D. J. Patterson, 1986, *Percolomonas cosmopolitus* (Ruinen) n.gen., a new type of filter feeding flagellate from marine plankton: *Journal of the Marine Biological Association of the United Kingdom* 66: 465.
- Field, M. C., and M. Carrington, 2009, The trypanosome flagellar pocket: *Nat Rev Microbiol*, v. 7, p. 775-86.

- Fisk, J. C., J. Li, H. Wang, J. M. Aletta, J. Qu, and L. K. Read, 2013, Proteomic analysis reveals diverse classes of arginine methylproteins in mitochondria of trypanosomes: *Mol Cell Proteomics*, v. 12, p. 302-11.
- Flavin, M., and T. A. Nerad, 1993, *Reclinomonas americana* N. G., N. Sp., a new freshwater heterotrophic flagellate: *J Eukaryot Microbiol*, v. 40, p. 172-9.
- Forestier, M., P. King, L. Zhang, M. Posewitz, S. Schwarzer, T. Happe, M. L. Ghirardi, and M. Seibert, 2003, Expression of two [Fe]-hydrogenases in *Chlamydomonas reinhardtii* under anaerobic conditions: *Eur J Biochem*, v. 270, p. 2750-8.
- Fowler, M., and R. F. Carter, 1965, Acute pyogenic meningitis probably due to *Acanthamoeba* sp.: a preliminary report: *Br Med J*, v. 2, p. 740-2.
- Franco, J. R., P. P. Simarro, A. Diarra, and J. G. Jannin, 2014, Epidemiology of human African trypanosomiasis: *Clin Epidemiol*, v. 6, p. 257-75.
- Fritz-Laylin, L. K., Z. J. Assaf, S. Chen, and W. Z. Cande, 2010a, *Naegleria gruberi* de novo basal body assembly occurs via stepwise incorporation of conserved proteins: *Eukaryot Cell*, v. 9, p. 860-5.
- Fritz-Laylin, L. K., and W. Z. Cande, 2010, Ancestral centriole and flagella proteins identified by analysis of *Naegleria* differentiation: *J Cell Sci*, v. 123, p. 4024-31.
- Fritz-Laylin, L. K., M. L. Ginger, C. Walsh, S. C. Dawson, and C. Fulton, 2011, The *Naegleria* genome: a free-living microbial eukaryote lends unique insights into core eukaryotic cell biology: *Res Microbiol*, v. 162, p. 607-18.
- Fritz-Laylin, L. K., S. E. Prochnik, M. L. Ginger, J. B. Dacks, M. L. Carpenter, M. C. Field, A. Kuo, A. Paredez, J. Chapman, J. Pham, S. Shu, R. Neupane, M. Cipriano, J. Mancuso, H. Tu, A. Salamov, E. Lindquist, H. Shapiro, S. Lucas, I. V. Grigoriev, W. Z. Cande, C. Fulton, D. S. Rokhsar, and S. C. Dawson, 2010b, The genome of *Naegleria gruberi* illuminates early eukaryotic versatility: *Cell*, v. 140, p. 631-42.
- Fukuhara, H., F. Sor, R. Drissi, N. Dinouel, I. Miyakawa, S. Rousset, and A. M. Viola, 1993, Linear mitochondrial DNAs of yeasts: frequency of occurrence and general features: *Mol Cell Biol*, v. 13, p. 2309-14.
- Fulton, C., 1977, Cell differentiation in *Naegleria gruberi*: *Annu Rev Microbiol*, v. 31, p. 597-629.
- Fulton, C., and C. Walsh, 1980, Cell differentiation and flagellar elongation in *Naegleria gruberi*. Dependence on transcription and translation: *J Cell Biol*, v. 85, p. 346-60.
- Fulton, C., C. Webster, and J. S. Wu, 1984, Chemically defined media for cultivation of *Naegleria gruberi*: *Proc Natl Acad Sci U S A*, v. 81, p. 2406-10.
- Gardner, M. J., N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M.-S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. A. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, G. I. McFadden, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser, and B. Barrell, 2002, Genome sequence of the human malaria parasite *Plasmodium falciparum*: *Nature*, v. 419, p. 498-511.
- Gerin, I., G. Noel, J. Bolsee, O. Haumont, E. Van Schaftingen, and G. T. Bommer, 2014, Identification of TP53-induced glycolysis and apoptosis regulator (TIGAR) as the phosphoglycolate-independent 2,3-bisphosphoglycerate phosphatase: *Biochem J*, v. 458, p. 439-48.
- Gfeller, R. P., and M. Gibbs, 1984, Fermentative Metabolism of *Chlamydomonas reinhardtii*: I. Analysis of Fermentative Products from Starch in Dark and Light: *Plant Physiol*, v. 75, p. 212-8.
- Ginger, M. L., 2006, Niche metabolism in parasitic protozoa: *Philos Trans R Soc Lond B Biol Sci*, v. 361, p. 101-18.

- Ginger, M. L., L. K. Fritz-Laylin, C. Fulton, W. Z. Cande, and S. C. Dawson, 2010, Intermediary metabolism in protists: a sequence-based view of facultative anaerobic metabolism in evolutionarily diverse eukaryotes: *Protist*, v. 161, p. 642-71.
- Goddard, J. M., and D. J. Cummings, 1975, Structure and replication of mitochondrial DNA from *Paramecium aurelia*: *J Mol Biol*, v. 97, p. 593-609.
- Goldberg, A. V., S. Molik, A. D. Tsaousis, K. Neumann, G. Kuhnke, F. Delbac, C. P. Vivares, R. P. Hirt, R. Lill, and T. M. Embley, 2008, Localization and functionality of microsporidian iron-sulphur cluster assembly proteins: *Nature*, v. 452, p. 624-8.
- Gray, M. W., 2012, Mitochondrial evolution: *Cold Spring Harb Perspect Biol*, v. 4, p. a011403.
- Gray, M. W., B. F. Lang, R. Cedergren, G. B. Golding, C. Lemieux, D. Sankoff, M. Turmel, N. Brossard, E. Delage, T. G. Littlejohn, I. Plante, P. Rioux, D. Saint-Louis, Y. Zhu, and G. Burger, 1998, Genome structure and gene content in protist mitochondrial DNAs, *Nucleic Acids Res*, v. 26, p. 865-78.
- Greenberg, R. A., B. Sobhian, S. Pathania, S. B. Cantor, Y. Nakatani, and D. M. Livingston, 2006, Multifactorial contributions to an acute DNA damage response by BRCA1/BARD1-containing complexes: *Genes Dev*, v. 20, p. 34-46.
- Gualdron-Lopez, M., A. Brennand, V. Hannaert, W. Quinones, A. J. Caceres, F. Bringaud, J. L. Concepcion, and P. A. Michels, 2012, When, how and why glycolysis became compartmentalised in the Kinetoplastea. A new look at an ancient organelle: *Int J Parasitol*, v. 42, p. 1-20.
- Guengerich, F. P., Z. Tang, S. G. Salamanca-Pinzon, and Q. Cheng, 2010, Characterizing proteins of unknown function: orphan cytochrome p450 enzymes as a paradigm: *Mol Interv*, v. 10, p. 153-63.
