Title page

Investigation of conserved flagellum proteins in *Trypanosoma brucei*

Samantha Jane Borrett

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

Awarded by Oxford Brookes University

Submission date 05/2015

Declaration

I declare that no material contained in this thesis has been used in any other submission for an academic award.

I confirm that all the research and findings presented in this thesis are my own work unless otherwise indicated through the use of a clear referencing system.

Minor collaborations are stated at the beginning of the relevant chapter (see sections 5.1 and 6.1)

“Even when we have fathomed out the riddle of the centriole, I suspect that the biologist will retain his or her fascination with this organelle, just as stars fascinate adults.”

D. N. Wheatley, 1982

Abstract

The single celled protozoan parasite *Trypanosoma brucei* is an excellent model organism to study eukaryotic cilia and flagella as it has a single flagellum that remains assembled throughout the cell cycle. The new flagellum assembles in a known position relative to the old flagellum, therefore creating a model system of identifiable organelle generations. In additional to a sequenced genome, there are many reverse genetics tools developed for *T. brucei* which makes the functional analysis of proteins possible. More than 300 proteins have been identified as components of the *T. brucei* flagellum but functional analysis of the majority of these proteins has not been carried out to date.

This project used a bioinformatics approach to identify potential flagellum proteins in *T. brucei* that were also conserved in *Homo sapiens,* thereby identifying potential ciliopathy candidates. Candidate proteins were confirmed as flagellum components through endogenous localisation techniques and co-localisation studies. Functional analysis was performed using inducible RNAi cell lines. Light and electron microscopy techniques were used for phenotypic analysis.

Through bioinformatics analysis a novel family of coiled-coil TPH domain-containing proteins were identified that are highly conserved in flagellated eukaryotes. There are three TPH domain-containing proteins conserved in *T. brucei* that all have a role in flagellum length control and cell morphogenesis. In all three cases protein ablation has a detrimental effect on cellular motility.

This work provides further understanding into the complexities of flagellum biogenesis in *T. brucei* and the downstream effects on cell motility and morphogenesis.

Acknowledgements

Firstly I would like to thank my supervisor, Dr Sue Vaughan, for giving me the opportunity to carry out my PhD in her lab. I would also like to thank members of the Gull lab for accommodating me when I needed to use various reagents and equipment.

Huge thanks to Katie Towers and Louise Hughes for your unwavering friendship, support and cups of tea!

Thank you to all members of S208a and WDO for listening ears and sharing cake ☺ Too many to name but you know who you are!

Mum and Dad, words cannot express my gratitude towards you for putting up with me over the last few years. To all my family and friends, thank you for your patience and feigned interest in trypanosomes!

Will, for endless support and consultant excel skills. You had faith in me when I had none, I could not have done it without you.

And lastly, thanks to the Nigel Groome studentship from Oxford Brookes University that funded the project.

This thesis is dedicated to my nan, June Rose Collins (1925 – 2015), who passed away shortly after my *viva voce*.

Table of contents

List of figures

List of tables

List of abbreviations used

Abbreviation	Term
µm	Micrometers
1F	A single flagellum
1K0N	One kinetoplast and no nucleus/zoid cell/anuclete
1K1N	One kinetoplast and one nucleus
1K1N1F	One kinetoplast, one nucleus and one flagellum
2D	Two dimensions
2F	Two flagella
2K1N	Two kinetoplasts and one nucleus
2K2N	Two kinetoplasts and two nuclei
2KDivN	Two kinetoplasts and a dividing nucleus
3D	Three dimensions
a.u.	Arbitrary units
aa	Amino acids
AMt	Axonemal microtubules
Ant.	Anterior end
A-tubule	A microtubule
BB	Basal body
BB/C	Basal body/centriole
BBS	Bardet-Biedl syndrome
BLAST	Basic local alignment search tool
BLASTp	Basic local alignment search tool of proteins
BLD	Bald mutant
bp	Base pairs
BSF	Bloodstream form
BtubA	Bacterial tubulin A
BtubB	Bacterial tubulin B
B-tubule	B microtubule
C-	Carboxyl [terminus]
C1	Central pair microtubule 1
C2	Central pair microtubule 2
CC	Coiled-coil
CCDC	Coiled-coil domain-containing
CDD	Conserved domain database

CEP/cep	Centrosomal protein
CFAP	Cilium or flagellum associated protein
CORD	Cone and rod dystrophy
CPC	Central pair complex
CrFP	*Chlamydomonas reinhardtii* flagellar proteome
C-tubule	C microtubule
DAPI	4,6-diamidino-2-phenylinidole
DivK1N	A dividing kinetoplast and one nucleus
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
DRC	Dynein regulatory complex
DTT	Dithiothreitol
DUF	Domain of unknown function
EM	Electron microscopy
FAP	Flagellum associated protein
FAZ	Flagellum attachment zone
FC	Flagellar connector
FCS	Fetal calf serum
FIB-SEM	Focussed ion beam SEM
FP	Flagellar pocket
FtsZ	Filamenting temperature-sensitive mutant Z
GFP	Green fluorescent protein
HAT	Human African trypanosomiasis
HMM	Hidden Markov model
hrs	Hours
IDA	Inner dynein arm
IFT	Intraflagellar transport
JME	Juvenile myoclonic epilepsy
K	Kinetoplast
KDa	Kilodalton
kDNA	Kinetoplast DNA
KFZ	Kinetoflagellar zone
LB	Luria Bertani
LCA	Lebers congenital amaurosis
LECA	Last eukaryotic common ancestor

LRR	Leucine-rich repeat
LRTP	Leucine-rich testis protein
MAPs	Microtubule associated proteins
MBB	Mature basal body
MD	Macular dystrophy
MKS	Meckel Gruber syndrome
Mt	Microtubule
MTOC	Microtubule organising centre
MtQ	Microtubule quartet
N	Nucleus
N-	Amino [terminus]
NAB	Nucleus-associated body
NCBI	National Centre for Biotechnology Information
N-DRC	Nexin-Dynein regulatory complex
NF	New flagellum
NIMA	Never in mitosis A
nm	Nanometers
NM	Nuclear mitosis
NP-40	Nonyl phenoxypolyethoxylethanol
NPHP	Nephronophthisis
NRK	NIMA related kinase
ns	Not significant
ODA	Outer dynein arm
OF	Old flagellum
OFD	Oral-facial-digital syndrome
OMIM	Online Mendelian inheritance of man
PBB	Pro-basal body
PBS	Phosphate buffered saline
PCD	Primary ciliary dyskinesia
PCF	Procyclic form
PCM	Pericentriolar material
PCR	Polymerase chain reaction
pEnT	Plasmid for endogenous tagging
pf	Protofilament
Pfam	Protein families database

PFR	Paraflagelllar rod
PKD	Polycystic kidney disease
PLK	Polo like kinase
Pos.	Posterior end
pPOT	Plasmid for PCR only tagging
R	Pearson correlation coefficient
RBH	Reciprocal best BLAST hit
Rib	Ribbon
RNA	Ribonucleic acid
RNAi	RNA interference
RP	Retinitis pigmentosa
RS	Radial spoke
s.d.	Standard deviation
SAS	Spindle assembly abnormal
SBF-SEM	Serial block face SEM
SCA	Spinocerebella ataxia
SDM-79	Semi defined media 1979
SDS	sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEM	Scanning electron microscopy
SmOx	Single marker Oxford
SMt	Subpellicular microtubule
SPAG	Sperm associated antigen
SPB	Spindle pole body
S-phase	Synthesis phase
STD	Standard deviation
TAC	Tripartite attachment complex
TbCMF	*Trypanosoma brucei* components of motile flagellar
TbFP	*Trypanosoma brucei* flagellar proteome
TEM	Transmission electron microscopy
TEMED	Tetramethylethylenediamine
TPH	Trichohyalin-plectin homology
UNIMOD	Universal module
US	Ushers syndrome
v/v	Volume/volume

VSG	Variant surface glycoprotein
w/v	Weight/volume
WDR	Tryptophan-aspartic acid repeats
WHO	World health organisation
WT	Wild type
YFP	Yellow fluorescent protein
γTuRC	Gamma tubulin ring complex

Please note;

The terms cilia and flagella refer to the same structure.

The terms centriole and basal body refer to the same structure.

The multiple terminologies for these structures has arisen from co-discovery in previously separate fields of research which now have a large degree of overlap.

1. Introduction

1.1 Microtubules

1.1.1 Tubulin

Tubulins are a superfamily of cytoskeletal proteins found in eukaryotes and are the main component of microtubules. Members of the tubulin superfamily are denoted α, β, γ, δ, ε, ζ, η, θ, ι and κ (alpha, beta, gamma, delta, epsilon, zeta, eta, theta, iota and kappa respectively) in the order of discovery (Chang and Stearns, 2000; Cleveland et al., 1978; Dutcher and Trabuco, 1998; Oakley and Oakley, 1989; Ruiz et al., 2000; Vaughan et al., 2000). There is an 11th known member of the tubulin super family which has been detected in *Xenopus laevis* and named 'cryptic tubulin' (K. Gull, personal communication in (Stearns, 2005)).

The most predominant forms of tubulin in eukaryotes are α, β and γ tubulins, which have been detected in all eukaryotes examined. The additional tubulins are not ubiquitously conserved, δ and ε tubulin has only been found in humans and some single celled organisms (Chang and Stearns, 2000; Garreau de Loubresse *et al.*, 2001; O'Toole *et al.*, 2003). The tubulins ζ and η have been reported in P*aramecium tetraurelia*, *Trypanosoma brucei* and *Leishmania major* (Ruiz *et al.*, 2000; Vaughan *et al.*, 2000) and most recently, ζ was localised to the basal foot apparatus in *Xenopus laevis* (Turk *et al.*, 2015). The most recently discovered tubulins, θ, ι and κ have only been described in *P. tetraurelia* and do not appear to be conserved in the closely related ciliate *Tetrahymena thermophila* (Libusova and Draber, 2006). The expansion of the tubulin family in *P. tetraurelia* could be due to three whole genome duplications (Aury *et al.*, 2006) or could be attributed to the extra microtubular structures with a specialised function in this organism, which include the post ciliary ribbon and kinetodesmal fibres. However, a recent analysis of tubulins commented that κ tubulin is actually an isoform of α tubulin and similarly that θ and ι tubulin are actually isoforms of β tubulin (Findeisen *et al.*, 2014). Findeisen and colleagues (2014) also propose the subfamily of η tubulin is actually part of the ζ tubulin subfamily.

It was though that ζ tubulin was restricted to kinetoplastids (Vaughan *et al.*, 2000) but a recent in-depth bioinformatics study of 504 organisms (Findeisen *et al.*, 2014) has detected ζ tubulin in ciliates and oomycetes amongst others but absent from apicomplexa, fungi and insects. It is interesting that in the 504 organisms analysed, it was common for multi copies of α, β and γ tubulin to be present but if δ, ε, or ζ were conserved, it was usually as a single copy.

Testament to how important tubulins are for basal bodies/centriole structure and function can be drawn from the degree of conservation and the analysis on mutant cell lines (Table 1). Knockdown or mutation of δ, ε, and η- tubulin in some single celled model organisms have caused phenotypes which indicate that the expanded tubulin family members are crucial for normal basal body formation (Dutcher et al., 2002; Dutcher and Trabuco, 1998; Garreau de Loubresse et al., 2001; O'Toole et al., 2003; Ross et al., 2013; Ruiz et al., 2000).

Table 1: Functions of tubulins

Collation of published experimental data on δ, ε, and η tubulin in protozoa. Adapted from (Stearns, 2005).

Tubulin	Localised to	Organism	Phenotype	Reference
δ	Basal body	*C. reinhardtii*	Loss of C microtubule from basal body	(Dutcher and Trabuco, 1998; O'Toole *et al.*, 2003)
	-	*P. tetraurelia*	Loss of C microtubule from basal body	(Garreau de Loubresse *et al.*, 2001)
	Basal body	*T. brucei*	Loss of basal body organisation with doublet and singlet microtubules. Immotile flagella	(Gadelha *et al.*, 2006)
ε	Basal body	*C. reinhardtii*	Loss of B and C microtubule from basal body	(Dutcher *et al.*, 2002)
	Basal body	*T. thermophila*	Incomplete microtubule triplets	(Ross *et al.*, 2013)
η	-	*P. tetraurelia*	Delocalisation of γ tubulin	(Ruiz *et al.*, 2000)

Here I have focused on mammalian and protozoan tubulins but it is important to note that tubulin is also conserved in plants. Only α, β, and γ tubulin have been reported in *Arabidopsis thaliana* to date (Kopczak *et al.*, 1992; Liu *et al.*, 1994; Snustad *et al.*, 1992). Absence of further tubulin family members is likely to be due to the lack of centrioles in higher plants.

Although tubulin is not directly conserved in prokaryotes there are functional homologs that are thought to derive from a common ancestor, which include a tubulin-related protein, FtsZ (Erickson, 1995; Erickson, 1997) and bacterial tubulin A (BtubA) and bacterial tubulin B (BtubB) (Jenkins et al., 2002). BtubA and BtubB can form a tubule structure in bacteria thought to be composed of 5 protofilaments (Pilhofer *et al.*, 2011; Schlieper *et al.*, 2005).

1.1.2 Microtubule structure

Microtubules are composed of α and β tubulin heterodimers which are assembled into protofilaments, sheets and then tubules (Figure 1; i→iv). Microtubules are typically composed of 13 tubulin protofilaments (Tilney *et al.*, 1973). Other numbers of protofilaments in microtubules can occur (Sui and Downing, 2010) but they shall not be discussed here. When the protofilament sheet (Figure 1; iii) becomes a tubule there is a 'seam' of α to β tubulin interactions (Figure 1; iv), which is the weakest point of the tubule (Simon and Salmon, 1990). Due to the helical arrangement of the protofilaments, the majority of tubulin interactions are 'α to α' or 'β to β' (Figure 1; v), which are stronger than 'α to β' bonds at the seam. Due to the composition of microtubules from α and β tubulin heterodimers, microtubules have an intrinsic polarity, which infers what end of the microtubule is the 'growing' end. The microtubule extends from the plus (+) end. Microtubules are nucleated from a γ-tubulin ring complex (γTuRC), which is always located at the minus (-) end of the microtubule (for review see (Kollman *et al.*, 2011)). Recently the conformation of γTuRC was solved (Kollman *et al.*, 2015).

Singlet microtubules, as stated previously, are made of 13 protofilaments but in axonemes and basal bodies there are doublet and triplet microtubule arrangements. Doublet or triplet microtubules are formed when microtubules 'share' walls of a tubule (Figure 2). In these cases the B and C tubule is comprised of 10 protofilaments instead of 13 (Figure 2). Microtubule doublet and triplets are not solely made up of α and β tubulin protofilaments, there are protofilament ribbon proteins (Nojima *et al.*, 1995) (Figure 2; green). Known ribbon proteins include Rib43a, Rib72 and tektins (Ikeda *et al.*, 2003; Linck *et al.*, 1985; Norrander *et al.*, 2000; Steffen and Linck, 1988). The inclusion of additional ribbon proteins in singlet microtubules has not been reported.

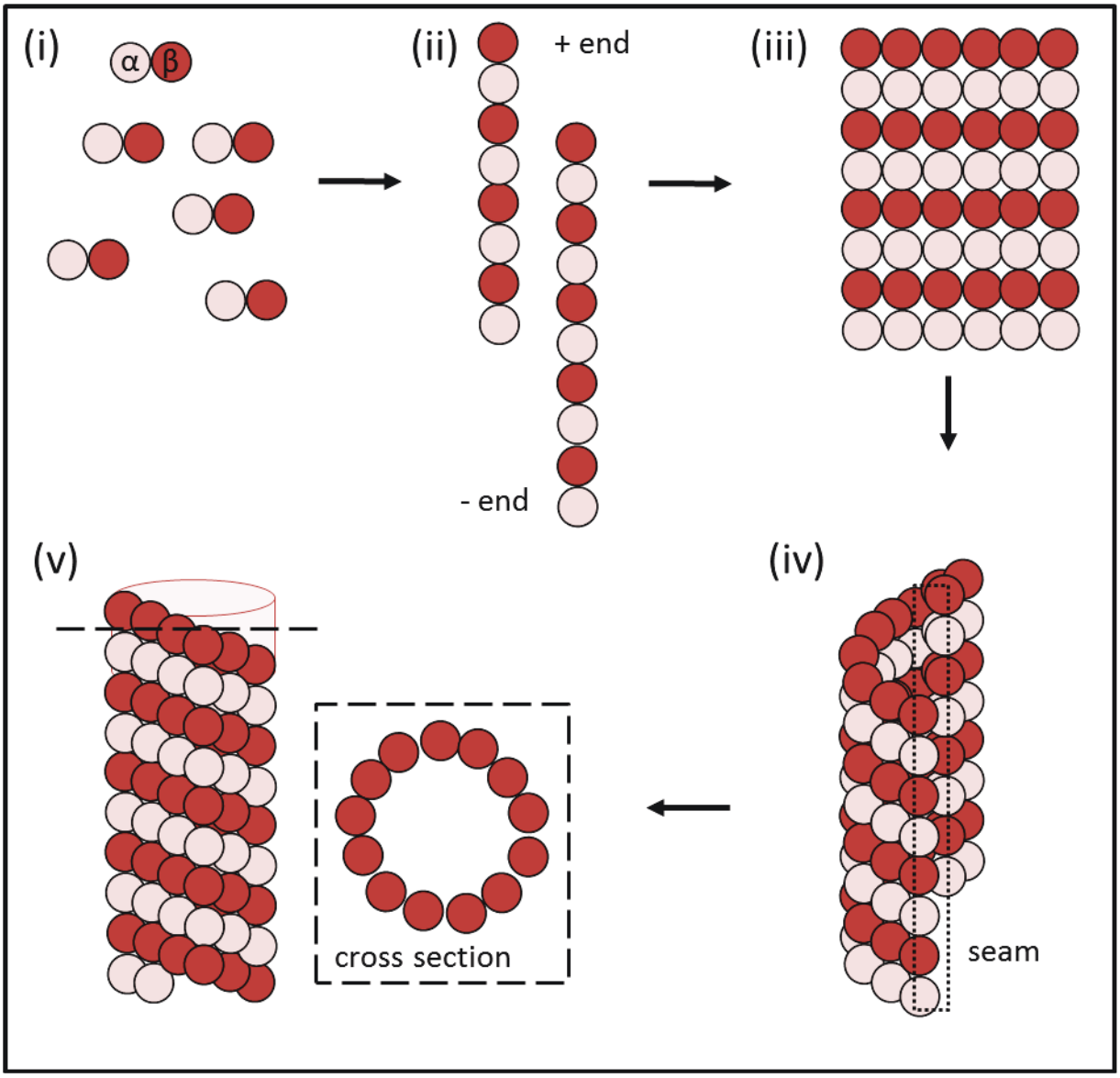


Figure 1: Microtubule assembly

Tubulin subunits exist as α and β heterodimers (i), which form tubulin protofilaments (ii). These tubulin protofilaments come together in a sheet array (iii) which can fold to become a microtubule (iv) with a seam (dotted line) of α and β tubulin interactions. (v) A microtubules with cross section view (dash line). Based on (Pampaloni and Florin, 2008).