- Haanstra, J. R., B. M. Bakker, and P. A. Michels, 2014, In or out? On the tightness of glycosomal compartmentalization of metabolites and enzymes in *Trypanosoma brucei*: *Mol Biochem Parasitol*, v. 198, p. 18-28.
- Haanstra, J. R., E. B. Gonzalez-Marcano, M. Gualdron-Lopez, and P. A. Michels, 2015, Biogenesis, maintenance and dynamics of glycosomes in trypanosomatid parasites: *Biochim Biophys Acta*.
- Haarmann, T., I. Ortel, P. Tudzynski, and U. Keller, 2006, Identification of the cytochrome P450 monooxygenase that bridges the clavine and ergoline alkaloid pathways: *Chembiochem*, v. 7, p. 645-52.
- Hackstein, J. H., A. Akhmanova, B. Boxma, H. R. Harhangi, and F. G. Voncken, 1999, Hydrogenosomes: eukaryotic adaptations to anaerobic environments, *Trends Microbiol*, v. 7: England, p. 441-7.
- Hampl, V., L. Hug, J. W. Leigh, J. B. Dacks, B. F. Lang, A. G. Simpson, and A. J. Roger, 2009, Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups": *Proc Natl Acad Sci U S A*, v. 106, p. 3859-64.
- Hampl, V., J. D. Silberman, A. Stechmann, S. Diaz-Trivino, P. J. Johnson, and A. J. Roger, 2008, Genetic evidence for a mitochondriate ancestry in the 'amitochondriate' flagellate *Trimastix pyriformis*: *PLoS One*, v. 3, p. e1383.
- Hanson, A. D., A. Pribat, J. C. Waller, and V. de Crecy-Lagard, 2010, 'Unknown' proteins and 'orphan' enzymes: the missing half of the engineering parts list--and how to find it: *Biochem J*, v. 425, p. 1-11.
- Happe, T., and A. Kaminski, 2002, Differential regulation of the Fe-hydrogenase during anaerobic adaptation in the green alga *Chlamydomonas reinhardtii*: *Eur J Biochem*, v. 269, p. 1022-32.
- He, D., O. Fiz-Palacios, C. J. Fu, J. Fehling, C. C. Tsai, and S. L. Baldauf, 2014, An alternative root for the eukaryote tree of life: *Curr Biol*, v. 24, p. 465-70.
- Hikosaka, K., Y. Watanabe, N. Tsuji, K. Kita, H. Kishine, N. Arisue, N. M. Palacpac, S. Kawazu, H. Sawai, T. Horii, I. Igarashi, and K. Tanabe, 2010, Divergence of the mitochondrial

- genome structure in the apicomplexan parasites, *Babesia* and *Theileria*, *Mol Biol Evol*, v. 27: United States, p. 1107-16.
- Horner, D. S., P. G. Foster, and T. M. Embley, 2000, Iron hydrogenases and the evolution of anaerobic eukaryotes: *Mol Biol Evol*, v. 17, p. 1695-709.
- Horton, P., K. J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier, and K. Nakai, 2007, WoLF PSORT: protein localization predictor: *Nucleic Acids Res*, v. 35, p. W585-7.
- Hug, L. A., and A. J. Roger, 2007, The impact of fossils and taxon sampling on ancient molecular dating analyses: *Mol Biol Evol*, v. 24, p. 1889-97.
- Hug, L. A., A. Stechmann, and A. J. Roger, 2010, Phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes: *Mol Biol Evol*, v. 27, p. 311-24.
- Huizinga, H. W., and G. L. McLaughlin, 1990, Thermal ecology of *Naegleria fowleri* from a power plant cooling reservoir: *Appl Environ Microbiol*, v. 56, p. 2200-5.
- Huynen, M. A., B. Snel, P. Bork, and T. J. Gibson, 2001, The phylogenetic distribution of frataxin indicates a role in iron-sulfur cluster protein assembly: *Hum Mol Genet*, v. 10, p. 2463-8.
- Jackson, A. P., T. D. Otto, M. Aslett, S. D. Armstrong, F. Bringaud, A. Schlacht, C. Hartley, M. Sanders, J. M. Wastling, J. B. Dacks, A. Acosta-Serrano, M. C. Field, M. L. Ginger, and M. Berriman, 2015, Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism: *Curr Biol*.
- Jackson, A. P., M. A. Quail, and M. Berriman, 2008, Insights into the genome sequence of a free-living Kinetoplastid: *Bodo saltans* (Kinetoplastida: Euglenozoa): *BMC Genomics*, v. 9, p. 594.
- Jamerson, M., K. Remmers, G. Cabral, and F. Marciano-Cabral, 2009, Survey for the presence of *Naegleria fowleri* amebae in lake water used to cool reactors at a nuclear power generating plant: *Parasitol Res*, v. 104, p. 969-78.
- Javaux, E. J., A. H. Knoll, and M. Walter, 2003, Recognizing and interpreting the fossils of early eukaryotes: *Orig Life Evol Biosph*, v. 33, p. 75-94.
- Jedelsky, P. L., P. Dolezal, P. Rada, J. Pyrih, O. Smid, I. Hrdy, M. Sedinova, M. Marcincikova, L. Voleman, A. J. Perry, N. C. Beltran, T. Lithgow, and J. Tachezy, 2011, The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*: *PLoS One*, v. 6, p. e17285.
- Jedrzejewski, M. J., 2000, Structure, function, and evolution of phosphoglycerate mutases: comparison with fructose-2,6-bisphosphatase, acid phosphatase, and alkaline phosphatase: *Prog Biophys Mol Biol*, v. 73, p. 263-87.
- Jerlstrom-Hultqvist, J., E. Einarsson, F. Xu, K. Hjort, B. Ek, D. Steinhauf, K. Hultenby, J. Bergquist, J. O. Andersson, and S. G. Svard, 2013, Hydrogenosomes in the diplomonad *Spironucleus salmonicida*: *Nat Commun*, v. 4, p. 2493.
- Jones, C. P., and C. Ingram-Smith, 2014, Biochemical and kinetic characterization of the recombinant ADP-forming acetyl coenzyme A synthetase from the amitochondriate protozoan *Entamoeba histolytica*: *Eukaryot Cell*, v. 13, p. 1530-7.
- Kadlec, V., L. Cerva, and J. Skvarova, 1978, Virulent *Naegleria fowleri* in an indoor swimming pool: *Science*, v. 201, p. 1025.
- Kairo, A., A. H. Fairlamb, E. Gobright, and V. Nene, 1994, A 7.1 kb linear DNA molecule of *Theileria parva* has scrambled rDNA sequences and open reading frames for mitochondrially encoded proteins: *EMBO J*, v. 13, p. 898-905.
- Kamikawa, R., M. Kolisko, Y. Nishimura, A. Yabuki, M. W. Brown, S. A. Ishikawa, K. Ishida, A. J. Roger, T. Hashimoto, and Y. Inagaki, 2014, Gene content evolution in Discobid mitochondria deduced from the phylogenetic position and complete mitochondrial genome of *Tsukubamonas globosa*: *Genome Biol Evol*, v. 6, p. 306-15.

- Katinka, M. D., S. Duprat, E. Cornillot, G. Metenier, F. Thomarat, G. Prensier, V. Barbe, E. Peyretilade, P. Brottier, P. Wincker, F. Delbac, H. El Alaoui, P. Peyret, W. Saurin, M. Gouy, J. Weissenbach, and C. P. Vivares, 2001, Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*: *Nature*, v. 414, p. 450-3.
- Katz, L. A., J. Grant, L. W. Parfrey, A. Gant, C. J. O'Kelly, O. R. Anderson, R. E. Molestina, and T. Nerad, 2011, *Subulatomonas tetraspora* nov. gen. nov. sp. is a member of a previously unrecognized major clade of eukaryotes: *Protist*, v. 162, p. 762-73.
- Kelly, S., J. Reed, S. Kramer, L. Ellis, H. Webb, J. Sunter, J. Salje, N. Marinsek, K. Gull, B. Wickstead, and M. Carrington, 2007, Functional genomics in *Trypanosoma brucei*: a collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci: *Mol Biochem Parasitol*, v. 154, p. 103-9.
- Kessel, M., and J. N. Eloff, 1975, The ultrastructure and development of the colonial sheath of *Microcystis marginata*: *Arch Microbiol*, v. 106, p. 209-14.
- Kim, D. I., K. C. Birendra, W. Zhu, K. Motamedchaboki, V. Doye, and K. J. Roux, 2014, Probing nuclear pore complex architecture with proximity-dependent biotinylation: *Proc Natl Acad Sci U S A*, v. 111, p. E2453-61.
- Kim, E., J. W. Harrison, S. Sudek, M. D. Jones, H. M. Wilcox, T. A. Richards, A. Z. Worden, and J. M. Archibald, 2011, Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life: *Proc Natl Acad Sci U S A*, v. 108, p. 1496-500.
- Kim, E., A. G. Simpson, and L. E. Graham, 2006, Evolutionary relationships of apusomonads inferred from taxon-rich analyses of 6 nuclear encoded genes: *Mol Biol Evol*, v. 23, p. 2455-66.
- Knoll, A. H., 2014, Paleobiological perspectives on early eukaryotic evolution: *Cold Spring Harb Perspect Biol*, v. 6.
- Knoll, A. H., E. J. Javaux, D. Hewitt, and P. Cohen, 2006, Eukaryotic organisms in Proterozoic oceans: *Philos Trans R Soc Lond B Biol Sci*, v. 361, p. 1023-38.
- Kobayashi, M., Y. Matsuo, A. Takimoto, S. Suzuki, F. Maruo, and H. Shoun, 1996, Denitrification, a novel type of respiratory metabolism in fungal mitochondrion: *J Biol Chem*, v. 271, p. 16263-7.
- Kobayashi, M., and H. Shoun, 1995, The Copper-containing Dissimilatory Nitrite Reductase Involved in the Denitrifying System of the Fungus *Fusarium oxysporum*.
- Kolisko, M., J. D. Silberman, I. Cepicka, N. Yubuki, K. Takishita, A. Yabuki, B. S. Leander, I. Inouye, Y. Inagaki, A. J. Roger, and A. G. Simpson, 2010, A wide diversity of previously undetected free-living relatives of diplomonads isolated from marine/saline habitats: *Environ Microbiol*, v. 12, p. 2700-10.
- Labute, P., 2008, The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area: *J Comput Chem*, v. 29, p. 1693-8.
- Lamb, D., S. Awramik, D. Chapman, and S. Zhu, 2009, Evidence for eukaryotic diversification in the ~1800 million-year-old Changzhougou Formation, North China, v. 173, p. 93-104.
- Lang, B. F., G. Burger, C. J. O'Kelly, R. Cedergren, G. B. Golding, C. Lemieux, D. Sankoff, M. Turmel, and M. W. Gray, 1997, An ancestral mitochondrial DNA resembling a eubacterial genome in miniature: *Nature*, v. 387, p. 493-497.
- Lara, E., A. Chatzinotas, and A. G. Simpson, 2006, Andalusia (n. gen.)--the deepest branch within jakobids (Jakobida; Excavata), based on morphological and molecular study of a new flagellate from soil: *J Eukaryot Microbiol*, v. 53, p. 112-20.
- Leander, B. S., 2004, Did trypanosomatid parasites have photosynthetic ancestors?: *Trends Microbiol*, v. 12, p. 251-8.
- Leander, B. S., H. J. Esson, and S. A. Breglia, 2007, Macroevolution of complex cytoskeletal systems in euglenids: *Bioessays*, v. 29, p. 987-1000.

- Letunic, I., T. Doerks, and P. Bork, 2015, SMART: recent updates, new developments and status in 2015: *Nucleic Acids Res*, v. 43, p. D257-60.
- Levitt, M., 1992, Accurate modeling of protein conformation by automatic segment matching: *J Mol Biol*, v. 226, p. 507-33.
- Li, H., and G. Jogl, 2009, Structural and biochemical studies of TIGAR (TP53-induced glycolysis and apoptosis regulator): *J Biol Chem*, v. 284, p. 1748-54.
- Lill, R., 2009, Function and biogenesis of iron-sulphur proteins: *Nature*, v. 460, p. 831-8.
- Lindmark, D. G., and M. Müller, 1973, Hydrogenosome, a Cytoplasmic Organelle of the Anaerobic Flagellate *Tritrichomonas foetus*, and Its Role in Pyruvate Metabolism.
- Lloyd, D., J. R. Ralphs, and J. C. Harris, 2002, *Giardia intestinalis*, a eukaryote without hydrogenosomes, produces hydrogen: *Microbiology*, v. 148, p. 727-33.
- Long, S., M. Jirku, F. J. Ayala, and J. Lukes, 2008a, Mitochondrial localization of human frataxin is necessary but processing is not for rescuing frataxin deficiency in *Trypanosoma brucei*: *Proc Natl Acad Sci U S A*, v. 105, p. 13468-73.
- Long, S., M. Jirku, J. Mach, M. L. Ginger, R. Sutak, D. Richardson, J. Tachezy, and J. Lukes, 2008b, Ancestral roles of eukaryotic frataxin: mitochondrial frataxin function and heterologous expression of hydrogenosomal *Trichomonas* homologues in trypanosomes: *Mol Microbiol*, v. 69, p. 94-109.
- Lukes, J., O. Flegontova, and A. Horak, 2015, Diplonemids: *Curr Biol*, v. 25, p. R702-4.
- Lukes, J., H. Hashimi, and A. Zikova, 2005, Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates: *Curr Genet*, v. 48, p. 277-99.
- Lukeš, J., D. Lys Guilbride, J. Votýpka, A. Zíková, R. Benne, and P. T. Englund, 2002, Kinetoplast DNA Network: Evolution of an Improbable Structure, *Eukaryot Cell*, v. 1, p. 495-502.