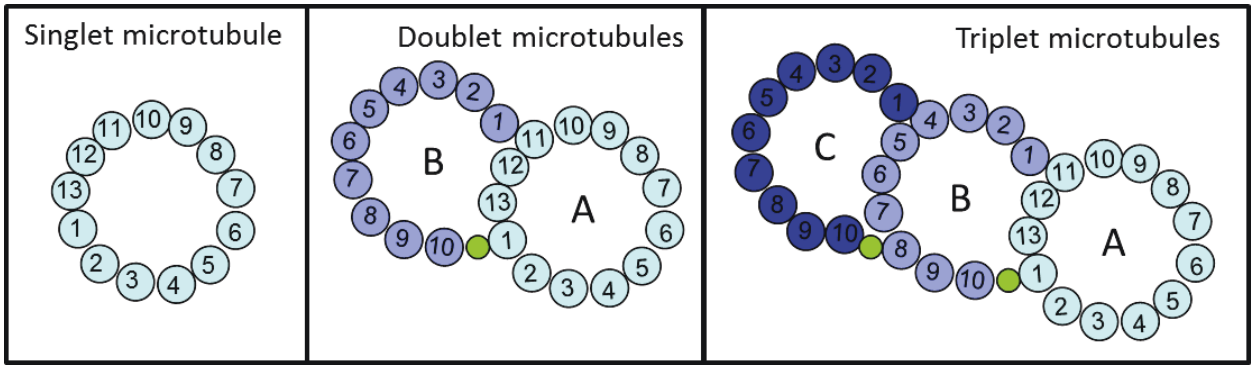


Figure 2: Microtubule arrangements

The most common conformations that microtubules occur in are singlets, doublets and triplets. If microtubules occur in doublets or triplets them the microtubules are referred to as 'A', 'B' or 'C' microtubules (shown as different shades of blue in diagram). The A microtubule typically has 13 tubulin protofilaments whilst the B and C microtubules only have 10 tubulin protofilaments. The B and C microtubules also incorporate a non-tubulin protofilament (green). Figure adapted from and protofilament numbering according to (Linck and Stephens, 2007).

1.1.3 Microtubule organising centres

In eukaryotes, microtubule organising centres (MTOCs) are the point from which microtubules are nucleated from by a γTuRC. Examples of MTOCS include basal bodies, centrosomes and spindle pole bodies - this list of MTOCs is not finite. There is also research to suggest that the Golgi apparatus is MTOC capable (Zhu and Kaverina, 2013). Mature basal bodies organise the axoneme microtubules, pro-basal bodies do not assemble an axoneme (Sherwin and Gull, 1989), the central pair microtubules of the axoneme (Figure 5) are nucleated from the basal plate (McKean *et al.*, 2003). Although structurally analogous to basal bodies, mature centrioles are part of an MTOC called a centrosome. The centrosome is composed of a pair of centrioles and a protein density termed the pericentriolar material (PCM). The PCM is recruited to the centrosome by the centrioles and is composed of conserved proteins that are important for centriole maturation and/or mitotic spindle assembly (see review (Fu *et al.*, 2015)). PCM components include kinases, structural coiled-coil proteins and microtubule associated proteins (MAPs). Twenty three novel centrosome proteins (Cep) were discovered in the original centrosome proteome (Andersen *et al.*, 2003). An extensive proteome of the

centrosome expanded the list of human centrosome proteins to 279, although not all have been functionally characterised (Jakobsen *et al.*, 2011).

Yeast are an important model organism for cell cycle analysis and many divergent orthologs of centrosome associated proteins such as polo like kinase (PLK) and centrin are conserved (Ohkura *et al.*, 1995; Spang *et al.*, 1993). However, yeast do not have a centriole or centrosome, instead their MTOC is known as a spindle pole body (SPB) and has a trilaminar plaque appearance, which is quite different to the 9-fold symmetry of a centriole.

There are several cases of unconventional MTOCs. The protozoan parasite, *Toxoplasma gondii,* in addition to centrioles has a separate MTOC called the apical polar ring that nucleates the subpellicular microtubules of the parasite (Russell and Burns, 1984). The slime mould, *Dictyostelium discoideum* has an MTOC termed the nucleus-associated body (NAB), which is structurally more similar to a SPB than a centriole (Kalt and Schliwa, 1996). A mammalian cell organelle, the deuterosome, is not well defined but must also be classed as an MTOC due to the ability to organise assembly of multiple pro-basal bodies (PBB) *de novo* (Klos Dehring *et al.*, 2013).

1.2 Kinases

Kinases are important regulators for a multitude of activities within the cell. Through phosphorylation they can activate or inhibit the activity of interacting partner proteins. Several kinases are known to be important for tightly regulated centriole/basal body maturation, duplication and segregation. There are 176 protein kinases within the *T. brucei* proteome (Parsons *et al.*, 2005) but only three kinase families will be discussed here for their relevance to basal bodies, cilia and the work presented in this thesis.

1.2.1 Never in mitosis A kinases

Never in mitosis A (NIMA) kinases are a superfamily of serine/threonine kinases, which are highly conserved in eukaryotes. The first NIMA kinase was discovered in the fungus, *Aspergillus nidulans* (Osmani *et al.*, 1988) and since then it has been established that

NIMA kinases and NIMA-related kinases (Nek) are highly conserved across eukaryotes. The defining feature of the NIMA/Nek kinase super group is a protein kinase domain at the N terminus but a non-conserved C terminus (O'Connell *et al.*, 2003). Within the NIMA/Nek superfamily there is a subgroup of kinases which are proposed to have co-evolved with centrioles/basal bodies in flagellated organisms (review by (Quarmby and Mahjoub, 2005). *C. reinhardtii* NEKs were detected in the flagellar proteome (Pazour *et al.*, 2005) and have been localised to the flagella (Bradley and Quarmby, 2005) and basal bodies (Mahjoub *et al.*, 2004). Nek2 orthologs are involved in regulating the G1-G2 tether between centrioles in a range of model organisms studied to date (Fry, 2002; Graf, 2002; Prigent *et al.*, 2005). Testament to their importance in centriole and flagellum biogenesis, Nek kinases have also been linked to a ciliopathy; polycystic kidney disease (Mahjoub *et al.*, 2005; Surpili *et al.*, 2003; Upadhya *et al.*, 2000).

There are 20 NEKs within the *T. brucei* kinome (Parsons *et al.*, 2005), one has been localised to the basal body (Pradel *et al.*, 2006). This Nek is named *T. brucei* NIMA-related kinase C (TbNRKC) after TbNRCA and TbNRKB (Gale *et al.*, 1994; Gale and Parsons, 1993). The kinase domain of TbNRKC has a high level of sequence identity with human Nek1 and Nek2. The overexpression of TbNRKC prevented basal body segregation (Pradel *et al.*, 2006), which is an observation that fits the role of Nek2 orthologs being involved in G1-G2 tether regulation (Fry, 2002).

1.2.2 Polo-like kinases

The first polo-like kinase (PLK) was discovered in *D. melanogaster* (Sunkel and Glover, 1988). Defining features of PLKs are the polo box domain(s), PEST motifs and, of course, a protein kinase domain. There are four clades of the polo-like kinase (PLK) family in chordata, PLK1-4 (Winkles and Alberts, 2005). PLK1 is a mitotic kinase and disruption of normal function can lead to cancer (Archambault *et al.*, 2015). PLK4 is involved in regulating and licensing centriole duplication (Kleylein-Sohn *et al.*, 2007) (Rodrigues-Martins *et al.*, 2007).

T. brucei has a single PLK, TbPLK (Hammarton *et al.*, 2007). TbPLK appears to be involved in regulating a range of functions including flagellum attachment (Lozano-Nunez *et al.*,

2013; Sun and Wang, 2011; Umeyama and Wang, 2008), basal body segregation (Hammarton *et al.*, 2007), flagellum inheritance (de Graffenried *et al.*, 2013; Ikeda and de Graffenried, 2012) and cytokinesis (Hammarton *et al.*, 2007; Li *et al.*, 2010). The localisation of TbPLK is transient along the flagellum path (Ikeda and de Graffenried, 2012; Umeyama and Wang, 2008) and therefore helps to explain the many and varied roles of this kinase. TbPLK is not involved in mitosis (Ikeda and de Graffenried, 2012).

1.2.3 Aurora kinases

Aurora kinases are a family of serine/threonine protein kinases that are highly conserved in eukaryotes. There are 3 subgroups of Aurora kinases; A, B and C. Aurora kinase A appears to be restricted to mammals whilst Aurora kinases B and C are also conserved in earlier branching eukaryotes (for reviews see (Bolanos-Garcia, 2005; Carmena and Earnshaw, 2003). Aurora kinases are associated with regulating mitosis and cytokinesis in several eukaryotes (Davids *et al.*, 2008; Li *et al.*, 2008; Reboutier *et al.*, 2013; Reininger *et al.*, 2011). Aurora A regulates ciliogenesis through interaction with trichoplein, cenexin and ninein in mammalian cells (Ibi *et al.*, 2011; Inoko *et al.*, 2012; Kasahara *et al.*, 2014). As well as being involved with the mature basal body and ciliogenesis, Aurora A is also involved with centrioles at the centrosome. Aurora A is part of the centriole duplication cycle by interacting with Cep192 (spd-2 in *D. melanogaster*), PLK1 (Joukov *et al.*, 2014) and centrin (Lukasiewicz *et al.*, 2011).

Aurora kinases have also been identified in several protozoa, including, *D. discoideum* (Li *et al.*, 2008), *Leishmania major* (Siman-Tov *et al.*, 2001), *T. brucei* (Tu *et al.*, 2006) and *Plasmodium falciparum* (Reininger *et al.*, 2011). To date, three Aurora kinases have been identified in *T. brucei*; TbAuk1, -2 and -3 (Li and Wang, 2006; Tu *et al.*, 2006). TbAuk1 is localised to the nucleus in procyclic form (PCF) cells (Li *et al.*, 2009; Tu *et al.*, 2006) and bloodstream form (BSF) cells (Li and Wang, 2006). The localisation of TbAuk2 and TbAuk3 is unknown. The Aurora kinases of *P. falciparum* are known as Pfark-1, -2, -3. These may have a similar role to the mammalian mitotic Aurora kinases as they localise to the SPB of the mitotic nuclei (Reininger *et al.*, 2011).

1.3 Centrin

1.3.1 Conservation of centrin

Centrin was first discovered in the algae, *Tetraselmis striata* (Salisbury et al., 1984) and has since been identified in many eukaryotes from protozoa (Mahajan *et al.*, 2008) to humans (Errabolu *et al.*, 1994). Centrins are conserved in land plants that lack centrioles but their localisation and function has shifted to plasmodesmata (Blackman *et al.*, 1999) and the cell plate (Harper *et al.*, 2000).

A universal feature of centrin proteins are calcium-binding EF-hand domains although the number can vary from 1 to 4 (Errabolu *et al.*, 1994; Gogendeau *et al.*, 2008; Weber *et al.*, 1994).The *C. reinhardtii* centrin ortholog (CrCenp) has been localised to several contractile fibres (Geimer *et al.*, 1998; Geimer and Melkonian, 2005; McFadden *et al.*, 1987), which are known to be calcium sensitive (Salisbury *et al.*, 1984).

Centrin orthologs have also been localised to several MTOCs in a range of eukaryotic organisms (for reviews see (Salisbury, 1995; Salisbury, 2007). The *Saccharomyces cerevisiae* ortholog, Cdc31p, localises to the SPB (Spang *et al.*, 1993) and in the absence of Cdc31p the SPB does not segregate (Baum *et al.*, 1986).

Centrins are highly conserved in eukaryotes but the number of orthologs can differ significantly between organisms. *C. reinhardtii* has a single centrin ortholog, mice have 4 centrin orthologs (Giessl *et al.*, 2006) and the ciliate *P. tetraurelia* has 49 (Gogendeau *et al.*, 2008). Expression of the 4 mice centrin orthologs is an example of multiple orthologs developing separate and overlapping functions and localisations. In the photoreceptor cells of mice, centrin 1 localises to the connecting cilium. Centrin 2 and 3 localise to the connecting cilium and the basal bodies but centrin 4 only localises to the basal bodies (Trojan *et al.*, 2008). In humans, centrin proteins localises to the centriole in a range of cell types (Laoukili *et al.*, 2000) and are known to be required for centriole duplication (Lutz *et al.*, 2001; Salisbury *et al.*, 2002).

There are 5 centrin proteins conserved in *Trypanosoma brucei* (TbCen1 – TbCen5) (Berriman *et al.*, 2005) with localisation as follows:

- TbCen1 localises to the basal bodies (He *et al.*, 2005).
- TbCen2 localises to the basal bodies, bilobe structure and flagellum (He *et al.*, 2005; Wang *et al.*, 2012).
- TbCen3 localises to the flagellum (Wei *et al.*, 2014).
- TbCen4 localises to the basal bodies and bilobe structure (Selvapandiyan *et al.*, 2007; Shi *et al.*, 2008).
- The localisation of TbCen5 is unknown.

In *T. brucei* when TbCen1 is knocked down there are defects in kinetoplast segregation and cytokinesis (Selvapandiyan *et al.*, 2012). It was purported that ablation of TbCen1 inhibited basal body duplication (He *et al.*, 2005). However, that analysis was performed using the antibody YL1/2 (Kilmartin *et al.*, 1982), which is specific for the mature basal body (Stephan *et al.*, 2007) and therefore does not label the pro-basal body. Ablation of TbCen2 has been reported to cause kinetoplast segregation defects (Selvapandiyan *et al.*, 2012), leads to defects in bilobe duplication and disrupts Golgi apparatus assembly (He *et al.*, 2005). TbCen3 interacts with the inner dynein arm protein complex (Wei *et al.*, 2014) and knockdown leads to a motility defect in PCF cells. TbCen4 has been shown to regulate recruitment of TbCen2 at the bilobe (Wang *et al.*, 2012) and has been implicated in coordinating cytokinesis in PCF cells (Shi *et al.*, 2008). No work has been published on TbCen5

Alterations in centrin levels can lead to abnormal kidney development in *D. rerio* (Delaval *et al.*, 2011) and hydrocephalus in knockout mice models (Ying *et al.*, 2014). These are both known ciliopathy phenotypes (see section 1.5.5 Defects in cilia: Ciliopathies), which highlight the importance of understanding the role of centrin proteins in basal body and flagella function.

1.3.2 Centrin interacting proteins

Many of the centrin interacting proteins have been identified by searching for a key motif; the centrin-binding repeats (CBRs) (Azimzadeh *et al.*, 2009; Kilmartin, 2003;

Stemm-Wolf *et al.*, 2013). Due to the high level of conservation of centrin proteins in eukaryotes, it is not surprising that centrin-interacting proteins are also highly conserved. Although there is likely a multitude of centrin-interacting proteins within eukaryotes, research to date has focussed on two groups; (1) The Sfi/Sfr family and (2) the POC5 family.

Sfi (suppressor of fermentation-induced loss of stress resistance protein 1) and Sfr (Sfi-repeat) proteins are highly conserved in eukaryotes and both possess CBRs. The *H. sapiens* Sfi ortholog, HsSfi1, contains 23 CBRs (Martinez-Sanz *et al.*, 2006), which HsCen2 has been shown to bind to (Martinez-Sanz *et al.*, 2010). The yeast ortholog, Sfi1p, has been shown to contain CBRs and binds with Cdc31p (Kilmartin, 2003). In the ciliate, *Tetrahymena thermophila,* there are 13 Sfr proteins, which contain 3 to 29 Sfi repeats. (Stemm-Wolf *et al.*, 2013).

The POC5 protein family also contain CBRs and interaction has been confirmed *in vitro* between HsPOC5 and HsCen2 (Azimzadeh *et al.*, 2009). POC5 is highly conserved with orthologs in *H. sapiens*, *T. brucei* and *C. reinhardtii* but not in *Drosophilla melanogaster*, *Caenorhabditis elegans, Saccharomyces cerevisiae* or *Plasmodium falciparum* (Azimzadeh *et al.*, 2009).

1.4 Centrioles and basal bodies

1.4.1 Structure and function

The microtubules of cilia/flagella extend from an organelle known as a basal body (Figure 3) (for review see (Marshall, 2008)). Cilia/flagella structure is discussed below, in section 1.5 of this literature review. Basal bodies and centrioles are structurally homologous, exhibiting 9 fold symmetry of triplet microtubules (9+0), or microtubule blades (Dippell, 1967; Dippell, 1968) composed of A, B and C microtubules (Anderson, 1972) (Figure 3). There are exceptions to the canonical structure with the basal body of *C. elegans* consisting of nine singlet microtubules instead of triplets (Pelletier *et al.*, 2006) and the basal body of *Acerentomon microrhinus* having 11 triplets (see review (Carvalho-Santos *et*

al., 2011)) but the most common structure is the nine fold triplet structure (Figure 3) that is conserved in humans and trypanosomes.

Although they are one and the same organelle, the differentiation between a basal body and a centriole is primarily the function of the organelle; When located in the cytoplasm, normally as part of a centrosome, it is called a centriole and if the structure is nucleating a cilium then it is known as a basal body (Hoyer-Fender, 2010; Kobayashi and Dynlacht, 2011). Basal bodies exist as a pair with each pair constituting a mature basal body and a pro-basal body (Renaud and Swift, 1964) the pro-basal body is also called an immature or daughter basal body in the literature. The pro-basal body forms orthogonal from the mature basal body (Alvey, 1986; Guichard *et al.*, 2010). The mature basal body extends a cilium/flagellum but the pro-basal body does not build an axoneme until it matures in the next cell cycle (Lange and Gull, 1996; Sherwin and Gull, 1989). During formation of an axoneme, the A and B microtubules of the mature BB extend and the plasma membrane remodels to encompass the growing axoneme (Dentler, 2005; Garcia-Gonzalo and Reiter, 2012).

Within the lumen of the basal body, at the proximal end is a 9 fold structure, the cartwheel (Figure 3; C). The protein spindle assembly abnormal-6 (SAS-6) is thought to be responsible for seeding the 9-fold symmetry of the cartwheel (Cottee *et al.*, 2011; Kitagawa *et al.*, 2011; Nakazawa *et al.*, 2007) but there are instances of basal body/centriole assembly in the absence of SAS-6 (Fong *et al.*, 2014). The sequential incorporation of A, then B, then C microtubules was originally devised from *Paramecium spp.* (Dippell, 1967; Dippell, 1968) and has since been confirmed in humans (Guichard *et al.*, 2010).