- Lynch, M., M. C. Field, H. V. Goodson, H. S. Malik, J. B. Pereira-Leal, D. S. Roos, A. P. Turkewitz, and S. Sazer, 2014, Evolutionary cell biology: two origins, one objective: *Proc Natl Acad Sci U S A*, v. 111, p. 16990-4.
- MacGregor, P., B. Szoor, N. J. Savill, and K. R. Matthews, 2012, Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act: *Nat Rev Microbiol*, v. 10, p. 431-8.
- Makiuchi, T., and T. Nozaki, 2014, Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa: *Biochimie*, v. 100, p. 3-17.
- Marande, W., J. Lukes, and G. Burger, 2005, Unique mitochondrial genome structure in diplomemids, the sister group of kinetoplastids: *Eukaryot Cell*, v. 4, p. 1137-46.
- Marciano-Cabral, F., 1988, Biology of *Naegleria* spp: *Microbiol Rev*, v. 52, p. 114-33.
- Marciano-Cabral, F., and G. A. Cabral, 2007, The immune response to *Naegleria fowleri* amebae and pathogenesis of infection: *FEMS Immunol Med Microbiol*, v. 51, p. 243-59.
- Martin, F. N., 1995, Linear mitochondrial genome organization in vivo in the genus *Pythium*: *Curr Genet*, v. 28, p. 225-34.
- Martin, W., 2003, Gene transfer from organelles to the nucleus: frequent and in big chunks: *Proc Natl Acad Sci U S A*, v. 100, p. 8612-4.
- Martincova, E., L. Voleman, J. Pyrih, V. Zarsky, P. Vondrackova, M. Kolisko, J. Tachezy, and P. Dolezal, 2015, Probing the Biology of *Giardia intestinalis* Mitosomes Using In Vivo Enzymatic Tagging: *Mol Cell Biol*, v. 35, p. 2864-74.
- Marumo, K., K. Nakada-Tsukui, K. Tomii, and T. Nozaki, 2014, Ligand heterogeneity of the cysteine protease binding protein family in the parasitic protist *Entamoeba histolytica*: *Int J Parasitol*, v. 44, p. 625-35.
- Matthews, K. R., 1999, Developments in the differentiation of *Trypanosoma brucei*: *Parasitol Today*, v. 15, p. 76-80.
- Matthews, K. R., R. McCulloch, and L. J. Morrison, 2015, The within-host dynamics of African trypanosome infections: *Philos Trans R Soc Lond B Biol Sci*, v. 370.

- McAllaster, M. R., K. N. Ikeda, A. Lozano-Nunez, D. Anrather, V. Unterwurzacher, T. Gossenreiter, J. A. Perry, R. Crickley, C. J. Mercadante, S. Vaughan, and C. L. de Graffenried, 2015, Proteomic identification of novel cytoskeletal proteins associated with TbPLK, an essential regulator of cell morphogenesis in *Trypanosoma brucei*: Mol Biol Cell, v. 26, p. 3013-29.
- McCulloch, R., 2004, Antigenic variation in African trypanosomes: monitoring progress: Trends Parasitol, v. 20, p. 117-21.
- Melis, A., and T. Happe, 2001, Hydrogen production. Green algae as a source of energy: Plant Physiol, v. 127, p. 740-8.
- Melis, A., L. Zhang, M. Forestier, M. L. Ghirardi, and M. Seibert, 2000, Sustained photobiological hydrogen gas production upon reversible inactivation of oxygen evolution in the green alga *Chlamydomonas reinhardtii*: Plant Physiol, v. 122, p. 127-36.
- Melkonian, M., B. Becker, and D. Becker, 1991, Scale formation in algae: J Electron Microscop Tech, v. 17, p. 165-78.
- Meuser, J. E., E. S. Boyd, G. Ananyev, D. Karns, R. Radakovits, U. M. Narayana Murthy, M. L. Ghirardi, G. C. Dismukes, J. W. Peters, and M. C. Posewitz, 2011, Evolutionary significance of an algal gene encoding an [FeFe]-hydrogenase with F-domain homology and hydrogenase activity in *Chlorella variabilis* NC64A: Planta, v. 234, p. 829-43.
- Michels, P. A., V. Hannaert, and F. Bringaud, 2000, Metabolic aspects of glycosomes in trypanosomatidae - new data and views: Parasitol Today, v. 16, p. 482-9.
- Minge, M. A., J. D. Silberman, R. J. Orr, T. Cavalier-Smith, K. Shalchian-Tabrizi, F. Burki, A. Skjaeveland, and K. S. Jakobsen, 2009, Evolutionary position of breviate amoebae and the primary eukaryote divergence: Proc Biol Sci, v. 276, p. 597-604.
- Mitchell, A., H. Y. Chang, L. Daugherty, M. Fraser, S. Hunter, R. Lopez, C. McAnulla, C. McMenamin, G. Nuka, S. Pesseat, A. Sangrador-Vegas, M. Scheremetjew, C. Rato, S. Y. Yong, A. Bateman, M. Punta, T. K. Attwood, C. J. Sigrist, N. Redaschi, C. Rivoire, I. Xenarios, D. Kahn, D. Guyot, P. Bork, I. Letunic, J. Gough, M. Oates, D. Haft, H. Huang, D. A. Natale, C. H. Wu, C. Orengo, I. Sillitoe, H. Mi, P. D. Thomas, and R. D. Finn, 2015, The InterPro protein families database: the classification resource after 15 years: Nucleic Acids Res, v. 43, p. D213-21.
- Mohammed, H., and J. S. Carroll, 2013, Approaches for assessing and discovering protein interactions in cancer: Mol Cancer Res, v. 11, p. 1295-302.
- Mootha, V. K., J. Bunkenborg, J. V. Olsen, M. Hjerrild, J. R. Wisniewski, E. Stahl, M. S. Bolouri, H. N. Ray, S. Sihag, M. Kamal, N. Patterson, E. S. Lander, and M. Mann, 2003, Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria: Cell, v. 115, p. 629-40.
- Moreira, D., P. Lopez-Garcia, and K. Vickerman, 2004, An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea: Int J Syst Evol Microbiol, v. 54, p. 1861-75.
- Morgan, G. W., D. Goulding, and M. C. Field, 2004, The single dynamin-like protein of *Trypanosoma brucei* regulates mitochondrial division and is not required for endocytosis: J Biol Chem, v. 279, p. 10692-701.
- Morozkina, E. V., and A. V. Kurakov, 2007, [Dissimilatory nitrate reduction in fungi under conditions of hypoxia and anoxia: a review]: Prikl Biokhim Mikrobiol, v. 43, p. 607-13.
- Morris, J. C., M. E. Drew, M. M. Klingbeil, S. A. Motyka, T. T. Saxowsky, Z. Wang, and P. T. Englund, 2001, Replication of kinetoplast DNA: an update for the new millennium, Int J Parasitol, v. 31: England, p. 453-8.
- Morrison, H. G., A. G. McArthur, F. D. Gillin, S. B. Aley, R. D. Adam, G. J. Olsen, A. A. Best, W. Z. Cande, F. Chen, M. J. Cipriano, B. J. Davids, S. C. Dawson, H. G. Elmendorf, A. B. Hehl, M. E. Holder, S. M. Huse, U. U. Kim, E. Lasek-Nesselquist, G. Manning, A.