SAS-6 is ubiquitously conserved across eukaryotes that form centrioles/basal bodies (Carvalho-Santos *et al.*, 2010). A divergent SAS-6 ortholog is present in the nematode, *Caenorhabditis elegans* (Leidel *et al.*, 2005). The number of cartwheels stacked at the proximal end of the basal body is not known although in most basal bodies there is a stack approximately 100nm high (Azimzadeh and Marshall, 2010; Geimer and Melkonian, 2004; Guichard *et al.*, 2012; Nakazawa *et al.*, 2007). The flagellate *Trichonympha* has an elongated cartwheel stack of 1500nm (Guichard *et al.*, 2012) which makes it a suitable

model organism to investigate cartwheel structures in, indeed it has been used to resolve the periodicity of the cartwheels in a stack as having a periodicity of 8.5nm at the central hub and 17nm at the periphery of the cartwheel (Guichard *et al.*, 2012).

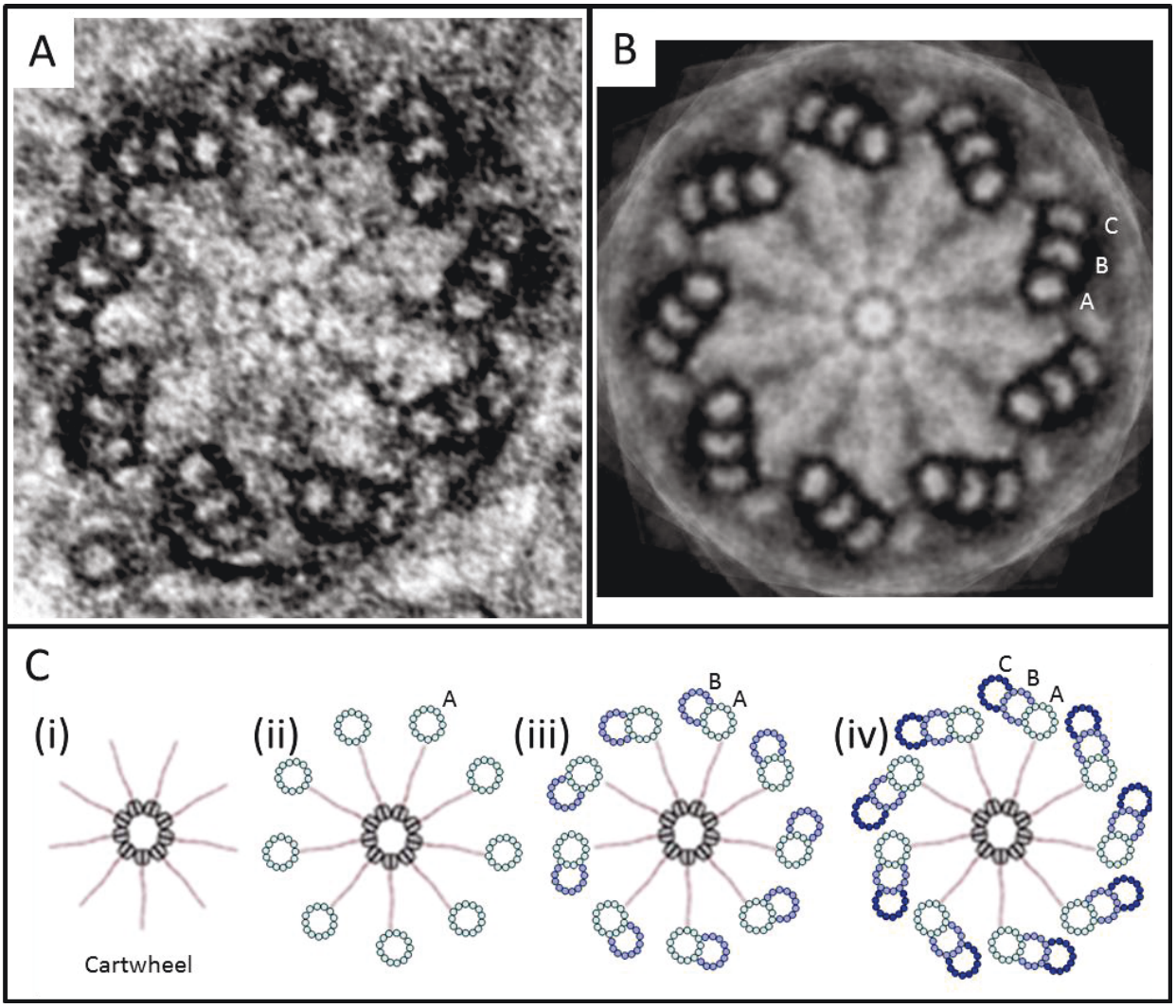


Figure 3: Basal body structure

(A) The 9 + 0 triplet microtubule structure of a basal body viewed in a proximal to distal orientation. The 9 microtubule triplets are arranged in a 9-fold symmetrical pattern. Micrograph of a basal body from *T. brucei* courtesy of Dr K Towers. (B) A 9-fold rotation of the basal body to demonstrate the 9-fold symmetry of the microtubule triplets. Original micrograph in A. (C) Diagram of a cartwheel (i) in the lumen of the basal body at the proximal end. (ii-iv) The microtubule triplets assemble onto the cartwheel scaffold in a step-wise manner. (iv) Each microtubule triplet consists of an A, B and C microtubule. (C) Based on (Dippell, 1968; Guichard *et al.*, 2010).

1.4.2 Basal body duplication

Basal bodies are highly conserved among eukaryotes and therefore it was proposed that the proteins responsible for basal body assembly would also be equally highly conserved. Identifying proteins that were highly conserved in flagellated organisms would therefore include the key to licensing basal body assembly and duplication. The core group of conserved proteins was termed the 'universal module' (UNIMOD) and consists of three proteins; SAS-6/BLD12, SAS-4/CPAP and CEP135/BLD10 (Carvalho-Santos *et al.*, 2010). Many advances in understanding basal body structure and function has been made in recent years but still, relatively little is known about basal body assembly. Lots of structural information came from early electron microscopy (EM) studies on *Chlamydomonas reinhardtii* and *Paramecium spp*. (Cavalier-Smith, 1974; Dippell, 1967; Dippell, 1968; Gould, 1975; Ringo, 1967). Each mature basal body templates the assembly of one pro-basal body to create a pair of basal bodies (Figure 4). This regulated replication ensures that each daughter cell with inherit a single pair of basal bodies. The pro-basal body is formed at an orthogonal angle to the mature basal body (Alvey, 1986; Kuriyama and Borisy, 1981).

Figure 4 illustrates the mammalian centriole duplication cycle with the centrioles numbered 1 to 4. This numbering is important as each centriole has a different age and thereby has different capabilities. In this example, centriole 1 is the oldest (exact age unknown) and centriole 2 is formed as the pro-centriole of centriole 1. The pro-centriole (2) then matures and forms a new pro-centriole (4). The oldest mature centriole (1) also forms a new pro-centriole (3). After G2 there are two pairs of centrioles; pair A has centrioles 1 and 3, pair B has centrioles 2 and 4. However, these centriole pairs are not the same, they are asymmetric. One daughter cell will inherit pair A (1 and 3) and the other daughter cell will inherit pair B (2 and 4). The mature basal body of pair A, centriole 1, is the oldest and it has been reported that the daughter cell which inherits the oldest mature centriole will form a primary cilium before the daughter cell that inherits the newest mature centriole (2) (Anderson and Stearns, 2009). A recent study found that the oldest mature centriole retains a portion of the ciliary membrane when the cilium is reabsorbed (Paridaen *et al.*, 2013). This retention of specialised membrane may help

prime the mature centriole to form a new cilium in the next cell cycle. Therefore the maturity of the centriole/basal body affects cell fate.

In most cells, canonical centriole duplication occurs when the centriole pair is faithfully duplicated once per cell cycle. This is known as centriologenesis (Figure 4). In terminally differentiated cell, such as multi-ciliated epithelia, non-canonical centriole duplication occurs. This facilitates the production of multiple centrioles, which then relocate to the plasma membrane where they dock as basal bodies and extend a cilia axoneme (Cohen *et al.*, 1988; Hagiwara *et al.*, 2004). Non-canonical centriole duplication can occur via two pathways (reviewed by (Nigg and Stearns, 2011). The centriolar pathway, where multiple centrioles are formed using the existing mature centriole as a template, or the acentriolar pathway, which does not use an existing centriole as a template, instead an organelle termed a deuterosome organises the assembly of multiple centrioles (Figure 4) (Klos Dehring *et al.*, 2013).

Studies have shown that the MBB and PBB are physically connected. A fibrous connection between mature and pro-centriole was seen when isolated human centrioles were examined by electron microscopy (EM) (Paintrand *et al.*, 1992) although it was several years until a potential component of the fibres, C-Nap1, was identified (Mayor *et al.*, 2000). This connection is now known as the 'G1-G2 tether' (Nigg, 2007; Sluder, 2005). A search for C-Nap1 interactors identified the coiled-coil protein, rootletin (Bahe *et al.*, 2005; Yang *et al.*, 2006) and the kinase, Nek2 (Fry *et al.*, 1998) whose role in regulating the tether is still ambiguous. This tether connection is lost prior to mitosis (Figure 4, blue). There is a second, more recently discovered connection between a mature centriole and an assembling pro-centriole, termed the stalk (Guichard *et al.*, 2010), or S-M linker (Figure 4, red). Components of the S-M linker are unidentified.

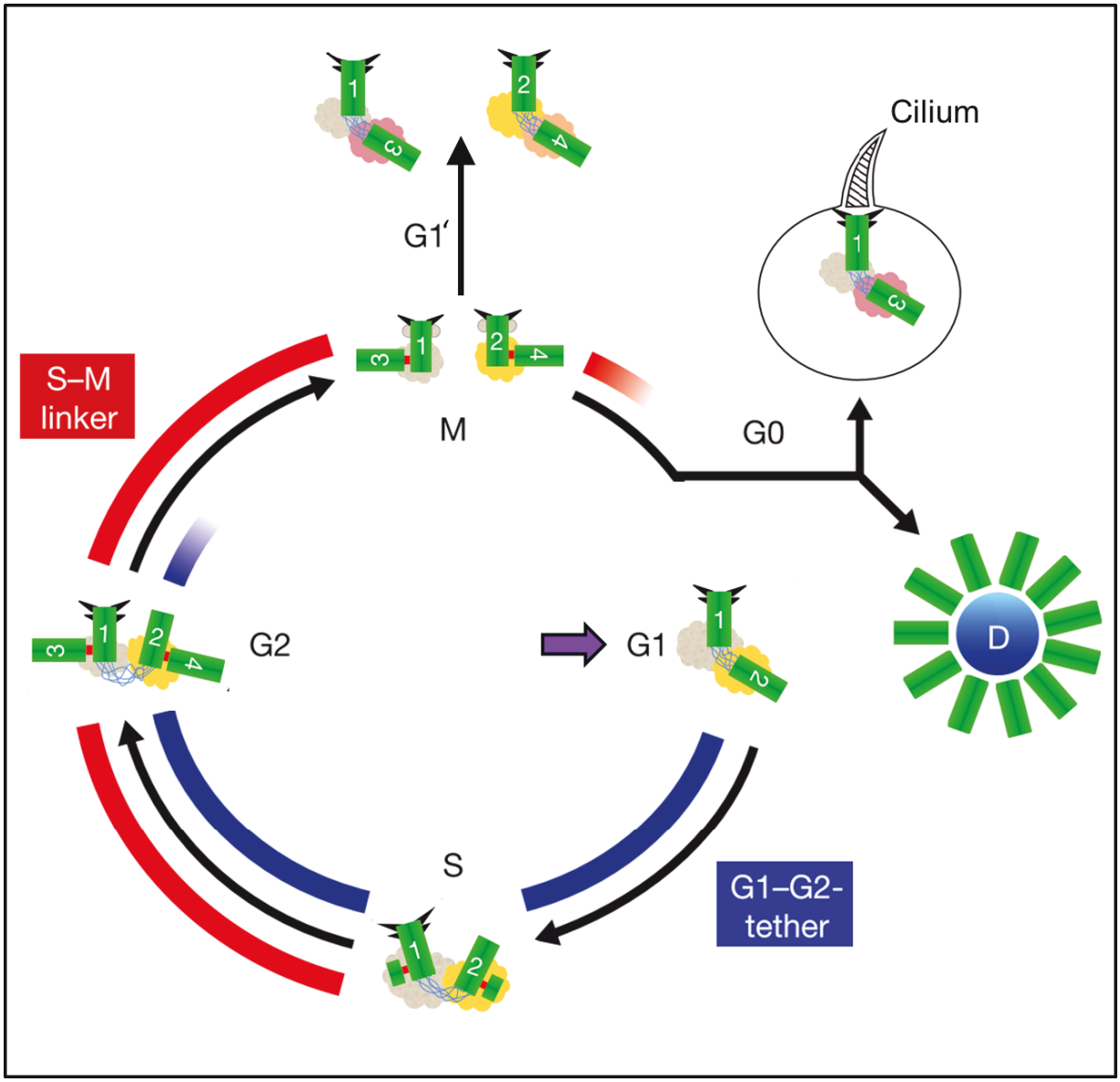


Figure 4: Basal body/ centriole duplication

In G1 (purple arrow) the cell has a single pair of centriole that encompasses a mature centriole (1) and a pro-centriole (2) surrounded by PCM (grey/yellow), which collectively is a centrosome. The mature centriole is connected to the pro-centriole via a fibrous G1- G2 tether (blue). The pro-centriole (2) matures and during S-phase both centrioles form a pro-centriole. The mature centrioles (1 and 2) are still connected via the tether and the pro-centrioles are attached via a S-M linker (red). By G2, each pro-centriole (3 and 4) has elongated. Prior to mitosis (M) the tether between the two mature centrioles (1 and 2) is resolved but the linker is still present in each pair of centrioles (1 to 3 and 2 to 4). The S-M linker is then lost and a new G1-G2 tether is built. Each pair of centrioles is inherited by a daughter cell and the centriole/basal body duplication cycle can begin again (G1') or the cell may enter G0 where the mature centriole/basal body nucleates a single cilium or the cell may terminally differentiate into a multi-ciliated cell via the acentriolar pathway, where multiple centrioles are nucleated from a deuterosome (D). Distal and sub-distal appendages (black) are shown at the distal end of the mature centrioles. Figure modified from (Nigg and Stearns, 2011).

1.4.3 Transition zone

The transition zone is distal to the basal body and proximal to the 9+2 axoneme. It can be identified ultrastructurally as it has doublet microtubules but lacks the central pair microtubules (Figure 5; C) and is therefore denoted as 9+0. There are more subtle structural features including transitional fibres and Y-shaped linkers (Czarnecki and Shah, 2012).

The transition zone has been identified as a key structure with respect to ciliopathy diseases as a number of important proteins have been localised to this area that are connected with ciliopathies including nephronophthisis (NPHP) and Meckel-Gruber syndrome (MKS) (Fliegauf *et al.*, 2006; Williams *et al.*, 2011), ciliopathies are discussed in section '1.5.5 Defects in cilia: Ciliopathies'.

It is expected that the specific structure of the transition zone will vary between organisms but there is currently a lack of ultrastructural information from most model organisms to confirm this. The best characterised transition zone is that of the biflagellate algae, *Chlamydomonas reinhardtii* which has many features in the lumen of the transition zone (O'Toole *et al.*, 2003) including the transitional plate, the stellate array and the acorn filament system. Many of the proteins which make up these sub-structures are unknown but centrin has been identified as a component of the acorn and other filaments in that basal apparatus region (Geimer and Melkonian, 2004; Geimer and Melkonian, 2005).

1.5 Cilia and flagella

Cilia are microtubule based structures which protrude from cells and occur almost ubiquitously in cells of the human body. The canonical structure of a cilium is highly conserved in many eukaryotes and can be formed by both single celled organisms and multi-cellular organisms (Vincensini *et al.*, 2011). Cilia/flagella can occur as a solitary cilium/flagellum (Plotnikova *et al.*, 2009) or many cilia from the same cell, for instance across the pellicle of protozoan ciliates such as *Paramecium tetraurelia* (Beisson *et al.*, 2010) or tracheal epithelial cells (Ostrowski *et al.*, 2002). In multi-ciliated cells the synchronisation of cilia beating is important for functions such as moving the egg cell

along ciliated oviduct epithelium (Shi *et al.*, 2011). Single motile cilia/flagella are important for sperm and protozoan motility (Hill, 2010; Inaba, 2007). Cilia can also occur in an immotile form which is often referred to as a primary cilium, which can be formed by almost all cell types in the human body (for review see; (Mitchell, 2004)).

1.5.1 Axoneme ultrastructure

As mentioned above, cilia can be divided into subcategories; motile and immotile. Immotile cilia lack the central pair of microtubules and other motility associated structures such as the radial spokes and dynein arms and are referred to as 9+0 (Figure 6; A), these are often called immotile or primary cilia. It is widely accepted that motile cilia contain 9 outer doublet microtubules and a central pair of microtubules. This configuration is known as 9+2 (Figure 6; B), and is conserved across motile flagella (Mitchell, 2004; Pigino *et al.*, 2012). In 9+2 and 9+0 axonemes, the outer doublet microtubules, A-tubule and B-tubule, extend from the basal body. The C tubule does not extend (Anderson, 1972).

1.5.2 Motile cilia

Motile cilia/flagella have dynein arms which connect to the A tubule of the outer doublets to the B tubule of the neighbouring doublet which gives rise to movement. Dynein arms on opposing sets of doublets are alternatively active (Brokaw, 2009; Summers and Gibbons, 1971) allowing bi-directional movement of the axoneme.

Motile cilia are important across tissue types because of their function in moving mucus, eggs and sperm (Morita and Shingyoji, 2004; Shi *et al.*, 2011). The beat of most motile cilia originates near the base of the axoneme and terminates at the tip (Brokaw, 1991; Gadelha *et al.*, 2007; Lindemann and Lesich, 2010). During embryonic development, it is vital that the motile cilia help to direct cerebrospinal fluid in and out of the brain ventricles otherwise fluid accumulation can lead to hydrocephalus (Figure 8; 3), (Lechtreck *et al.*, 2008; Olbrich *et al.*, 2012).

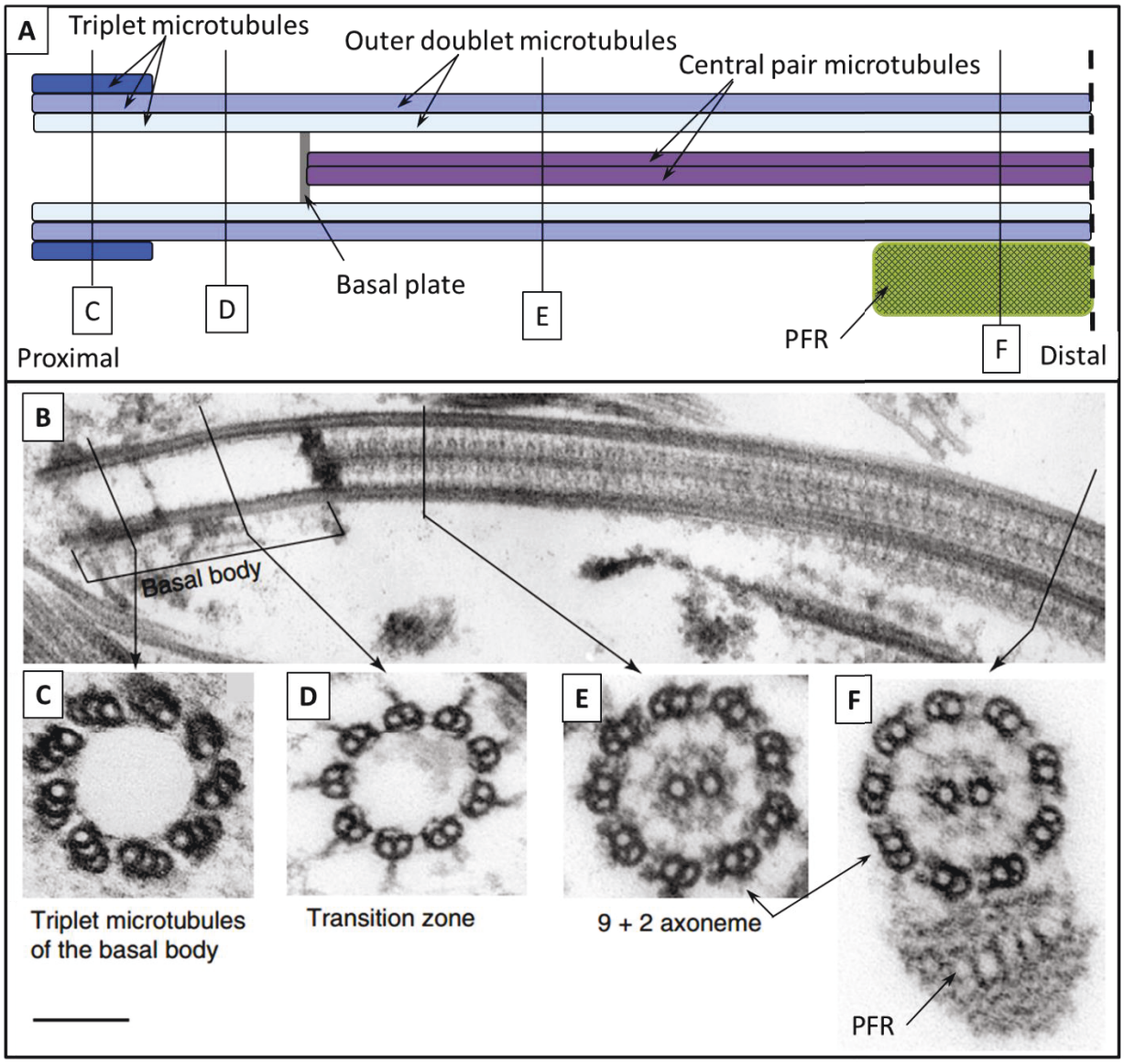


Figure 5: Flagellum structure

(A) Cartoon of flagellum structure showing the arrangement of the triplet microtubules of the basal body (also, C), the outer doublet microtubules of the axoneme and the central pair microtubules (purple). (B) micrograph overview of a flagellum showing the spatial relationship between the basal body (C), transition zone (D), axoneme (E) and PFR (F). The paraflagellar rod (PFR, green) is an extra structure unique to *Euglena* and trypanosomatids and will be discussed later. (C-E) viewed in a proximal to distal orientation. (Modified from (Vaughan *et al.*, 2006)) Scale bar = 100nm

1.5.3 Immotile cilia

Immotile cilia are also referred to as primary or sensory cilia. As previously mentioned, immotile cilia lack the central pair microtubules and associated structures (Figure 6; A). Immotile cilia/flagella are important for many processes and have a crucial role in embryo formation when the cilia of the node cells influence the left/right asymmetry of the embryo patterning (Afzelius, 1998). When the nodal cilia malfunction it can result in the left right asymmetry of the body patterning being inverted, a condition known as *situs inversus* (Figure 8; 1), (reviewed in (Tobin and Beales, 2009). Immotile cilia have important sensory roles and indeed it is being discovered motile cilia/flagella also have sensory roles (Gluenz *et al.*, 2010). When cells have a single cilium it is referred to as the primary cilium and its mechano-sensory roles include detecting liquid flow in the kidneys and detecting stress loading on bone (Anderson *et al.*, 2008) which helps to maintain a healthy bone density and prevent osteoporosis. There are specialised examples of immotile cilia such as the connecting cilium in photoreceptor cells of the eye or the kinocilia which connect the surface of hair cells in the ear to transmit vibrations. Leber congenital amaurosis and other optical ciliopathies (Adams *et al.*, 2007) are caused by disruption to the connecting cilium of the rod cells in the eye. A ciliopathy known as Usher's syndrome affects the cilia in both photoreceptors and hair cells of the ear, causing sufferers to become blind and deaf as the disease progresses (Kremer *et al.*, 2006; van Wijk *et al.*, 2009).

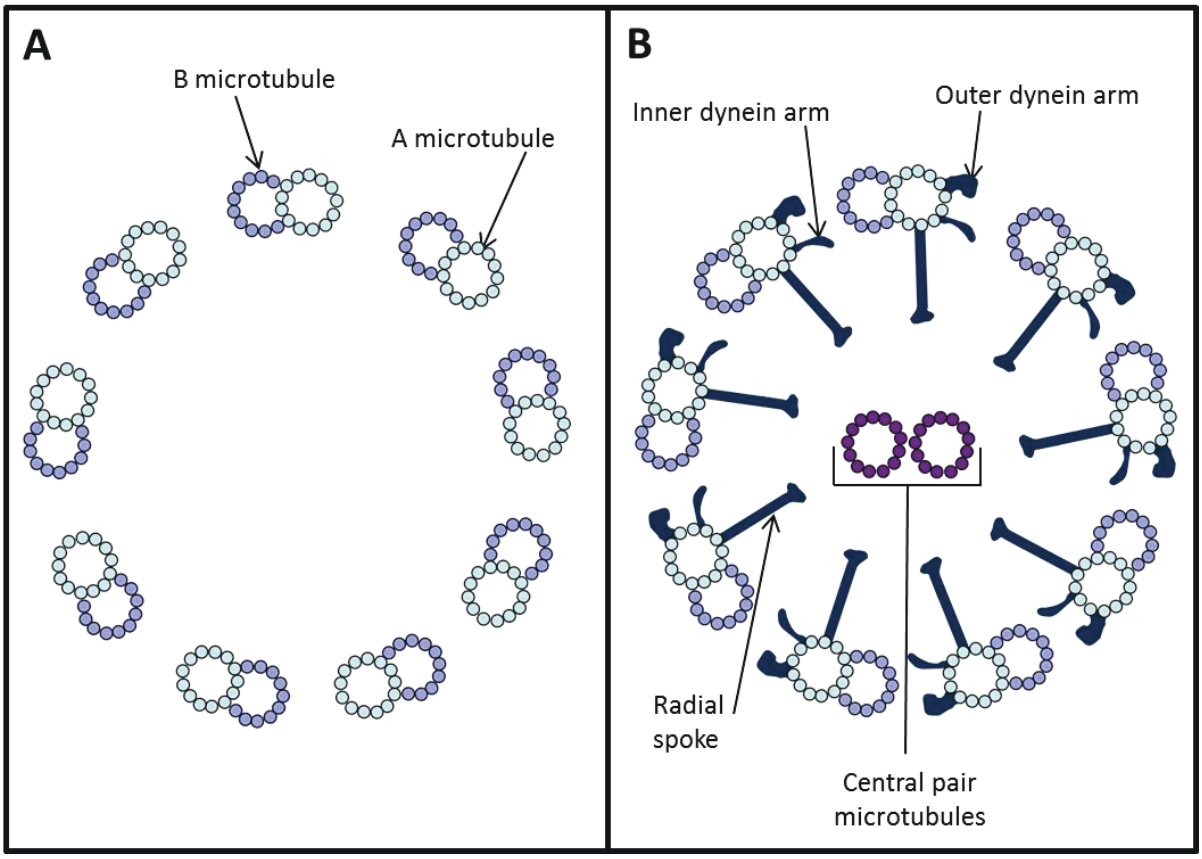


Figure 6: Comparison of motile and immotile cilia

(A) Cross section of an immotile cilium showing possession of the 9 outer doublet microtubules but lacking a central pair and associated apparatus. In comparison the cross section of a motile cilium (B) which shows the same 9 outer doublet microtubules but in motile cilia the doublets have outer and inner dynein arms to generate and propagate motile forces along the length of the cilium (Smith, 2002; Smith and Lefebvre, 1997b) motile cilia also contain a central pair of microtubules (purple) which give the microtubule configuration of 9+2. Cilia viewed in a proximal to distal orientation.

1.5.4 Intraflagellar transport

Components required to build and maintain the axoneme are moved up the cilium/flagellum via a process known as intraflagellar transport (IFT). Although originally described in *C. reinhardtii* (Kozminski *et al.*, 1993) it is now recognised as being a conserved process for cilia/flagella formation and maintenance (Absalon *et al.*, 2008; Baldari and Rosenbaum, 2010; Ishikawa and Marshall, 2011). IFT allows constant growth and turnover of the microtubules at the distal tip of the cilium/flagellum (Marshall and Rosenbaum, 2001; Mitchison and Kirschner, 1984). The IFT particles group into IFT trains which use the outer doublet microtubules of the axoneme as a track (Figure 7). Dynein 2 transports cargo in a distal to proximal direction (retrograde transport) and Kinesin-2 transports cargo in a proximal to distal direction (anterograde transport) (Scholey, 2008). Anterograde IFT occurs at 2 µm/second and retrograde transport occurs at 3.5 µm/second (Kozminski *et al.*, 1993) implying that the two directions of movement must be caused by two motor systems due to the difference in transport rates. Dysregulation of IFT can also cause ciliopathies including Leber congenital amaurosis (Figure 8; B), which causes patients to lose their sight (Boldt *et al.*, 2011).

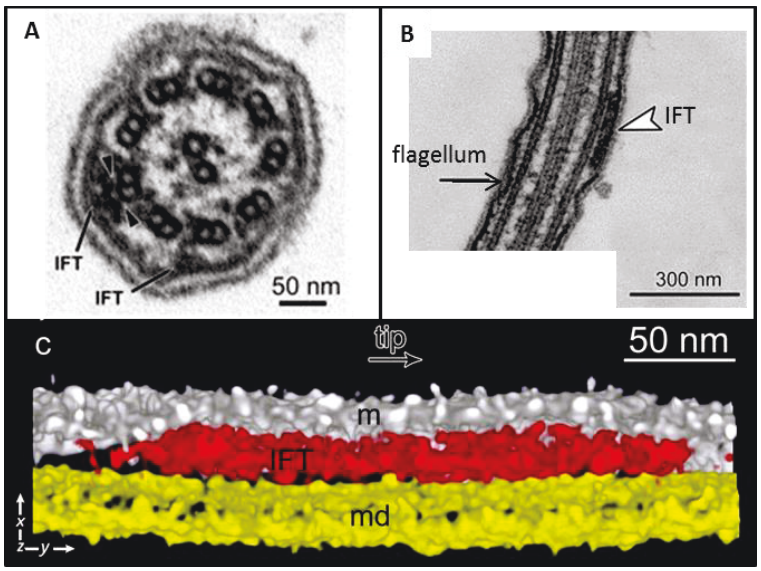


Figure 7: Intraflagellar transport

(A) TEM micrograph of a cross section through a *C. reinhardtii* flagellum. The 9+2 configuration of axoneme microtubules is evident and two IFT trains are visible. (B) TEM micrograph of a longitudinal view of the flagellum to the IFT train (white arrowhead). The length of an IFT train can vary. (C) 3D reconstruction of an IFT train (red) travelling along a microtubule doublet (md, yellow) modelled from a tomogram. Flagellum membrane is shown in white. Adapted from (Pigino *et al.*, 2009).

1.5.5 Defects in cilia: Ciliopathies

Defects in cilia cause a class of diseases known as the ciliopathies. Ciliopathies result from cilia dysfunction and can be caused by a variety of mutations in genes encoding proteins that are involved in the cilia assembly or maintenance (for review see (Hildebrandt *et al.*, 2011; Waters and Beales, 2011)) and include diseases such as Jeune syndrome, primary ciliary dyskinesia, Meckel syndrome and Alström syndrome (Badano *et al.*, 2006). Ciliopathies have many overlapping clinical features despite the variety of genetic mutations due to an effect termed mutational load (Lee and Weatherbee, 2010; Zaghloul and Katsanis, 2010). Shared phenotypes are many and varied, examples include polydactyly (multiple digits) (Figure 8; 9), kidney cysts (Figure 8; 7), blindness (Figure 8; 4), obesity (Figure 8; 8) and skeletal defects such as short ribs (Figure 8; 10). Currently there are 15 recognised ciliopathies including Bardet-Biedl syndrome, Meckel-Gruber syndrome and Senior-Løken syndrome although it is worth noting that there are approximately 80 other diseases that are predicted to also be ciliopathies (van Reeuwijk *et al.*, 2011). Although the first ciliopathy, Bardet-Biedl syndrome was first officially diagnosed in 1866 it has taken over a century for this area of cell biology to come to the forefront of research interest. This may be partly due to the complicated nature of ciliopathies as a disease group (Baker and Beales, 2009). The specific ciliopathy and severity of the disease is due to how many and which genes are mutated. The complexity of the gene interactions is summarised by the gene network devised by van Reeuwijk and colleagues (2011). In some cases a mutation in a single gene is enough to cause a ciliopathy, in other patient cases mutations in a combination of several genes is needed to result in a clinical syndrome – a variation in the specific genes with a mutation may significantly alter the severity of the disease even within the same syndrome.

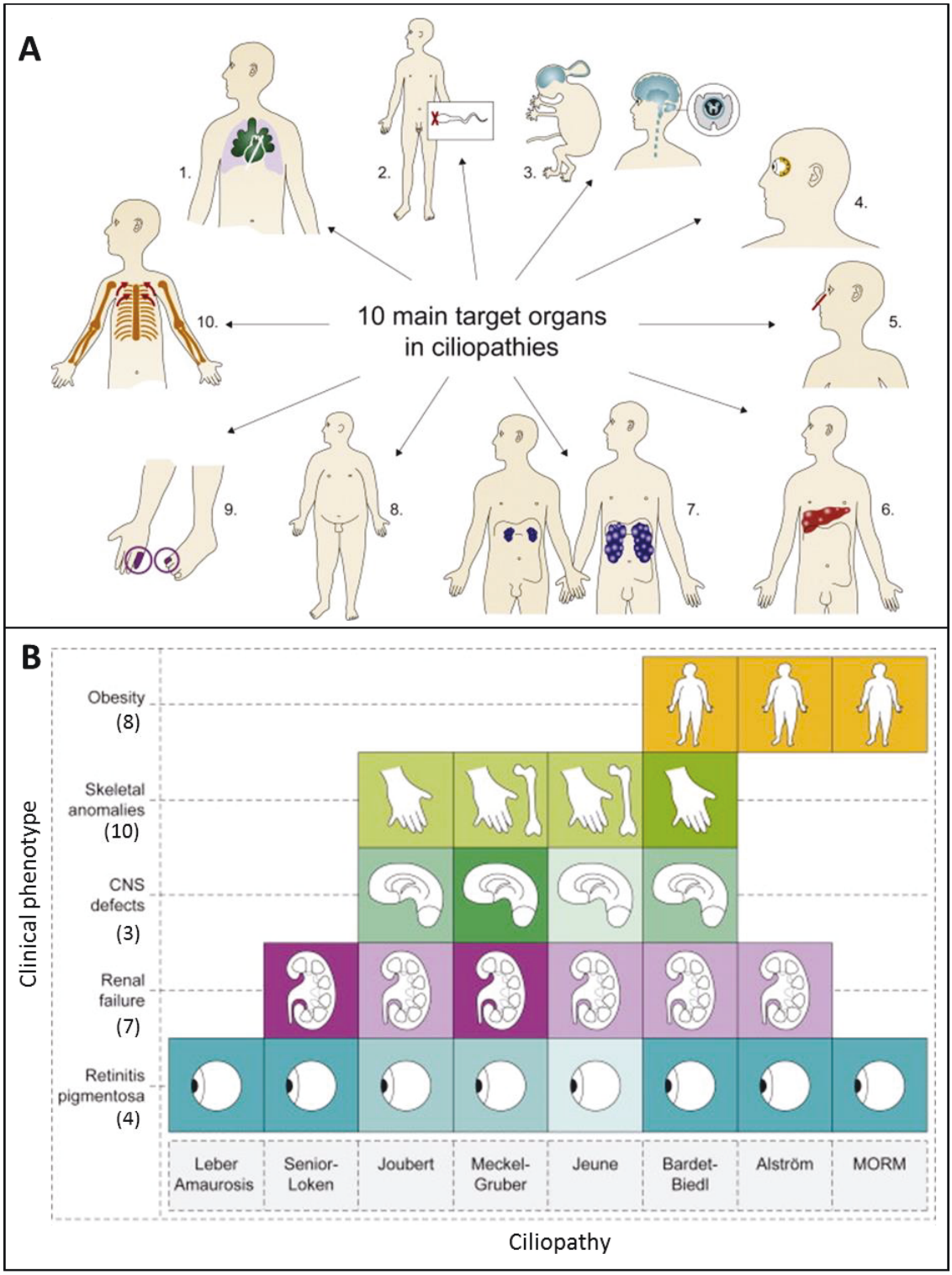


Figure 8: Ciliopathy phenotypes

(A) A pictorial representation of selected but no exhaustive clinical phenotypes associated with ciliopathy diseases. (1) Situs inversus, (2) Infertility, (3) Hydrocephalus, (4) Retinitis pigmentosa, (5) Anosmia, (6) Cystic liver, (7) Cystic kidneys, (8) Obesity, (9) Polydactyly and (10) Skeletal deformities e.g. short ribs. (B) A chart to demonstrate how these clinical phenotypes can occur independently or in combination with one another. Leber congenital amaurosis sufferers are blind whilst Bardet-Biedl syndrome patients have multiple ciliopathic traits. Figure modified from (Mockel *et al.*, 2011).

1.5.6 Ciliopathy disease: Model organisms

Model organisms such as C*. reinhardtii* and *T. brucei* are used to study the effect of protein depletion on the structure and function of cilia/flagella (Ostrowski *et al.*, 2011; Vincensini *et al.*, 2011) to elucidate the cause of diseases such as the ciliopathies.