- Nigam, J. E. Nixon, D. Palm, N. E. Passamaneck, A. Prabhu, C. I. Reich, D. S. Reiner, J. Samuelson, S. G. Svard, and M. L. Sogin, 2007, Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*: *Science*, v. 317, p. 1921-6.
- Morriswood, B., K. Havlicek, L. Demmel, S. Yavuz, M. Sealey-Cardona, K. Vidilaseris, D. Anrather, J. Kostan, K. Djinovic-Carugo, K. J. Roux, and G. Warren, 2013, Novel bilobe components in *Trypanosoma brucei* identified using proximity-dependent biotinylation: *Eukaryot Cell*, v. 12, p. 356-67.
- Mulder, D. W., E. M. Shepard, J. E. Meuser, N. Joshi, P. W. King, M. C. Posewitz, J. B. Broderick, and J. W. Peters, 2011, Insights into [FeFe]-hydrogenase structure, mechanism, and maturation, *Structure*, v. 19: United States, 2011 Elsevier Ltd, p. 1038-52.
- Muller, M., 1993, The hydrogenosome: *J Gen Microbiol*, v. 139, p. 2879-89.
- Muller, M., M. Mentel, J. J. van Hellemond, K. Henze, C. Woehle, S. B. Gould, R. Y. Yu, M. van der Giezen, A. G. Tielens, and W. F. Martin, 2012, Biochemistry and evolution of anaerobic energy metabolism in eukaryotes: *Microbiol Mol Biol Rev*, v. 76, p. 444-95.
- Mus, F., A. Dubini, M. Seibert, M. C. Posewitz, and A. R. Grossman, 2007, Anaerobic acclimation in *Chlamydomonas reinhardtii*: anoxic gene expression, hydrogenase induction, and metabolic pathways: *J Biol Chem*, v. 282, p. 25475-86.
- Müller, M., M. Mentel, J. J. v. Hellemond, K. Henze, C. Woehle, S. B. Gould, R.-Y. Yu, M. v. d. Giezen, A. G. M. Tielens, and W. F. Martin, 2012, Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes.
- Nagovitsin, K., 2009, *Tappania*-bearing association of the Siberian platform: Biodiversity, stratigraphic position and geochronological constraints, v. 173, p. 137-145.
- Nakahara, K., and H. Shoun, 1996, N-terminal processing and amino acid sequence of two isoforms of nitric oxide reductase cytochrome P450_{nor} from *Fusarium oxysporum*: *J Biochem*, v. 120, p. 1082-7.
- Nguyen, S., D. C. Jones, S. Wyllie, A. H. Fairlamb, and M. A. Phillips, 2013, Allosteric activation of trypanosomatid deoxyhypusine synthase by a catalytically dead paralog: *J Biol Chem*, v. 288, p. 15256-67.
- Niemann, M., S. Wiese, J. Mani, A. Chanfon, C. Jackson, C. Meisinger, B. Warscheid, and A. Schneider, 2013, Mitochondrial outer membrane proteome of *Trypanosoma brucei* reveals novel factors required to maintain mitochondrial morphology: *Mol Cell Proteomics*, v. 12, p. 515-28.
- Nixon, J. E., J. Field, A. G. McArthur, M. L. Sogin, N. Yarlett, B. J. Loftus, and J. Samuelson, 2003, Iron-dependent hydrogenases of *Entamoeba histolytica* and *Giardia lamblia*: activity of the recombinant entamoebic enzyme and evidence for lateral gene transfer: *Biol Bull*, v. 204, p. 1-9.
- Nowack, E. C., and M. Melkonian, 2010, Endosymbiotic associations within protists: *Philos Trans R Soc Lond B Biol Sci*, v. 365, p. 699-712.
- O'Kelly, C. J., J. D. Silberman, L. A. Amaral Zettler, T. A. Nerad, and M. L. Sogin, 2003, *Monopylocystis visvesvarai* n. gen., n. sp. and *Sawyeria marylandensis* n. gen., n. sp.: two new amitochondrial heterolobosean amoebae from anoxic environments: *Protist*, v. 154, p. 281-90.
- Ogino, T., M. Matsubara, N. Kato, Y. Nakamura, and T. Mizuno, 1998, An *Escherichia coli* protein that exhibits phosphohistidine phosphatase activity towards the HPT domain of the ArcB sensor involved in the multistep His-Asp phosphorelay: *Mol Microbiol*, v. 27, p. 573-85.
- Ong, S. E., B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, and M. Mann, 2002, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics: *Mol Cell Proteomics*, v. 1, p. 376-86.
- Ong, S. E., and M. Mann, 2005, Mass spectrometry-based proteomics turns quantitative: *Nat Chem Biol*, v. 1, p. 252-62.

- Opperdoes, F. R., 1987, Compartmentation of carbohydrate metabolism in trypanosomes: *Annu Rev Microbiol*, v. 41, p. 127-51.
- Opperdoes, F. R., and P. Borst, 1977, Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: the glycosome: *FEBS Lett*, v. 80, p. 360-4.
- Opperdoes, F. R., J. F. De Jonckheere, and A. G. Tielens, 2011, *Naegleria gruberi* metabolism: *Int J Parasitol*, v. 41, p. 915-24.
- Panek, T., J. D. Silberman, N. Yubuki, B. S. Leander, and I. Cepicka, 2012, Diversity, evolution and molecular systematics of the Psalteriomonadidae, the main lineage of anaerobic/microaerophilic heteroloboseans (excavata: discoba): *Protist*, v. 163, p. 807-31.
- Panigrahi, A. K., Y. Ogata, A. Zikova, A. Anupama, R. A. Dalley, N. Acestor, P. J. Myler, and K. D. Stuart, 2009, A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome: *Proteomics*, v. 9, p. 434-50.
- Park, J. S., and A. G. Simpson, 2015, Diversity of Heterotrophic Protists from Extremely Hypersaline Habitats: *Protist*, v. 166, p. 422-37.
- Park, J. S., A. G. Simpson, S. Brown, and B. C. Cho, 2009, Ultrastructure and molecular phylogeny of two heterolobosean amoebae, *Euplaesiobystra hypersalinica* gen. et sp. nov. and *Tulamoeba peronaphora* gen. et sp. nov., isolated from an extremely hypersaline habitat: *Protist*, v. 160, p. 265-83.
- Patterson, D. J., 1999, The Diversity of Eukaryotes: *Am Nat*, v. 154, p. S96-s124.
- Peixoto, L., F. Chen, O. S. Harb, P. H. Davis, D. P. Beiting, C. S. Brownback, D. Ouloguem, and D. S. Roos, 2010, Integrative genomic approaches highlight a family of parasite-specific kinases that regulate host responses: *Cell Host Microbe*, v. 8, p. 208-18.
- Philippe, H., and A. Germot, 2000, Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution: *Mol Biol Evol*, v. 17, p. 830-4.
- Pillay, D., A. F. Boulange, V. Coustou, T. Baltz, and T. H. Coetzer, 2013, Recombinant expression and biochemical characterisation of two alanyl aminopeptidases of *Trypanosoma congolense*: *Exp Parasitol*, v. 135, p. 675-84.