There are a number of factors that make *T. brucei* an excellent and well used model organism to study the biology of cilia and flagella. The cytoskeleton of *T. brucei* remains intact during cell division and the parasite contains many single copy organelles including the flagellum and basal body. Duplication and segregation of these structures is temporally and spatially co-ordinated in order to produce two identical daughter cells and these events can be easily followed using a number of established antibody and morphology markers (Birkett *et al.*, 1985; Kilmartin *et al.*, 1982; Kohl *et al.*, 1999; Woods *et al.*, 1989b). The new flagellum grows alongside the old flagellum in a precisely defined position that can be easily assessed by light and electron microscopy (Gull *et al.*, 1990; Sherwin and Gull, 1989) and so defects in assembly of a new flagellum can be compared to the existing flagellum. In addition, the genome of *T. brucei* has been sequenced (Berriman *et al.*, 2005) and a proteomic study found that 123 *T. brucei* cytoskeletal flagellum proteins are conserved in *H. sapiens* and other eukaryotes (Broadhead *et al.*, 2006), which makes experimental findings in *T. brucei* applicable to human disease research.

As a well-established model, *T. brucei* has been used to investigate the function of proteins linked to ciliopathies including polycystic kidney disease (PKD) (Morgan *et al.*, 2005), Primary ciliary dyskinesia (PCD) (Baron *et al.*, 2007a; Springer *et al.*, 2011), retinitis pigmentosa (Stephan *et al.*, 2007) and hydrocephalus (Dawe *et al.*, 2007). Depletion of the protein, TbLRTP by RNA interference (RNAi) resulted in an increase in basal body duplication and overexpression of this protein resulted in a decrease in basal body duplication (Morgan *et al.*, 2005). The human homolog of TbLRTP was later found to be a gene involved in causing PKD when mutated (van Rooijen *et al.*, 2008) and the mutant phenotype has been reproduced in other model organisms (Serluca *et al.*, 2009).

Ultrastructural analysis of tissue from patients with PCD found that dynein arms were missing in the axoneme (Papon *et al.*, 2010). Studies in *T. brucei* using RNAi to reduce

expression of the outer and inner arm dynein proteins (Baron *et al.*, 2007a; Springer *et al.*, 2011) produces phenotypes where the cells were immotile (Gunay-Aygun, 2009) or in which motility is severely disrupted (Baron *et al.*, 2007a). Abnormalities seen in axoneme ultrastructure in the trypanosome mutants mirror the dynein defects seen in the cilia of PCD patients (Papon *et al.*, 2010).

Retinitis pigmentosa is a disease of the eye which can occur as an independent ciliopathy or as a clinical symptom of a systemic ciliopathy (Adams *et al.*, 2007). *T. brucei* was used as a model to investigate the function of the homolog of the human gene XRP2 (Stephan *et al.*, 2007). Immunofluorescence and immunogold labelling showed that the protein was localised to the distal end of the basal body. An inducible RNAi cell line was used to deplete the protein and thereby study the function of this gene. Induced cells produced 9+2 axonemes that were shorter than wild type cells and with defects in the arrangement of microtubules. It was proposed that the effect on axonemal assembly may be due to disruption of protein importation into the flagellum (Stephan *et al.*, 2007).

1.6 *Trypanosoma brucei*

1.6.1 The *T. brucei* life cycle

Trypanosoma brucei subspecies are protozoan parasites that causes African sleeping sickness in humans and nagana in cattle in sub-saharan Africa (for review see (Steverding, 2008). Sleeping sickness or human African trypanosomiasis (HAT) is caused by two subspecies of *Trypanosoma brucei*, *Trypanosoma brucei rhodesiense* and *Trypanosoma brucei gambiense,* which causes 95% of reported HAT cases (WHO, 2012). The third subspecies, *Trypanosoma brucei brucei,* does not infect humans but can infect cattle. *Trypanosoma brucei* subspecies are transmitted between mammalian hosts via the Tsetse fly insect vector. Trypanosomes are such a successful extracellular parasite because of their variant surface glycoprotein (VSG) coat which they can switch almost unlimitedly at a rapid rate (Taylor and Rudenko, 2006), thereby evading the hosts immune system. This causes problems for attempts to create an effective vaccine (La Greca and Magez, 2011).

The procyclic form of the trypanosome does not have a VSG coat as it resides within the Tsetse fly; instead it has a procyclin coat (Roditi *et al.*, 1998).

T. brucei has a complex lifecycle which involves two hosts and multiple cellular forms (Figure 9). *T. brucei* is transmitted between the mammalian host (red) and the Tsetse fly insect vector (Figure 9; green). There are two proliferative forms of the parasite that are used as experimental model organisms in the laboratory. The long slender bloodstream form (BSF) and the procyclic form (PCF) are both used for research due to the fact that they are responsive to *in vitro* growth conditions and therefore is it relatively easy to scale up growth cultures for experiments. The BSF cell is the cell type which causes parasitemia in the mammalian host and has a protective VSG coat to survive in the bloodstream (for review see (Taylor and Rudenko, 2006)). The bloodstream form cells can also cross the blood brain barrier in the later stages of mammalian disease. The next life cycle form is the short stumpy form cell. The stumpy form cell can be found in the mammalian bloodstream and it is poised for uptake by the insect vector when the Tsetse fly bites the infected mammal for a blood meal. The stumpy form cell is ingested in the blood meal and differentiates into a PCF which does not need a VSG coat and instead the cell surface has a layer of procyclin proteins. The PCF cells migrate from the Tsetse fly midgut to the proventriculus where the PCF cells differentiate into the epimastigote form, which is capable of attaching to the salivary gland epithelium via the flagellum (Tetley and Vickerman, 1985). The epimastigote forms can rapidly divide to facilitate an efficient volume of *T. brucei* cells which then differentiate into metacyclic form cells ready to be regurgitated by the Tsetse fly at the next blood meal and begin an infection within a new mammalian host. The flagellum remains assembled throughout the lifecycle of *T. brucei* and plays several important roles including; establishing an infection in the mammalian host, traversing the Tsetse fly midgut to reach the salivary glands and anchoring the epimastigote form to the epithelial cells.

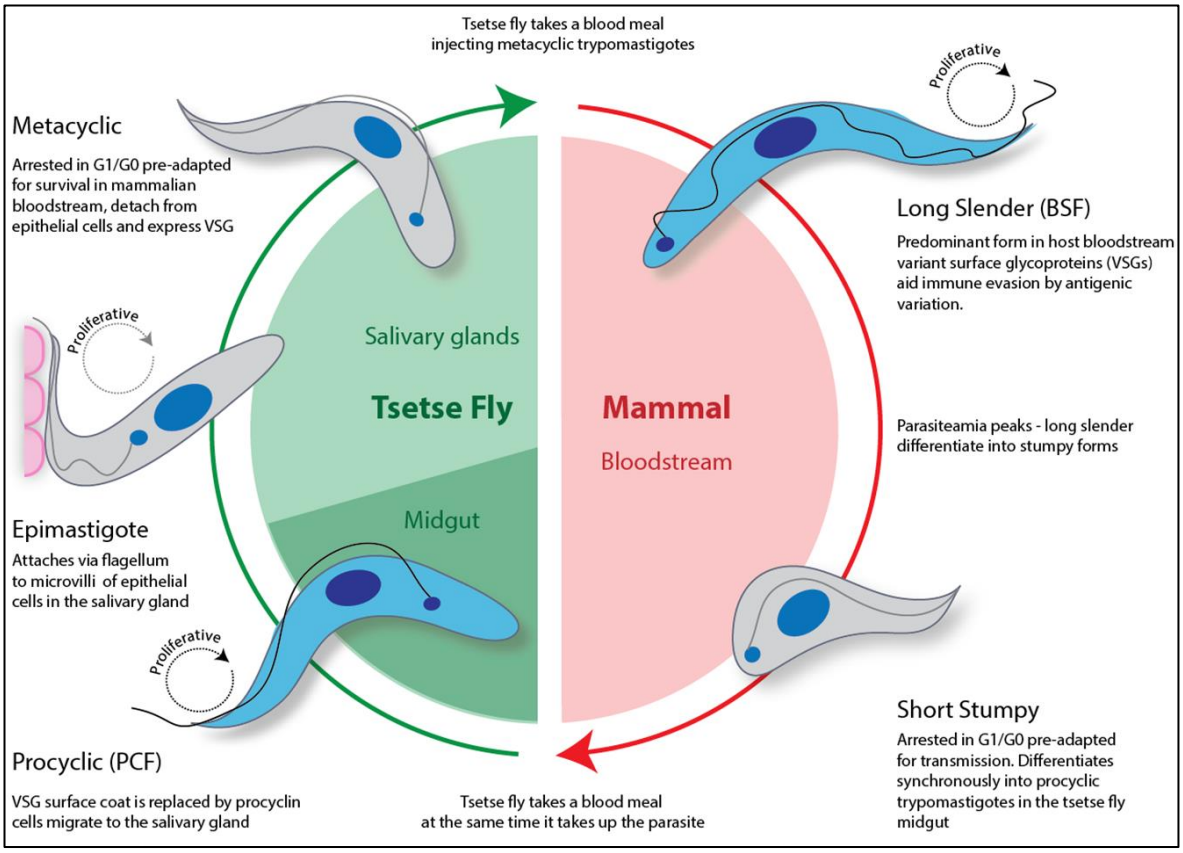


Figure 9: Life cycle of *Trypanosoma brucei*

The lifecycle of *Trypanosoma brucei* subspecies, which cycles between a mammalian host (red section) and the insect vector, the Tsetse fly (green section). The diagram shows 5 distinct lifecycle forms although there are other intermediate cell types and variations not included here for simplicity. Diagram from (Towers, 2010)

1.6.2 Microtubules in *T. brucei*

The main cytoskeleton of *T. brucei* is made up of microtubules (Seebeck *et al.*, 1988). Microtubules within *T. brucei*, like other eukaryotes are predominantly composed of α and β tubulin heterodimers. Although tektin protofilament proteins are not conserved within trypanosomatids, genes encoding Rib72 and Rib43 were found in the genome (Berriman *et al.*, 2005).

Microtubules within *T. brucei* belong to two main classes; the axonemal microtubules (AMt) or the subpellicular microtubules (SMt). There are a small number of additional microtubules in the basal body, microtubule quartet, neck microtubule and mitotic spindle. The AMt and SMt contain the same form of α tubulin (Schneider *et al.*, 1987) but are made up of a different β tubulin, which has been subject to post translational modification (Gallo and Anderton, 1983; Gallo *et al.*, 1988). This modification of β tubulin seems to be a conserved feature of motile flagella (Nielsen *et al.*, 2001).

In addition to the AMt and SMt, trypanosomatids have a microtubule quartet (MtQ) that extends from the basal bodies to the anterior end of the cell body, inserted between SMt on the dorsal side of the cell (Figure 11; B and C). The neck microtubule is a singular microtubule which extends from the flagellar pocket, past the pocket collar and then a short distance along the cell body in both the PCF and BSF cell (Gadelha *et al.*, 2009; Lacomble *et al.*, 2009). The polarity of the neck microtubule is unknown.

The SMt are so named as they are located directly underneath the cell membrane, the pellicle. This is a feature common to many protozoans (Morrissette and Sibley, 2002). The SMt in gives the cell body its shape (Figure 10) as the number of microtubules and spacing between is precisely regulated (Sherwin and Gull, 1989). The SMt remains intact during cell division (Sherwin and Gull, 1989) and provides a framework for organelle positioning (Robinson *et al.*, 1995).

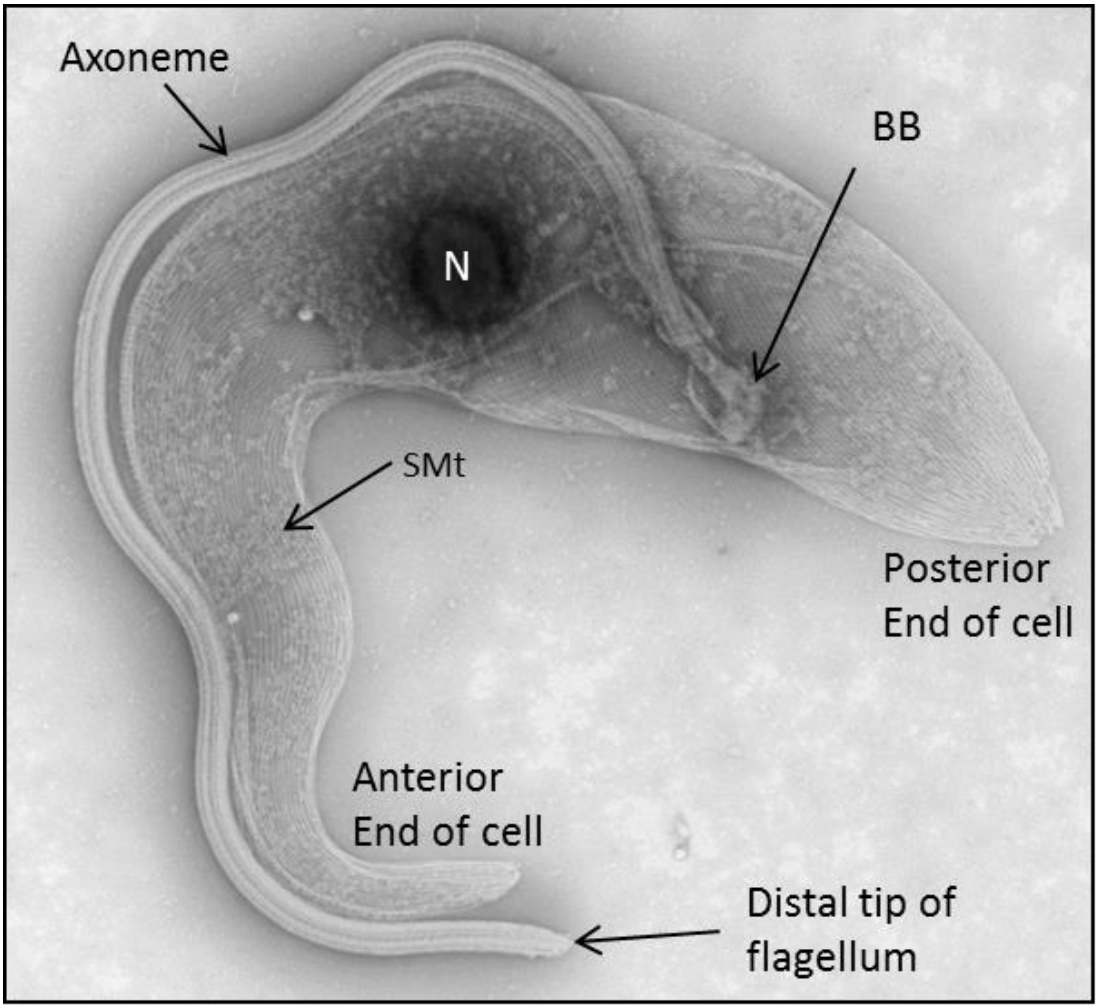


Figure 10: *Trypanosoma brucei* cytoskeleton

Whole mount cytoskeleton of *Trypanosoma brucei* showing the subpellicular corset microtubules (SMt), the basal body (BB) and axoneme microtubules. The black mass in the middle of the cell is the nucleus (N). The kinetoplast would be located in a proximal location to the BB in whole cells but this cytoskeleton has been treated with detergent and the membrane of the kinetoplast has been extracted. Micrograph courtesy of Dr K Towers.

The outer doublet microtubules of the axoneme are nucleated from a known MTOC, the basal body, therefore the positive end of the AMt is at the distal tip of the axoneme (Figure 11). The central pair microtubules (purple) are nucleated from the basal plate (Figure 5; A). In *T. brucei* the negative end of the SMt are at the anterior end of the cell body and therefore the positive end is at the posterior end of the cell body (Figure 11; C). Therefore the SMt have the opposite polarity to the AMt. The MtQ has the same polarity as the AMt and is therefore opposite to the SMt as well.

The MTOC for the SMt is still structurally uncharacterised but is positioned at the anterior end of the cell body as localised by immunolabelling cells with an anti-γ tubulin antibody (Scott *et al.*, 1997), which also labels the basal body and the mitotic spindle. Knockdown of γ tubulin in *T. brucei* is ultimately lethal as the cells are immotile but EM analysis of flagella revealed that the central pair microtubules were absent but surprisingly the outer doublet microtubules appeared normal (McKean *et al.*, 2003). This is evidence that the triplets of the mature basal body can act as a template for outer doublet extension in the absence of γ tubulin but assembly of the central pair microtubules requires γ tubulin.

Figure 11, figure legend: (A) A cartoon overview of a trypanomastigote cell with the basal body (BB) nucleating the proximal end of the flagellum. The distal tip of the flagellum is at the anterior end of the cell body. (B) A partial cross section through the dorsal side of the cell body at position indicated by dashed line (A) in a proximal to distal orientation as indicated by red arrow. The flagellum (blue) is positioned on top of the cell body (red). The flagellum contains the axoneme microtubules (AMt) and the paraflagellar rod (PFR, green). Underneath the cell body membrane are the subpellicular corset microtubules (SMt, red). A specialised set of four microtubules, the microtubule quartet (MtQ, orange) are also positioned in parallel to the SMt in the region of the flagellum. There is a non-microtubule structure, the FAZ filament (FAZ_f, yellow) that occupies a space running the length of the flagellum between the MtQ and SMt. (C) The polarity of the SMt is indicated as positive (+) at the posterior end of the cell and negative (-) at the anterior end of the cell. The polarity of the MtQ is the opposite to the SMt with the negative end of the MtQ located at the BB and the positive end of the MtQ is at the anterior end of the cell body. (D) The polarity of the axonemal microtubules (AMt) is the same as that of the MtQ with the positive end of the AMts at the distal tip of the flagellum and the negative end at their MTOC, the basal body.

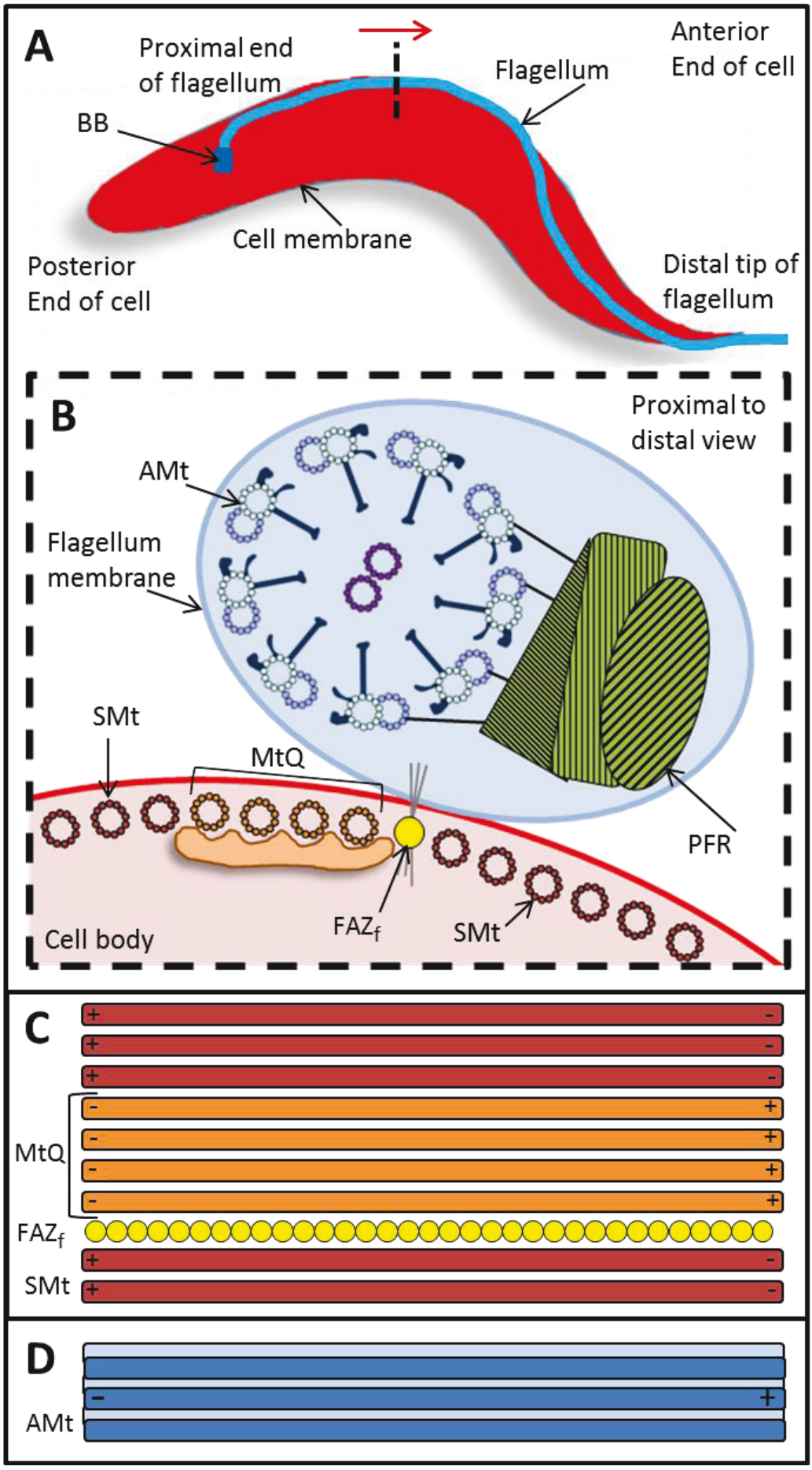


Figure 11: The polarity of the *Trypanosoma brucei* microtubule cytoskeleton

1.6.3 Actin filaments

When the genome of *Trypanosoma brucei* was sequenced and assembled it came to light that genes encoding for actin proteins were present that were previously assumed not to be conserved within trypanosomatids. Whilst *T. brucei* does have actin, functional analysis concluded it does not appear to be essential as knockdown is only lethal in bloodstream form cells (Garcia-Salcedo et al., 2004). This is thought to be due to an interruption to endocytosis. It is not lethal when knocked down by RNAi in procyclic form cells, although defects in Golgi apparatus and vesicle formation were reported (Garcia-Salcedo et al., 2004).

1.6.4 Intermediate filaments

Based on the sequencing of genomes no intermediate filaments are thought to be conserved in in *T. brucei* (Berriman *et al.*, 2005) or the related kinetoplastids, *T. cruzi* and *Leishmania major* (El-Sayed *et al.*, 2005; Ivens *et al.*, 2005). Genes encoding for lamin, keratin, vimentin, septin, desmoplakin, plectin1, collagen and integrins were not found (Berriman *et al.*, 2005). Although a candidate may have been found in *Leishmania* spp. because an anti-vimentin antibody labels the poles of the cell body (Kratzerova *et al.*, 2001). Vimentin is known to be associated with centrosomes, the MTOC, in mammalian cells (Trevor *et al.*, 1995). If there truly is a vimentin-like protein conserved in *Leishmania* it is likely to be a divergent form and therefore was not identified by traditional bioinformatics approaches. It is important to note that unconventional orthologs are being discovered in trypanosomatids such as kinetochore proteins (Akiyoshi and Gull, 2014) and lamins (DuBois *et al.*, 2012), which were previously thought to be not conserved. There are many non-microtubule structures in trypanosomatids that have been extensively structurally characterised such as the PFR (Fuge, 1969; Hughes *et al.*, 2012), the FC (Briggs *et al.*, 2004b; Moreira-Leite *et al.*, 2001), the FAZ (Vickerman, 1969), the TAC (Bonhivers *et al.*, 2008a; Ogbadoyi *et al.*, 2003) and bilobe (Esson *et al.*, 2012; Morriswood *et al.*, 2009) but not much is known about their molecular composition. Filamentous structures have also been recorded in other closely related kinetoplastid species including *Phytomonas* and *Crithidia* spp. (Page and Lagnado, 1998; Page and Lagnado, 2000).

1.6.5 The *T. brucei* cell cycle

Trypanosoma brucei are single celled organisms with a single flagellum that belong to the order kinetoplastida. The trypomastigote cell contains many single copy organelles including a nucleus, flagellum, mitochondrion and kinetoplast. The organelles are duplicated once per cell cycle in a regulated order. The timings of these cell cycle events has been well characterised in procyclic and bloodstream form cells (Gull *et al.*, 1990; Matthews *et al.*, 1995; Ploubidou *et al.*, 1999; Sherwin and Gull, 1989; Woodward and Gull, 1990). This well characterised cell cycle progression means that during phenotypic analysis of mutant cell lines it is relatively easy to detect deviation from the normal cell cycle. (Specific organelles are covered in the following sections 1.6.5 – 1.6.11).

Light microscopy is commonly used to assess *T. brucei* cell cycle progression using a selection of antibodies to immunolabel structures within the cell (Bastin et al., 1996; Kohl et al., 1999; Woods et al., 1989b; Woodward et al., 1995) in addition to staining the DNA-containing organelles with 4',6-diamidino-2-phenylindole (DAPI). This catalogue of community made antibodies has helped to refine cell cycle timings and events (Figure 12).

Figure 12 displays the cell division cycle for procyclic form *T. brucei* trypanomastigotes which start the cell cycle with a single copy kinetoplast and a single copy nucleus (Figure 12; A, 1K1N). The next cell cycle stage in this diagram (Figure 12; B) shows a kinetoplast prior to mitochondrial DNA division and the cell has a single nucleus (Figure 12; B, DivK1N). The basal bodies have duplicated and segregated with the new basal body nucleating the new flagellum (Figure 12; NF, green). The new flagellum is positioned more posterior than the old flagellum (Figure 12; OF, blue). The NF is connected to the OF via a structure called the flagella connector (Figure 12; FC, coral) (see section 1.5.7). The kinetoplast divides to form two kinetoplasts and therefore the cell is denoted as 2K1N (Figure 12; C). The NF continues to extend and has formed a new flagellum attachment zone (FAZ) (see section 1.5.5). Nuclear division then occurs (Figure 12; D, 2KDivN) to form a cell with two kinetoplasts and two nuclei (Figure 12; E, 2K2N). This means the cell is ready to undergo cytokinesis and the cleavage furrow is visible (Figure 12; F), the FC is resolved. The cell undergoes cytokinesis and the final point of abscission at the posterior end of the connected daughter cells (Figure 12; G) is resolved to produced two daughter

cells primed to enter the cell cycle again (Figure 12; A', 1K1N). One daughter cell inherits the old flagellum (blue) with the associated basal body pair and the other daughter cell inherits the new flagellum (green) with the associated basal body pair (Wheeler *et al.*, 2013b).

Figure 12: Figure legend.

A diagram to represent the duplication of the DNA-containing organelles in the *T. brucei* procyclic form cell cycle. (A) *T.brucei* cell which has one flagellum (blue), one kinetoplast (1K) and one nucleus (1N) and is therefore denoted 1K1N. (B) The basal bodies (purple) have duplicated and the newly mature basal body is assembling a new flagellum (green), which is connected to the old flagellum (blue) by the flagella connector (coral). The kinetoplast is dividing and the cell still has a single nucleus (DivK1N). (C) The two kinetoplasts have segregated and the cell has a one nucleus (2K1N). (D) In addition to two kinetoplasts, the nucleus undergoes mitosis (karyokinesis) (2KDivN). (E) The cell has two kinetoplasts and two nuclei (2K2N). (F) The 2K2N cell undergoes cytokinesis with the cleavage furrow originating at the anterior end of the cell and travelling towards the posterior end. (G) The final cell cycle stage prior to abscission, the cells are still connected at the posterior ends. (A') Cytokinesis is complete to produce two 1K1N daughter cells. Modified from (Langousis and Hill, 2014).

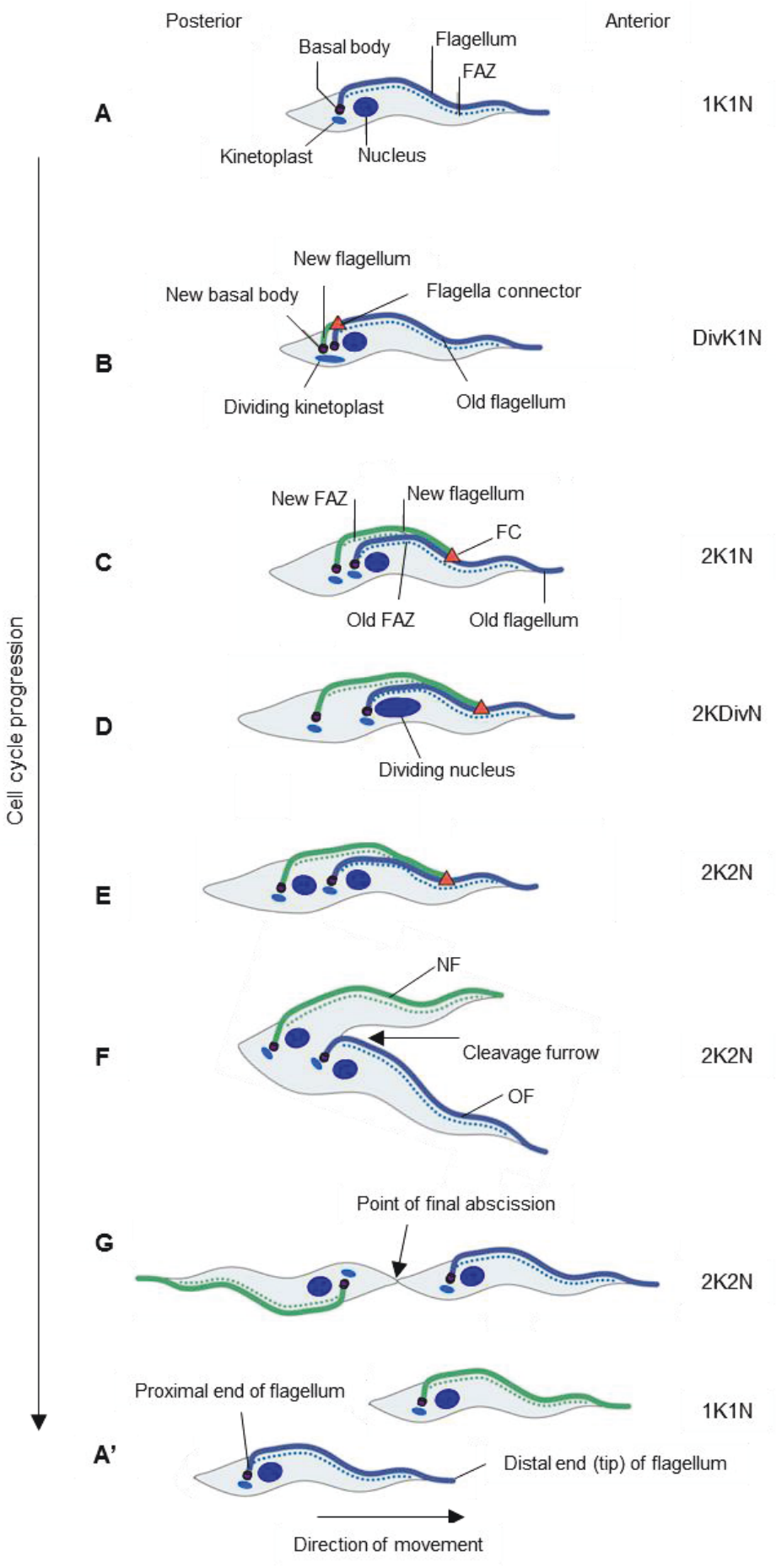


Figure 12: Cell cycle of procyclic form *Trypanosoma brucei*

1.6.6 Kinetoplast and TAC

The kinetoplast is a specialist organelle that contains the mitochondrial DNA and gives the kinetoplastids their name (Vickerman *et al.*, 1988). The kinetoplast contains the mitochondrial genome, referred to as kinetoplast DNA (kDNA), that is separate from the nuclear genome (Figure 13; A). The kinetoplast is connected to the basal body (BB) via a filamentous set of linkages, the tripartite attachment complex (TAC) (Ogbadoyi *et al.*, 2003) (Figure 13; B and C), which is thought to aid kinetoplast segregation during basal body separation during the cell cycle (Gluenz *et al.*, 2011; Liu *et al.*, 2005; Robinson and Gull, 1991). Few components of the TAC have been identified to date (Schnarwiler *et al.*, 2014; Zhao *et al.*, 2008).

The mitochondrial DNA forms minicircles and maxicircles which are compacted into the disk shaped organelle, situated below the basal body (Figure 13; C). Duplication and separation of the kinetoplast DNA occurs separately from nuclear mitosis (Ploubidou *et al.*, 1999). Kinetoplast duplication and segregation is well characterised (Gluenz *et al.*, 2011; Robinson and Gull, 1991) and the copy number of kinetoplasts within a cell is used to assess cell cycle progression (Woodward and Gull, 1990).

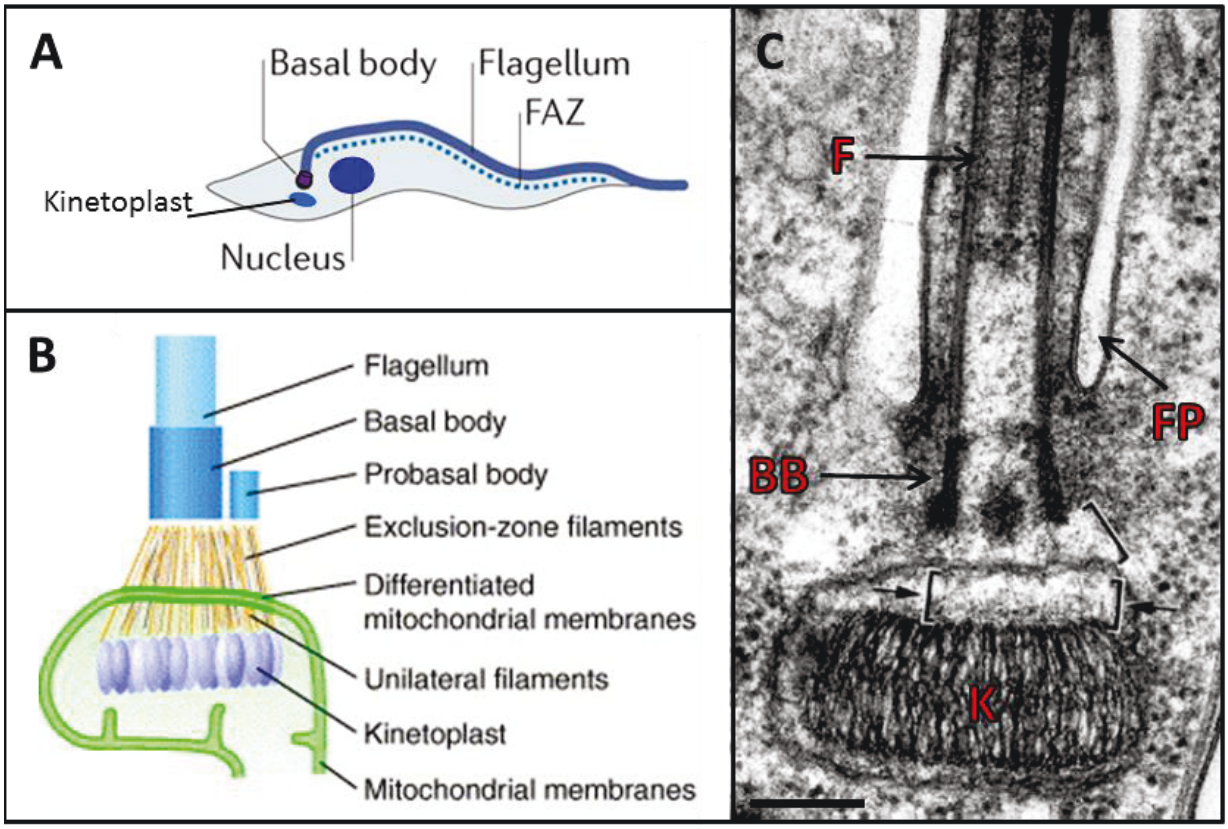


Figure 13: The kinetoplast

(A) A trypanomastigote cell showing the position of the kinetoplast within the cell, at the proximal end of the flagellum (modified from(Langousis and Hill, 2014). (B) The kinetoplast is attached to the basal body via a set of filaments (adapted from (Liu *et al.*, 2005). (C) A micrograph of the kinetoplast (K) from *Crithidia fasciculata* (taken from (Ogbadoyi *et al.*, 2003) scale bar = 250nm. Basal body; BB, Flagellum; F, Flagellar pocket; FP.

1.6.7 Flagellar pocket

The proximal end of the flagellum in trypanosomatids is enveloped within a specialised membrane invagination of the cell body membrane termed the flagellar pocket (FP; Figure 14). The FP is the exclusive site of endocytosis and exocytosis in trypanosomatids which makes it essential for normal cell cycle progression and therefore it is an attractive target for vaccine development (reviewed by (La Greca and Magez, 2011). The FP is the only area of trypanosome cell membrane where SMt are absent (see review (Landfear and Ignatushchenko, 2001). The FP of *T. brucei* is structurally analogous to the mammalian ciliary pocket (Ghossoub *et al.*, 2011; Molla-Herman *et al.*, 2010). In *T. brucei*, the basal body docks to the flagellar pocket membrane like the centriole docks to the cell surface membrane of mammalian cells. In *T. brucei*, once the basal body has docked, it

remains there indefinitely whereas in mammalian cells the centriole docks and separates from the ciliary pocket/cell membrane cyclically.

The boundary of the FP is defined by the flagellar pocket collar (FPC; Figure 14), which is identifiable in electron micrographs as a dense band of material at the neck of the FP (Lacomble *et al.*, 2009; Sherwin and Gull, 1989). There has only been one protein component of the FPC identified, BILBO1 (Bonhivers *et al.*, 2008b). The FPC is important for structural integrity of the FP. In BSF *T. brucei*, 3D reconstruction of a tomogram demonstrates a neck channel (Gadelha *et al.*, 2009), presumably to facilitate material trafficking in and out of the pocket. This neck area is also present in PCF cells (Lacomble *et al.*, 2009).