- Pombert, J. F., N. A. Blouin, C. Lane, D. Boucias, and P. J. Keeling, 2014, A lack of parasitic reduction in the obligate parasitic green alga *Helicosporidium*: *PLoS Genet*, v. 10, p. e1004355.
- Poon, S. K., L. Peacock, W. Gibson, K. Gull, and S. Kelly, 2012, A modular and optimized single marker system for generating *Trypanosoma brucei* cell lines expressing T7 RNA polymerase and the tetracycline repressor: *Open Biol*, v. 2, p. 110037.
- Porter, S. M., R. Meisterfeld, and A. Knoll, 2003, Vase-shaped Microfossils from the Neoproterozoic Chuar Group, Grand Canyon: A Classification Guided by Modern Testate Amoebae.
- Price, D. C., C. X. Chan, H. S. Yoon, E. C. Yang, H. Qiu, A. P. Weber, R. Schwacke, J. Gross, N. A. Blouin, C. Lane, A. Reyes-Prieto, D. G. Durnford, J. A. Neilson, B. F. Lang, G. Burger, J. M. Steiner, W. Loffelhardt, J. E. Meuser, M. C. Posewitz, S. Ball, M. C. Arias, B. Henrissat, P. M. Coutinho, S. A. Rensing, A. Symeonidi, H. Doddapaneni, B. R. Green, V. D. Rajah, J. Boore, and D. Bhattacharya, 2012, *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants: *Science*, v. 335, p. 843-7.
- Putz, S., P. Dolezal, G. Gelius-Dietrich, L. Bohacova, J. Tachezy, and K. Henze, 2006, Fe-hydrogenase maturases in the hydrogenosomes of *Trichomonas vaginalis*: *Eukaryot Cell*, v. 5, p. 579-86.
- Rees, J. S., X. W. Li, S. Perrett, K. S. Lilley, and A. P. Jackson, 2015, Protein Neighbors and Proximity Proteomics: *Mol Cell Proteomics*, v. 14, p. 2848-56.
- Regoes, A., D. Zourmpanou, G. Leon-Avila, M. van der Giezen, J. Tovar, and A. B. Hehl, 2005, Protein import, replication, and inheritance of a vestigial mitochondrion: *J Biol Chem*, v. 280, p. 30557-63.

- Reid, H., S. Kibona, A. Rodney, B. McPherson, C. Sindato, I. Malele, S. Kinung'hi, M. Jennaway, J. Chagalucha, B. Blake, and A. Vallely, 2012, Assessment of the burden of human African trypanosomiasis by rapid participatory appraisal in three high-risk villages in Urambo District, Northwest Tanzania, *Afr Health Sci*, v. 12, p. 104-13.
- Rigden, D. J., 2008, The histidine phosphatase superfamily: structure and function: *Biochem J*, v. 409, p. 333-48.
- Rigden, D. J., I. Bagyan, E. Lamani, P. Setlow, and M. J. Jedrzejewski, 2001, A cofactor-dependent phosphoglycerate mutase homolog from *Bacillus stearothermophilus* is actually a broad specificity phosphatase: *Protein Sci*, v. 10, p. 1835-46.
- Rigden, D. J., L. V. Mello, P. Setlow, and M. J. Jedrzejewski, 2002, Structure and mechanism of action of a cofactor-dependent phosphoglycerate mutase homolog from *Bacillus stearothermophilus* with broad specificity phosphatase activity: *J Mol Biol*, v. 315, p. 1129-43.
- Robinson, B. S., P. E. Christy, and J. F. De Jonckheere, 1989, A temporary flagellate (mastigote) stage in the vahlkampfiid amoeba *Williaertia magna* and its possible evolutionary significance: *Biosystems*, v. 23, p. 75-86.
- Roger, A. J., and L. A. Hug, 2006, The origin and diversification of eukaryotes: problems with molecular phylogenetics and molecular clock estimation: *Philos Trans R Soc Lond B Biol Sci*, v. 361, p. 1039-54.
- Roger, A. J., O. Sandblom, W. F. Doolittle, and H. Philippe, 1999, An evaluation of elongation factor 1 alpha as a phylogenetic marker for eukaryotes: *Mol Biol Evol*, v. 16, p. 218-33.
- Rosano, G. L., and E. A. Ceccarelli, 2014, Recombinant protein expression in *Escherichia coli*: advances and challenges: *Front Microbiol*, v. 5, p. 172.
- Rotureau, B., I. Subota, J. Buisson, and P. Bastin, 2012, A new asymmetric division contributes to the continuous production of infective trypanosomes in the tsetse fly: *Development*, v. 139, p. 1842-50.
- Roux, K. J., D. I. Kim, M. Raida, and B. Burke, 2012, A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells: *J Cell Biol*, v. 196, p. 801-10.
- Rudinger, M., L. Fritz-Laylin, M. Polsakiewicz, and V. Knoop, 2011, Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*: *Rna*, v. 17, p. 2058-62.
- Sadatom, D., S. Tanimura, K.-i. Ozaki, and K. Takeda, 2013, Atypical Protein Phosphatases: Emerging Players in Cellular Signaling: *International Journal of Molecular Sciences*, v. 14, p. 4596-4612.
- Schopf, J. W., 2006, Fossil evidence of Archaean life: *Philos Trans R Soc Lond B Biol Sci*, v. 361, p. 869-85.
- Schultz, J., F. Milpetz, P. Bork, and C. P. Ponting, 1998, SMART, a simple modular architecture research tool: identification of signaling domains: *Proc Natl Acad Sci U S A*, v. 95, p. 5857-64.
- Schuster, F., 1975, Ultrastructure of Cysts of *Naegleria* spp: A Comparative Study*: *Journal of Eukaryotic Microbiology*, v. 22.
- Shu, C. J., L. E. Ulrich, and I. B. Zhulin, 2003, The NIT domain: a predicted nitrate-responsive module in bacterial sensory receptors: *Trends Biochem Sci*, v. 28, p. 121-4.
- Simarro, P. P., G. Cecchi, J. R. Franco, M. Paone, A. Diarra, G. Priotto, R. C. Mattioli, and J. G. Jannin, 2015, Monitoring the Progress towards the Elimination of *Gambiense* Human African Trypanosomiasis: *PLoS Negl Trop Dis*, v. 9, p. e0003785.
- Simarro, P. P., G. Cecchi, J. R. Franco, M. Paone, A. Diarra, J. A. Ruiz-Postigo, E. M. Fevre, R. C. Mattioli, and J. G. Jannin, 2012, Estimating and mapping the population at risk of sleeping sickness: *PLoS Negl Trop Dis*, v. 6, p. e1859.
- Simpson, A. G., 2003, Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota): *Int J Syst Evol Microbiol*, v. 53, p. 1759-77.

- Simpson, A. G., Y. Inagaki, and A. J. Roger, 2006a, Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes: *Mol Biol Evol*, v. 23, p. 615-25.
- Simpson, A. G., and D. J. Patterson, 2001, On core jakobids and excavate taxa: the ultrastructure of *Jakoba incarcerata*: *J Eukaryot Microbiol*, v. 48, p. 480-92.
- Simpson, A. G., and A. J. Roger, 2004, The real 'kingdoms' of eukaryotes: *Curr Biol*, v. 14, p. R693-6.
- Simpson, A. G., J. R. Stevens, and J. Lukes, 2006b, The evolution and diversity of kinetoplastid flagellates: *Trends Parasitol*, v. 22, p. 168-74.
- Small, I., N. Peeters, F. Legeai, and C. Lurin, 2004, Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences: *Proteomics*, v. 4, p. 1581-90.
- Sogin, M. L., and J. H. Gunderson, 1987, Structural diversity of eukaryotic small subunit ribosomal RNAs. Evolutionary implications: *Ann N Y Acad Sci*, v. 503, p. 125-39.
- Stark, K., and F. P. Guengerich, 2007, Characterization of orphan human cytochromes P450: *Drug Metab Rev*, v. 39, p. 627-37.
- Stechmann, A., K. Hamblin, V. Perez-Brocal, D. Gaston, G. S. Richmond, M. van der Giezen, C. G. Clark, and A. J. Roger, 2008, Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes: *Curr Biol*, v. 18, p. 580-5.
- Sutherland, C. S., J. Yukich, R. Goeree, and F. Tediosi, 2015, A literature review of economic evaluations for a neglected tropical disease: human African trypanosomiasis ("sleeping sickness"): *PLoS Negl Trop Dis*, v. 9, p. e0003397.
- Suyama, Y., and K. Miura, 1968, SIZE AND STRUCTURAL VARIATIONS OF MITOCHONDRIAL DNA*: *Proc Natl Acad Sci U S A*, v. 60, p. 235-42.
- Swarbreck, D., C. Wilks, P. Lamesch, T. Z. Berardini, M. Garcia-Hernandez, H. Foerster, D. Li, T. Meyer, R. Muller, L. Ploetz, A. Radenbaugh, S. Singh, V. Swing, C. Tissier, P. Zhang, and E. Huala, 2008, The Arabidopsis Information Resource (TAIR): gene structure and function annotation: *Nucleic Acids Res*, v. 36, p. D1009-14.
- Szpara, M. L., D. Gatherer, A. Ochoa, B. Greenbaum, A. Dolan, R. J. Bowden, L. W. Enquist, M. Legendre, and A. J. Davison, 2014, Evolution and diversity in human herpes simplex virus genomes: *J Virol*, v. 88, p. 1209-27.
- Taanman, J.-W., 1999, The mitochondrial genome: structure, transcription, translation and replication, v. 1410, p. 103-123.
- Takeda, K., Y. Komuro, T. Hayakawa, H. Oguchi, Y. Ishida, S. Murakami, T. Noguchi, H. Kinoshita, Y. Sekine, S. Iemura, T. Natsume, and H. Ichijo, 2009, Mitochondrial phosphoglycerate mutase 5 uses alternate catalytic activity as a protein serine/threonine phosphatase to activate ASK1: *Proc Natl Acad Sci U S A*, v. 106, p. 12301-5.
- Tasker, M., M. Timms, E. Hendriks, and K. Matthews, 2001, Cytochrome oxidase subunit VI of *Trypanosoma brucei* is imported without a cleaved presequence and is developmentally regulated at both RNA and protein levels: *Mol Microbiol*, v. 39, p. 272-85.
- Tielens, A. G., C. Rotte, J. J. van Hellemond, and W. Martin, 2002, Mitochondria as we don't know them, *Trends Biochem Sci*, v. 27: England, p. 564-72.
- Tielens, A. G., and J. J. van Hellemond, 2009, Surprising variety in energy metabolism within Trypanosomatidae: *Trends Parasitol*, v. 25, p. 482-90.
- Tiengwe, C., L. Marcello, H. Farr, C. Gadelha, R. Burchmore, J. D. Barry, S. D. Bell, and R. McCulloch, 2012, Identification of ORC1/CDC6-interacting factors in *Trypanosoma brucei* reveals critical features of origin recognition complex architecture: *PLoS One*, v. 7, p. e32674.
- Timmis, J. N., M. A. Ayliffe, C. Y. Huang, and W. Martin, 2004, Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes: *Nat Rev Genet*, v. 5, p. 123-35.

- Tovar, J., A. Fischer, and C. G. Clark, 1999, The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*: *Mol Microbiol*, v. 32, p. 1013-21.
- Tovar, J., G. León-Avila, L. B. Sánchez, R. Sutak, J. Tachezy, M. v. d. Giezen, M. Hernández, M. Müller, and J. M. Lucocq, 2003, Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation: *Nature*, v. 426, p. 172-176.
- Tsaousis, A. D., E. Nyvtova, R. Sutak, I. Hrdy, and J. Tachezy, 2014, A nonmitochondrial hydrogen production in *Naegleria gruberi*: *Genome Biol Evol*, v. 6, p. 792-9.
- Tu, S., S. J. Guo, C. S. Chen, C. X. Liu, H. W. Jiang, F. Ge, J. Y. Deng, Y. M. Zhou, D. M. Czajkowsky, Y. Li, B. R. Qi, Y. H. Ahn, P. A. Cole, H. Zhu, and S. C. Tao, 2015, YcgC represents a new protein deacetylase family in prokaryotes: *Elife*, v. 4.
- Uchimura, H., H. Enjoji, T. Seki, A. Taguchi, N. Takaya, and H. Shoun, 2002, Nitrate reductase-formate dehydrogenase couple involved in the fungal denitrification by *Fusarium oxysporum*: *J Biochem*, v. 131, p. 579-86.
- Unsel, M., J. R. Marienfeld, P. Brandt, and A. Brennicke, 1997, The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides: *Nature Genetics*, v. 15, p. 57-61.
- Urbaniak, M. D., M. L. Guther, and M. A. Ferguson, 2012, Comparative SILAC proteomic analysis of *Trypanosoma brucei* bloodstream and procyclic lifecycle stages: *PLoS One*, v. 7, p. e36619.
- Urbaniak, M. D., D. M. Martin, and M. A. Ferguson, 2013, Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of *Trypanosoma brucei*: *J Proteome Res*, v. 12, p. 2233-44.
- Vahrenholz, C., G. Riemen, E. Pratje, B. Dujon, and G. Michaelis, 1993, Mitochondrial DNA of *Chlamydomonas reinhardtii*: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication: *Curr Genet*, v. 24, p. 241-7.
- Van Den Abbeele, J., G. Caljon, K. De Ridder, P. De Baetselier, and M. Coosemans, 2010, *Trypanosoma brucei* modifies the tsetse salivary composition, altering the fly feeding behavior that favors parasite transmission: *PLoS Pathog*, v. 6, p. e1000926.
- van der Giezen, M., K. A. Sjollema, R. R. Artz, W. Alkema, and R. A. Prins, 1997, Hydrogenosomes in the anaerobic fungus *Neocallimastix frontalis* have a double membrane but lack an associated organelle genome: *FEBS Lett*, v. 408, p. 147-50.