1.6.8 Flagellum attachment zone

The flagellum is attached to the cell body via the flagellum attachment zone (FAZ) (Kohl *et al.*, 1999) which is a set of maculae running along the cell body from the flagellar pocket to the anterior end of the cell (Woods *et al.*, 1989a). The FAZ is known to be required for correct cell morphology and cell division (Lacomble *et al.*, 2011; Vaughan *et al.*, 2008; Zhou *et al.*, 2011). The FAZ is a structure that runs the length of the flagellum (except distal tip where the flagellum is no longer attached to the cell body) with electron dense structures occurring with a periodicity of ~95nm (Vickerman, 1969) called the FAZ filament (FAZ_f, Figure 11). In addition to the FAZ filament, 'staple' like structures have also been described in procyclic form cells (Hoog *et al.*, 2012) and are proposed to play a role in attaching the flagellar and cellular membranes to each other.

1.6.9 Bilobe

The bilobe is a cytoskeletal structure only identified in *T. brucei,* located near the flagellum, distal to the FPC (Figure 14) (Esson *et al.*, 2012). The function of the bilobe is unknown although it may be involved in Golgi apparatus inheritance (de Graffenried *et al.*, 2008; Zhou *et al.*, 2010). The bilobe has 11 identified components (Morriswood *et al.*, 2013) although only 4 have been functionally characterised; TbMORN1 (Morriswood *et al.*, 2009), TbLRRP1 (Zhou *et al.*, 2010), TbCentrin2 and TbCentrin4 (Wang *et al.*, 2012).

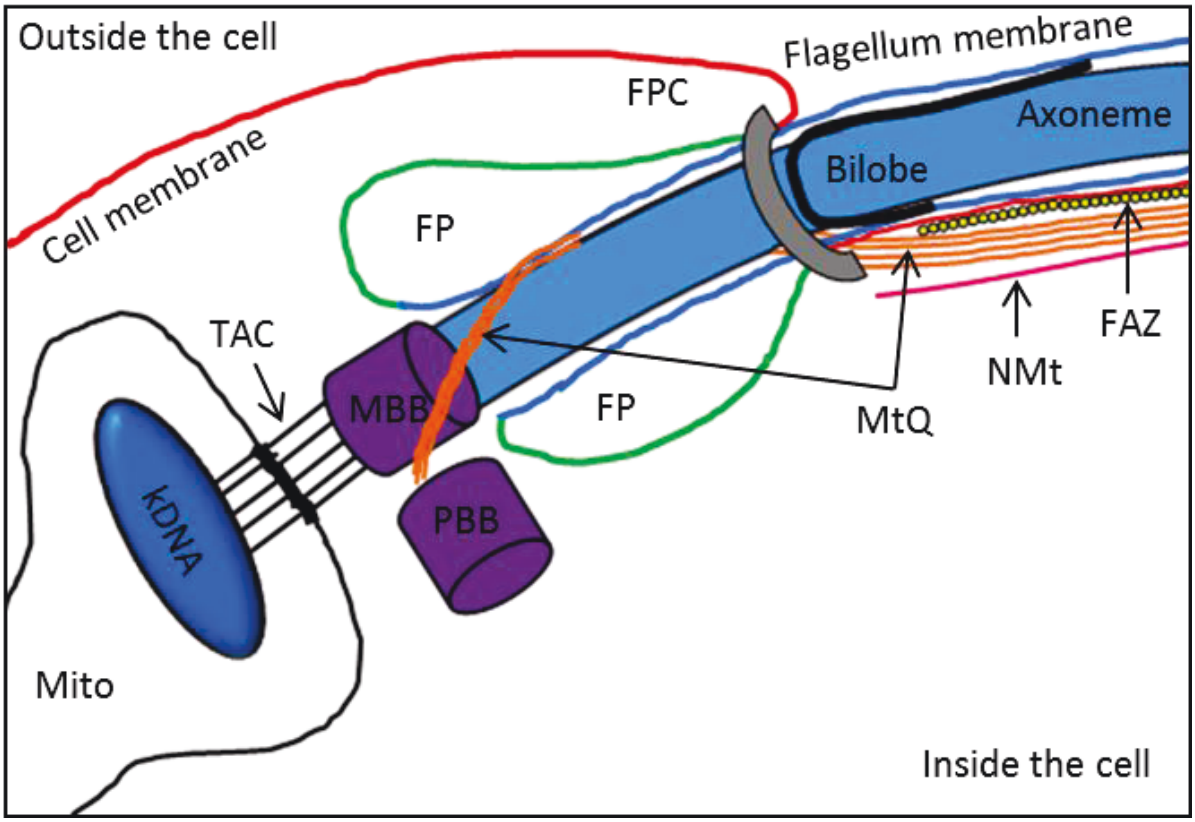


Figure 14: The flagellar pocket and associated structures

A simplified diagram of the main cytoskeletal features occurring at the proximal end of the flagellum in *T. brucei.* The kinetoplast DNA (kDNA) is positioned near the mature basal body (MBB) but is within the mitochondrion (Mito). The kDNA and MBB are connected via the filamentous tripartite attachment complex (TAC). The pro-basal body is positioned orthogonally to the MBB, which has assembled a flagellum (blue). The flagellum axoneme extends out of the flagellar pocket (FP, green), which is defined by the flagellar pocket collar (FPC, grey). The flagellum exits the pocket and extends along the cell body (red), attached via the flagellum attachment zone (FAZ, yellow). Subpellicular microtubules are not included in this diagram. The microtubule quartet (MtQ, orange) originate at the basal bodies, wrap around the axoneme once and then follow the same path of the FAZ along the cell body. The neck microtubule (NMt) is evident from the vicinity of the FPC and extends a short way along the cell dorsal side. The bilobe (black) is a fish-hook shaped structure positioned distal to the FPC. The three distinct membrane zones are demonstrated in this diagram; the FP membrane (green), the flagellum membrane (blue) and the cell membrane (red). The junction between the cell membrane and the FP membrane is at the FPC. The junction between the FP and the flagellum membrane is at the point of axoneme nucleation, the basal bodies.

1.6.10 Flagellum positioning

A structure, termed the flagellar connector (FC), is only present in the procyclic form of *T. brucei.* The FC physically connects the new flagellum (NF) to the old flagellum (OF) (Briggs *et al.*, 2004b; Moreira-Leite *et al.*, 2001) (Figure 15). It was thought that the new flagellum physically pushed its way along the cell body, using the old flagellum as a guide via the flagellar connector but further research has revealed the extension of the new flagellum may be more passive than that as when axoneme assembly does not occur the flagellum connector still moves up the old flagellum like a flag up a flag pole (Davidge *et al.*, 2006). This suggests a motor system within the FC or the old flagellum. The FC made up of three distinct zones which look like plates either side of the flagellum (Buisson and Bastin, 2010) (Figure 15; A, B and E). The FC is present before the NF exits the flagellar pocket (Figure 15; C) and resolves just prior to cleave furrow ingression (Figure 15; D). The FC is a stable structure as it remains connected despite the OF and NF being extremely motile although the structure does not remain at the distal tip of the fully extended flagellum. The FC also remains attached to both flagella when the cells are treated with a detergent to extract the membrane; therefore the FC is a cytoskeletal structure of unknown composition (Figure 15; E).

In BSF *T. brucei* the NF grows alongside the OF but without physical attachment. The distal tip of the new flagellum is thought to grow 'burrowed' in a membrane invagination which emerges at the anterior end of the cell (Hughes *et al.*, 2013). The emergence of the NF tip is labelled in BSF by the DOT1 antibody (Woods *et al.*, 1989b) as a ring/horseshoe structure (Hughes *et al.*, 2013). This labelling pattern is also seen in the related *T. cruzi* (Deane and Milder, 1973).

Figure 15: Figure legend. (A) TEM image of an assembling new flagellum (NF) physically attached to the old flagellum (OF) via the plate-like flagellar connector (FC) (∗). (B) Diagram of the FC to show that the FC is attached to outer doublets 7, 8 and 9 of the OF. The doublets that the FC is attached to in the NF are unknown. (C) The FC is formed between the NF and the OF before the NF exits the FP. (D) The FC persists as the NF extends until cytokinesis occurs. Adapted from (Langousis and Hill, 2014). (E) Examination of whole mount cytoskeletons reveals the FC is stable and remains attached. A, B, C and E Adapted from (Briggs *et al.*, 2004b). Scale bars = 200nm.

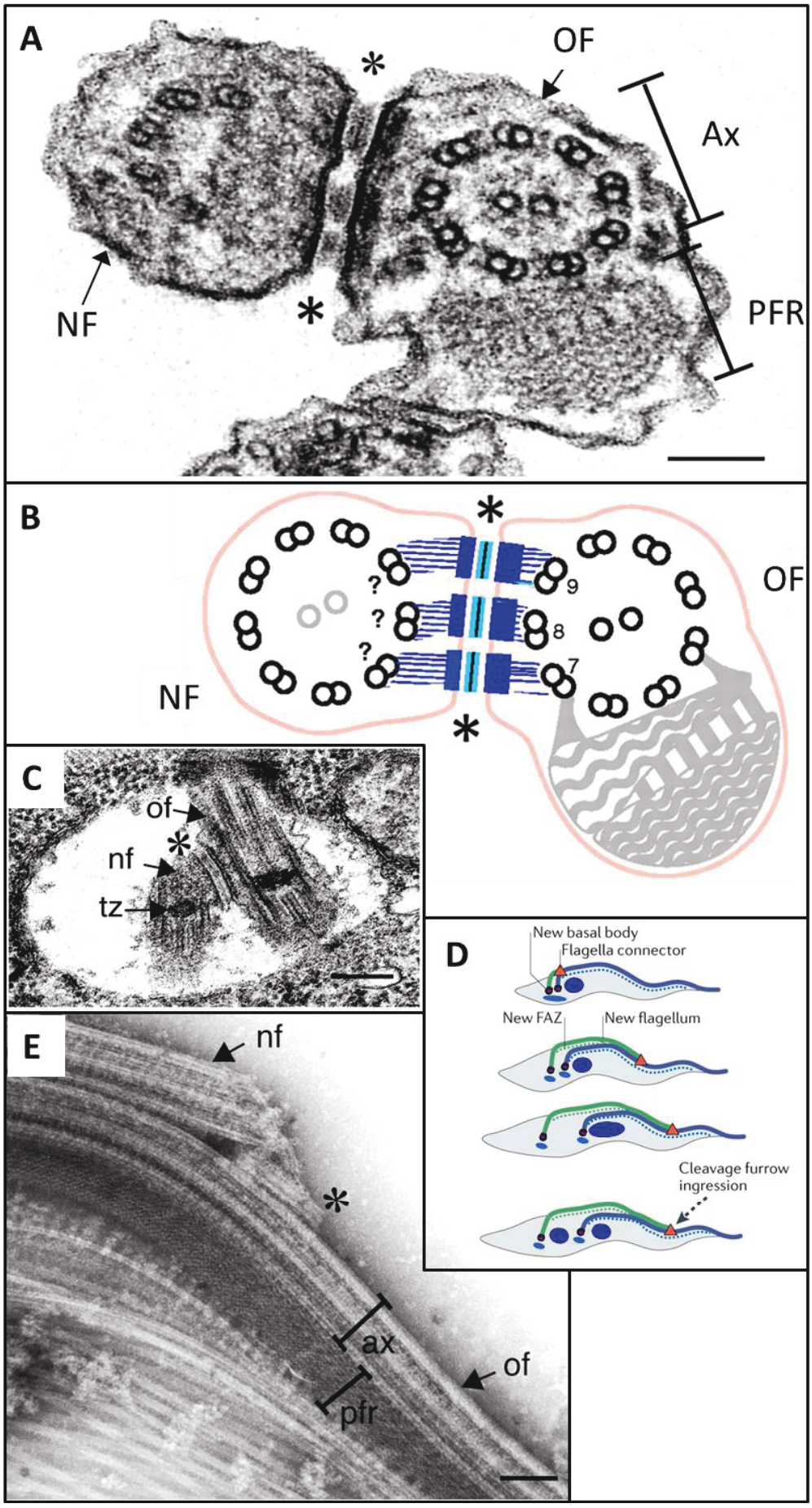


Figure 15: The flagellar connector

1.6.11 Paraflagellar rod

The paraflagellar rod (PFR) is a feature of flagella in Kinetoplastids and Euglenoids (Gallo and Schrevel, 1985; Hyams, 1982). In trypanosomes the axoneme is attached to the PFR by filamentous connectors between the outer doublets 4-7 (Hughes *et al.*, 2012) (Figure 16). It has been shown that the PFR in *Trypanosoma brucei* is essential for cell motility (Bastin *et al.*, 1998), it is not known if the same is true for other organisms with a PFR. The PFR of Trypanosomatids and Euglenoids has been shown to contain shared proteins (Gallo and Schrevel, 1985). The PFR is a complex lattice structure composed of 3 distinct zones; proximal, intermediate and distal (Bastin *et al.*, 1999a; Hughes *et al.*, 2012). In *T. brucei* the PFR structure has a repeat of 57.2nm (Hughes *et al.*, 2012).

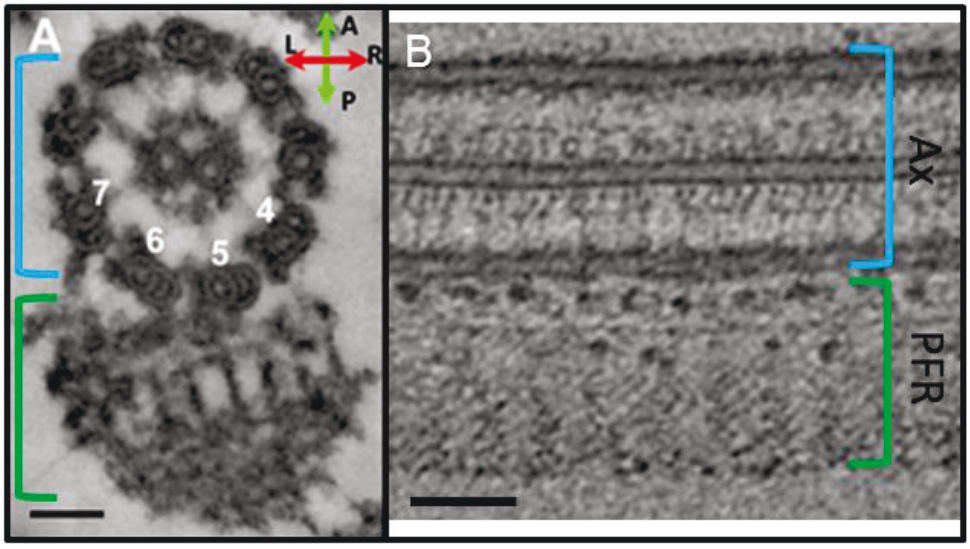


Figure 16: The paraflagellar rod

Transmission electron micrographs from detergent extracted *T. brucei* cytoskeletons. (A) Cross section through the flagellum shows the paraflagellar rod (PFR) is attached to the axoneme by connections from outer doublets 4 – 7. (B) A longitudinal view through the flagellum to show the axoneme and PFR run in parallel. Scale bars (A) 50nm, (B) 100nm. Figure adapted from (Hughes *et al.*, 2012).

1.6.12 Motility

The *T. brucei* is an unusual flagellate as the direction of movement is led by the flagellum-pulling the cell body through (Hill, 2003) whereas in other cell types the cell body in propelled by the flagellum from behind, like sperm cells (Brokaw, 1991). In addition to cell movement, the flagellum in essential for sensory functions (Maric *et al.*, 2010), cell body morphogenesis (Broadhead *et al.*, 2006; Kohl *et al.*, 2003; Vaughan, 2010), cytokinesis (Ralston *et al.*, 2006) and pathogenicity. The flagellum is crucial for disease transmission as it facilitates attachment of the parasite to the epithelium in the salivary gland of the tsetse fly vector (Rotureau *et al.*, 2014; Vickerman *et al.*, 1988). When flagella function is compromised the cells are rapidly cleared from the blood of the mammalian host (Griffiths *et al.*, 2007); forward motility is essential to force antibody-bound VSG to the posterior end of the cell and therefore towards the FP for endocytosis and degradation. In mutant *T. brucei* cell lines that swim in the opposite direction the VSG (with host immune system antibodies bound to it) pools at the anterior end of the cell and is therefore not able to reach the FP (Engstler *et al.*, 2007). This phenotype would quickly lead to host clearance of the infection.

T. brucei trypanomastigotes beat in a tip to base direction where the beat originates near the tip of the axoneme and is terminated by the restraint of the basal body (Gadelha *et al.*, 2007) which may be due to the fact that the flagellum is attached to the cell body rather than being 'free' as in *C. reinhardtii* or sperm cells. *T. brucei* trypanomastigotes are also able to beat in a base to tip direction to facilitate cell reorientation when the cell needs to perform a 'U-turn' (Branche *et al.*, 2006). In addition to the propulsive direction, the speed of the cell is also important. Cell speed in two-dimensions (2D) has been analysed using microscopy for many years (Gadelha *et al.*, 2007; Hutchings *et al.*, 2002) but recent work has allowed *T. brucei* cell motility to be observed in three-dimensions (3D) (Weisse *et al.*, 2012). Another advancement in cell speed analysis has been the introduction of heated stages (and CO_2 control for BSF cells) (Weisse *et al.*, 2012). By assessing *T. brucei* cell motility in 3D a 'quasi two state' model of motile behaviour has been proposed (Figure 17) where the cells swim, tumble or a mixture of the two.

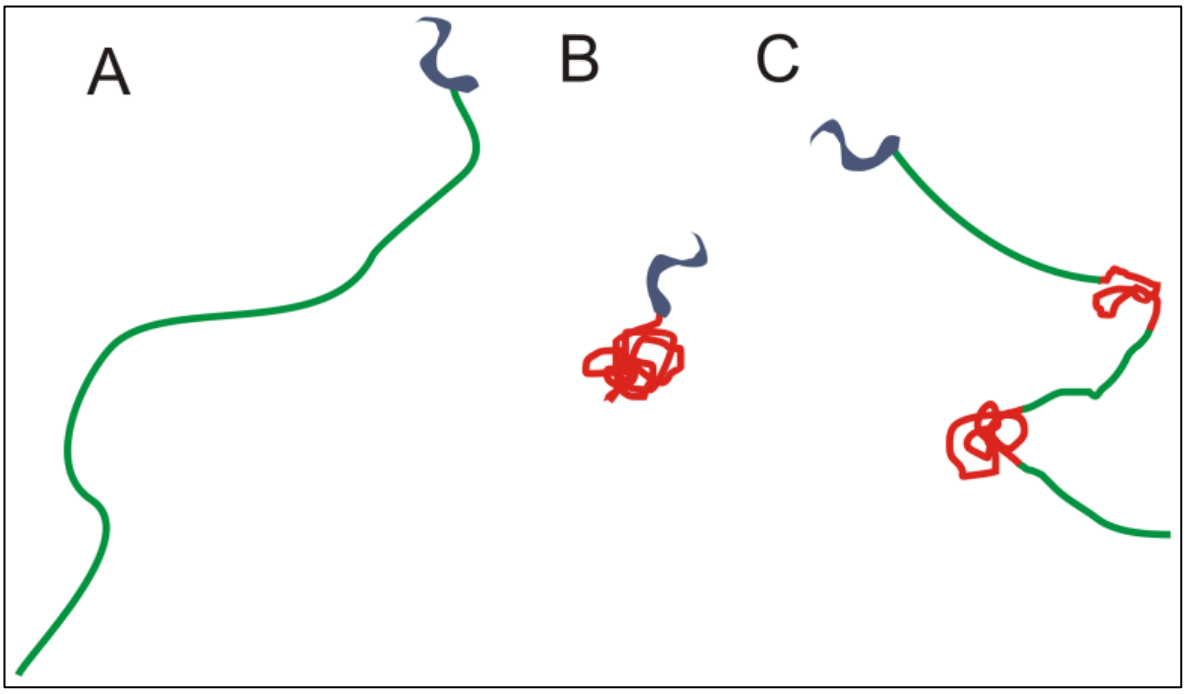


Figure 17: Cell motility of *Trypanosoma brucei*

Cartoon representation of the 'quasi two state' motility model of *T. brucei*. Figure from (Weisse *et al.*, 2012). A trypanomastigote form cell (A) swimming (green track), (B) tumbling (red track) and (C) interchanging between the two behaviours. Swimming behaviour results in the cell moving in the X and Y axis whereas tumbling behaviour usually presents the cell moving in the Z axis.