- van der Wielen, P. W., H. Bolhuis, S. Borin, D. Daffonchio, C. Corselli, L. Giuliano, G. D'Auria, G. J. de Lange, A. Huebner, S. P. Varnavas, J. Thomson, C. Tamburini, D. Marty, T. J. McGenity, and K. N. Timmis, 2005, The enigma of prokaryotic life in deep hypersaline anoxic basins: *Science*, v. 307, p. 121-3.
- Van Hellemond, J. J., M. Klockiewicz, C. P. Gaasenbeek, M. H. Roos, and A. G. Tielens, 1995, Rhodoquinone and complex II of the electron transport chain in anaerobically functioning eukaryotes: *J Biol Chem*, v. 270, p. 31065-70.
- van Hellemond, J. J., F. R. Opperdoes, and A. G. Tielens, 2005, The extraordinary mitochondrion and unusual citric acid cycle in *Trypanosoma brucei*: *Biochem Soc Trans*, v. 33, p. 967-71.
- van Ooijen, G., K. Knox, K. Kis, F. Y. Bouget, and A. J. Millar, 2012, Genomic transformation of the picoeukaryote *Ostreococcus tauri*: *J Vis Exp*, p. e4074.
- Velez, N., C. A. Brautigam, and M. A. Phillips, 2013, *Trypanosoma brucei* S-adenosylmethionine decarboxylase N terminus is essential for allosteric activation by the regulatory subunit prozyme: *J Biol Chem*, v. 288, p. 5232-40.
- Vickerman, K., 1969, On the surface coat and flagellar adhesion in trypanosomes: *J Cell Sci*, v. 5, p. 163-93.
- Vickerman, K., 1985, Developmental cycles and biology of pathogenic trypanosomes: *Br Med Bull*, v. 41, p. 105-14.

- Vignais, P. M., and B. Billoud, 2007, Occurrence, classification, and biological function of hydrogenases: an overview: *Chem Rev*, v. 107, p. 4206-72.
- Vincent, K. A., A. Parkin, and F. A. Armstrong, 2007, Investigating and exploiting the electrocatalytic properties of hydrogenases: *Chem Rev*, v. 107, p. 4366-413.
- Visvesvara, G. S., H. Moura, and F. L. Schuster, 2007, Pathogenic and opportunistic free-living amoebae: *Acanthamoeba* spp., *Balamuthia mandrillaris*, *Naegleria fowleri*, and *Sappinia diploidea*: *FEMS Immunol Med Microbiol*, v. 50, p. 1-26.
- von der Heyden, S., E. E. Chao, K. Vickerman, and T. Cavalier-Smith, 2004, Ribosomal RNA phylogeny of bodonid and diplomonid flagellates and the evolution of euglenozoa: *J Eukaryot Microbiol*, v. 51, p. 402-16.
- Wang, J., P. Cieplak, and P. Kollman, 2000a, How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules?: *Journal of Computational Chemistry*, v. 21.
- Wang, J., P. Cieplak, and P. A. Kollman, 2000b, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?: *Journal of Computational Chemistry*, v. 21, p. 1049-1074.
- Wang, Y., L. Liu, Z. Wei, Z. Cheng, Y. Lin, and W. Gong, 2006, Seeing the process of histidine phosphorylation in human bisphosphoglycerate mutase: *J Biol Chem*, v. 281, p. 39642-8.
- Weiss, W., F. Weiland, and A. Gorg, 2009, Protein detection and quantitation technologies for gel-based proteome analysis: *Methods Mol Biol*, v. 564, p. 59-82.
- White, R. R., S. Miyata, E. Papa, E. Spooner, K. Gounaris, M. E. Selkirk, and K. Artavanis-Tsakonas, 2011, Characterisation of the *Trichinella spiralis* deubiquitinating enzyme, TsUCH37, an evolutionarily conserved proteasome interaction partner: *PLoS Negl Trop Dis*, v. 5, p. e1340.
- Wickstead, B., K. Ersfeld, and K. Gull, 2002, Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei*: *Mol Biochem Parasitol*, v. 125, p. 211-6.
- Wickstead, B., K. Ersfeld, and K. Gull, 2003, The frequency of gene targeting in *Trypanosoma brucei* is independent of target site copy number: *Nucleic Acids Res*, v. 31, p. 3993-4000.
- Wickstead, B., K. Ersfeld, and K. Gull, 2004, The small chromosomes of *Trypanosoma brucei* involved in antigenic variation are constructed around repetitive palindromes: *Genome Res*, v. 14, p. 1014-24.
- Willert, E. K., R. Fitzpatrick, and M. A. Phillips, 2007, Allosteric regulation of an essential trypanosome polyamine biosynthetic enzyme by a catalytically dead homolog: *Proc Natl Acad Sci U S A*, v. 104, p. 8275-80.
- Wirtz, E., S. Leal, C. Ochatt, and G. A. Cross, 1999, A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*: *Mol Biochem Parasitol*, v. 99, p. 89-101.
- Wisniewski, J. R., D. F. Zielinska, and M. Mann, 2011, Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method: *Anal Biochem*, v. 410, p. 307-9.
- Wisniewski, J. R., A. Zougman, N. Nagaraj, and M. Mann, 2009, Universal sample preparation method for proteome analysis: *Nat Methods*, v. 6, p. 359-62.
- Woese, C. R., O. Kandler, and M. L. Wheelis, 1990, Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya: *Proc Natl Acad Sci U S A*, v. 87, p. 4576-9.
- Yabuki, A., T. Nakayama, N. Yubuki, T. Hashimoto, K. Ishida, and Y. Inagaki, 2011, *Tsukubamonas globosa* n. gen., n. sp., a novel excavate flagellate possibly holding a key for the early evolution in "Discoba": *J Eukaryot Microbiol*, v. 58, p. 319-31.

- Yabuki, A., and A. Tame, 2015, Phylogeny and Reclassification of *Hemistasia phaeocysticola* (Scherffel) Elbrachter & Schnepf, 1996: *J Eukaryot Microbiol*, v. 62, p. 426-9.
- Yubuki, N., V. P. Edgcomb, J. M. Bernhard, and B. S. Leander, 2009, Ultrastructure and molecular phylogeny of *Calkinsia aureus*: cellular identity of a novel clade of deep-sea euglenozoans with epibiotic bacteria: *BMC Microbiol*, v. 9, p. 16.
- Yubuki, N., A. G. Simpson, and B. S. Leander, 2013, Reconstruction of the feeding apparatus in *Postgaardia mariagerensis* provides evidence for character evolution within the Symbiontida (Euglenozoa): *Eur J Protistol*, v. 49, p. 32-9.
- Zhao, B., and M. R. Waterman, 2007, Novel properties of P450s in *Streptomyces coelicolor*: *Drug Metab Rev*, v. 39, p. 343-52.
- Zilberstein, D., 2015, Proteomic analysis of posttranslational modifications using iTRAQ in *Leishmania*: *Methods Mol Biol*, v. 1201, p. 261-8.
- Zuckerandl, E., and L. Pauling, 1965, Molecules as documents of evolutionary history: *J Theor Biol*, v. 8, p. 357-66.