There are several cytoskeletal influences on *T. brucei* cell speed/ flagellum motility. In *C. reinhardtii* the central pair microtubules and associated protein complex rotates (Mitchell and Nakatsugawa, 2004; Omoto *et al.*, 1999) but central pair rotation does not occur in *T. brucei* (Gadelha *et al.*, 2006). It is not known what function the rotation of the central pair may have. The flagellum of *T. brucei* also includes the PFR , which aids motility (Hughes *et al.*, 2012) as mutant cell lines lacking PFR components are immotile (Bastin *et al.*, 1999b; Ginger *et al.*, 2013). The cellular architecture is also important and has evolved for efficient motility (Alizadehrad *et al.*, 2015). The BSF cells need to navigate through the host blood stream and PCF cells are required to traverse the Tsetse fly midgut to reach the salivary glands. A recent analysis on trypanomastigote morphology has evaluated many cell shape factors including the correlation between cell length and width (Wheeler *et al.*, 2013a).

Thesis Aims

The overall objective of the work described in this thesis was to identify and characterise unknown flagellum proteins in *T. brucei* and to further characterise known flagellum proteins. The specific aims of each chapter are as follows:

- To identify novel flagellum proteins in *Trypanosoma brucei* using a bioinformatics approach (Chapter 3).
- To discover members of the TPH domain-containing protein family and assess conservation of the proteins across eukaryotes (Chapter 4).
- To determine the localisation of the trypanosomal TPH domain-containing proteins using fluorescent protein tags and investigate function using RNAi ablation (Chapter 5).
- To determine the function of TbCentrin2 using inducible RNAi and endogenous localisation (Chapter 6).

2: Materials and Methods

2.1 Materials

2.1.1 Antibiotics

Table 2: Antibiotics

Antibiotic	**Working Concentration (µg/ml)**	**Catalog number**	**Supplier**
Ampicillin sodium Salt	100	A0104	Melford Laboratories Ltd.
Puromycin dihydrochloride	1	A1113803	Gibco®, Invitrogen™
Phelomycin	5	P9564	Scientific laboratory supplies Ltd.
Blasticidine S hydrochloride	20	15205	Sigma-Aldrich Co. LLC
Hygromycin B	20	A2175.0005	VWR
Doxycycline hydrochloride	1	BP2653	Fisher Scientific UK Ltd.

2.1.2 Antibodies

Table 3: Primary antibodies

Primary antibody	Class	Origin and clonality	Labels	Application	Dilution	Concentration of stock	Catalogue number (Supplier)	Reference (where appropriate)
L8C4	IgG	Mouse monoclonal	PFR2	Light microscopy	1:50	Unknown: Hybridoma supernatant	N/A	(Kohl *et al.*, 1999)
				Western blotting	1:1000			
BBA4	IgM	Mouse monoclonal	Basal body	Light microscopy	1:100	Unknown: Hybridoma supernatant	N/A	(Woodward *et al.*, 1995)
KMX-1	IgG	Mouse monoclonal	β tubulin	Light microscopy	1:50	Unknown: Hybridoma supernatant	N/A	(Birkett *et al.*, 1985)
Anti-YFP	IgG	Rabbit polyclonal	GFP	Western blotting	1:2000	0.500mg/ml	AB6556 (Abcam)	N/A

Table 4: Secondary antibodies

Secondary antibody	**Origin and clonality**	**Antigen**	**Application**	**Dilution**	**Concentration of stock**	**Catalogue number (Supplier)**	**Used to detect**
Goat anti mouse IgG-TRITC	Goat polyclonal	Mouse IgG	Light microscopy	1:200	1.5mg/ml	115-025-166-JIR (Stratech)	L8C4, KMX1
Goat anti mouse IgM-TRITC	Goat polyclonal	Mouse IgM	Light microscopy	1:200	1.5mg/ml	115-025-075-JIR (Stratech)	BBA4
Peroxidase-AffiniPure Donkey Anti-Rabbit IgG	Donkey polyclonal	Rabbit IgG	Western blotting	1:10,000	0.8mg/ml	711-035-152-JIR (Stratech)	Anti-YFP (AB6556)
Peroxidase-AffiniPure Donkey Anti-Mouse IgG	Donkey polyclonal	Mouse IgG	Western blotting	1:10,000	0.8mg/ml	715-035-150-JIR (Stratech)	L8C4

2.1.3 Plasmids for endogenous tagging

In this thesis two methods were used to insert a fluorescent tag at the endogenous locus of a target gene.

The first method used the vector, pEnT5-Y to insert a YFP at the endogenous locus using homologous recombination. Primers were designed to have the required restriction enzyme recognition site so the ORF and 5' UTR fragments were ligated into the pEnT5-Y vector backbone in the correct orientation for N terminus endogenous allele replacement. Primers were designed according to (Kelly et al., 2007) a linearisation site of NotI was used to prepare the vector for transfection.

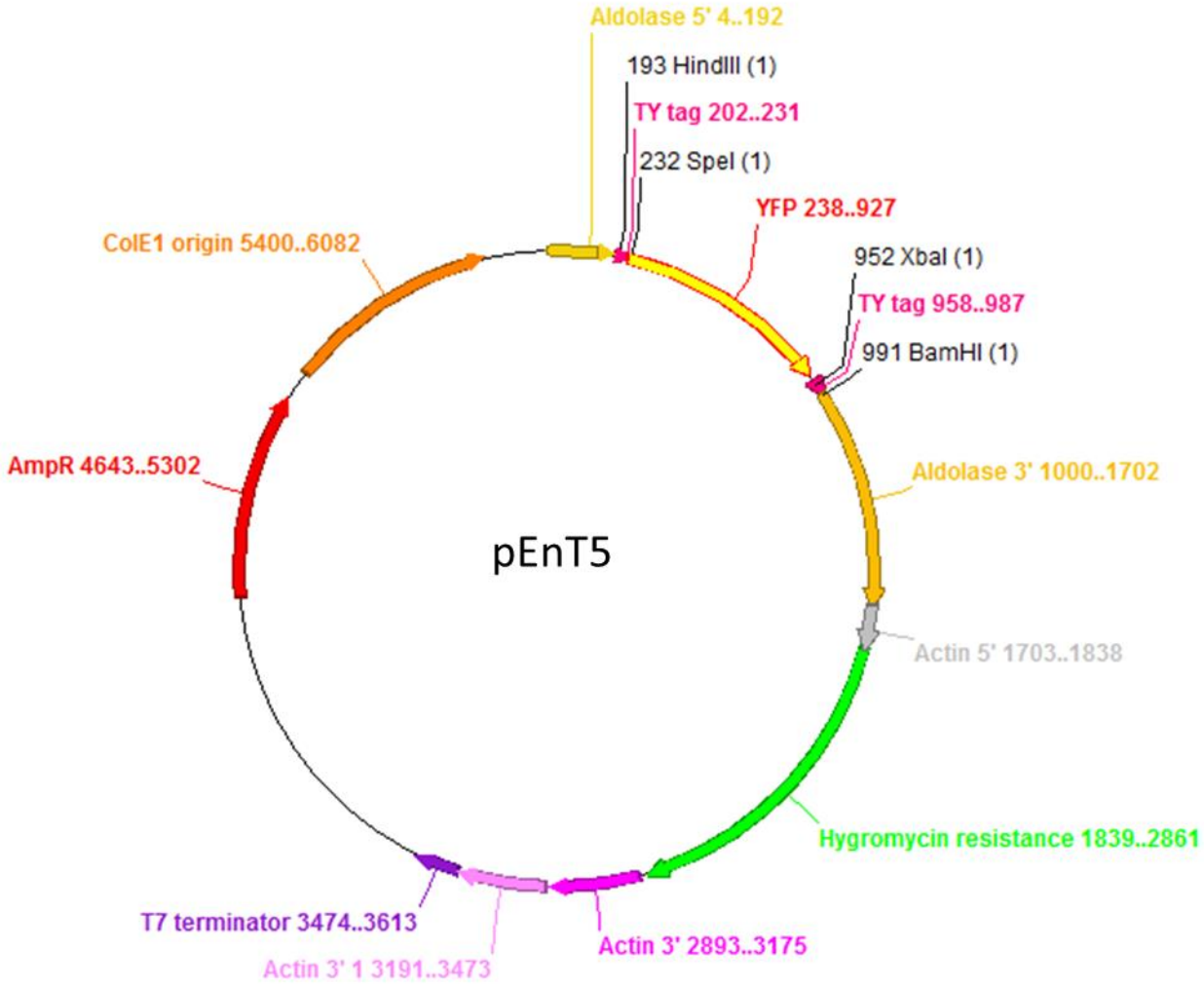


Figure 18: Map of pEnt5 plasmid for endogenous tagging.

Vector map was generated with A plasmid Editior (ApE). Vector sequence is freely available from http://www.wicksteadlab.co.uk/vectors.shtml. Vector published (Kelly et al., 2007).

The second method used the principle of PCR-based tagging which was first conceptualised in a yeast system (Janke et al., 2004) and later adapted to *T. brucei* (Arhin et al., 2004). Recently, the PCR-based tagging strategy in *T. brucei* has been simplified by the combination of an automated in silico primer design programme and a single template which can be used to amplify PCR products for tagging at the N or C terminus (Dean *et al.*, 2015).

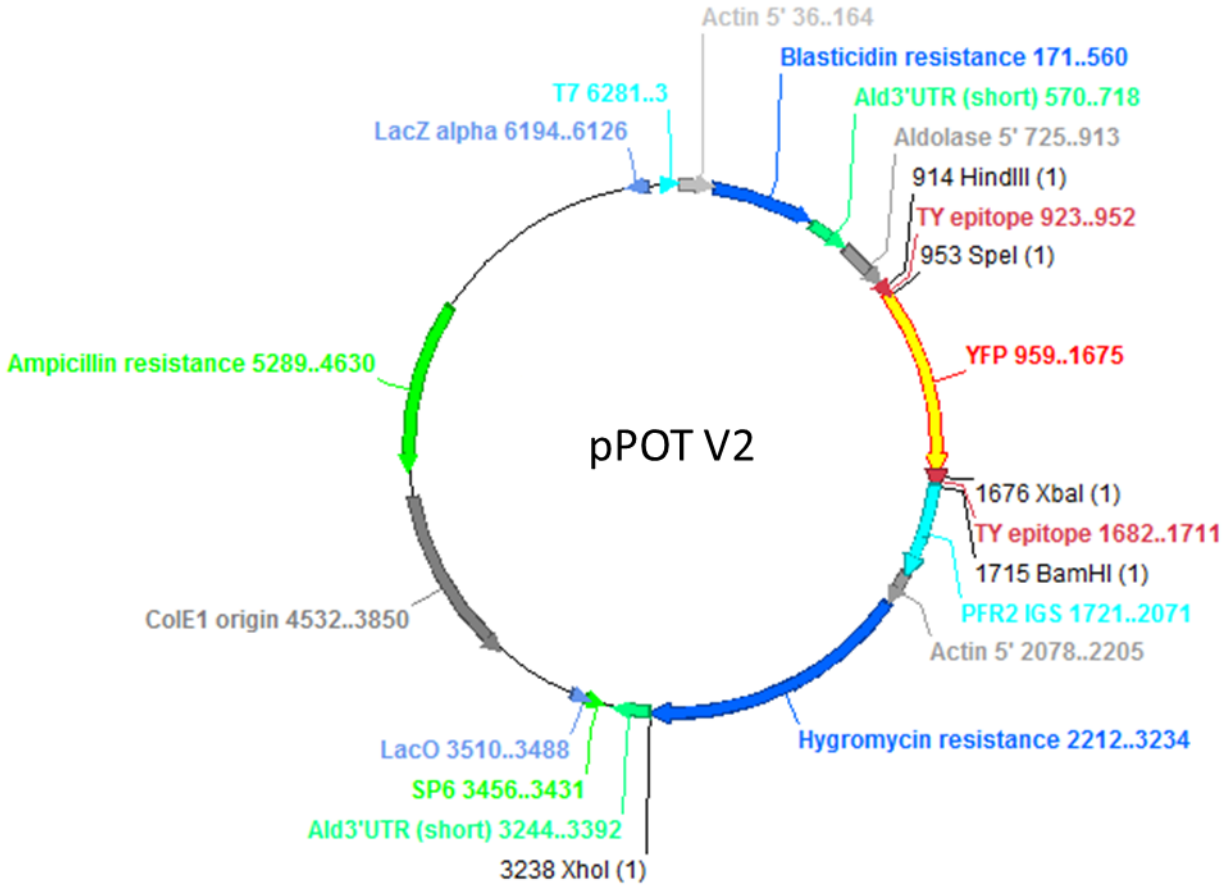


Figure 19: Map of pPOTv2 for long primer PCR

This version of pPOT had YFP as the marker with the option to use blasticidin resistance or hygromycin resistance for selection depending on which terminus the YFP was fused to. Vector map was generated with A plasmid Editior (ApE). Vector published (Dean *et al.*, 2015).

2.1.4 Plasmid for exogenous expression

The pDex777-YFP vector (Poon *et al.*, 2012) was used to achieve exogenous expression of target proteins. A linearisation site of NotI was used to prepare the vector for transfection.

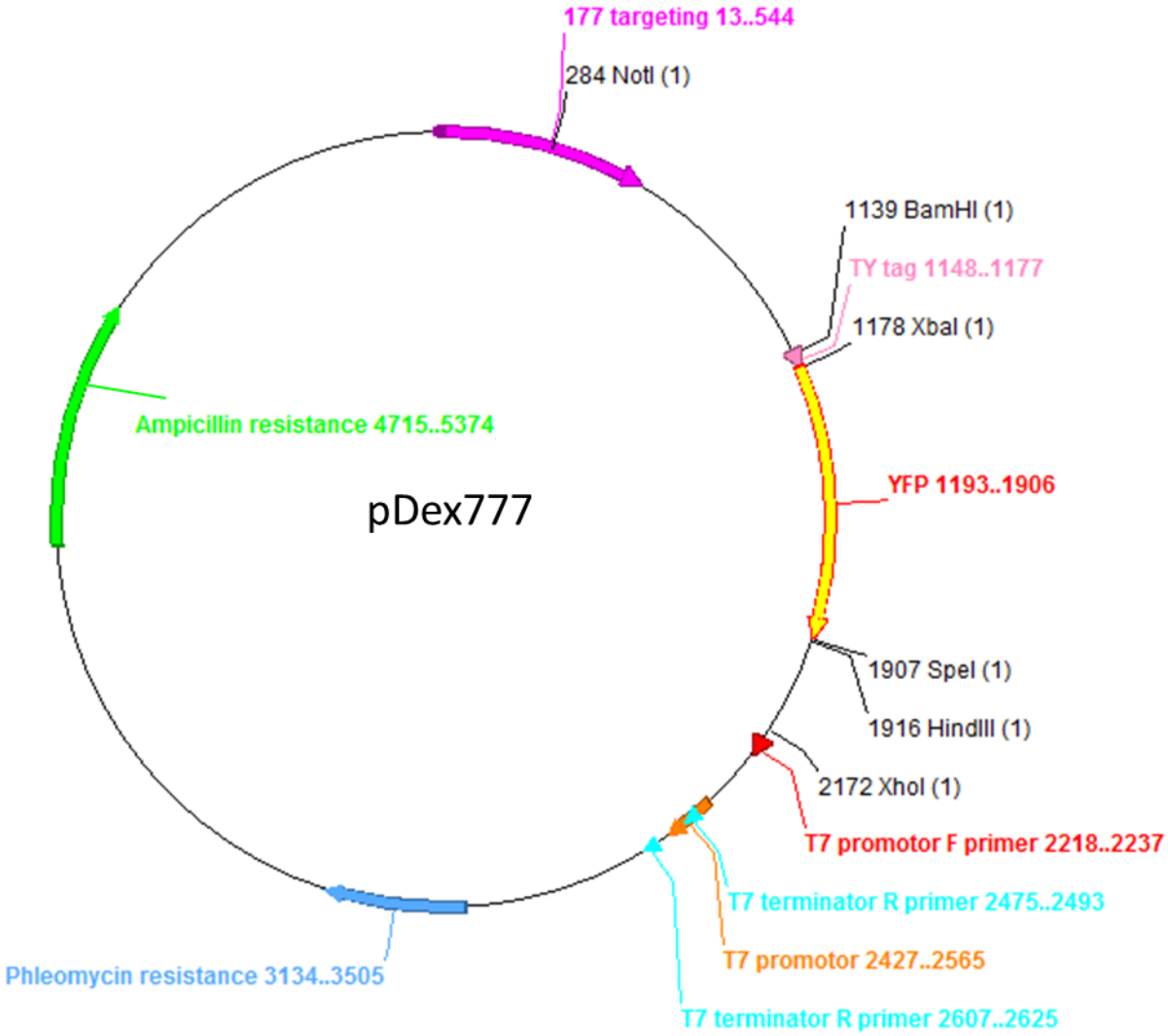


Figure 20: Map of pDex777-YFP plasmid for exogenous expression

Vector map was generated with A plasmid Editior (ApE). Vector sequence is freely available from http://www.wicksteadlab.co.uk/vectors.shtml. Vector published (Poon *et al.*, 2012)

2.1.5 Plasmid for RNA interference

The plasmid p2T7-177 (Wickstead et al., 2002) was used to create constructs for inducible RNA interference (RNAi) a linearisation site of NotI was used to prepare the vector for transfection.. Vector map and sequence is available from http://www.wicksteadlab.co.uk/vectors.shtml.

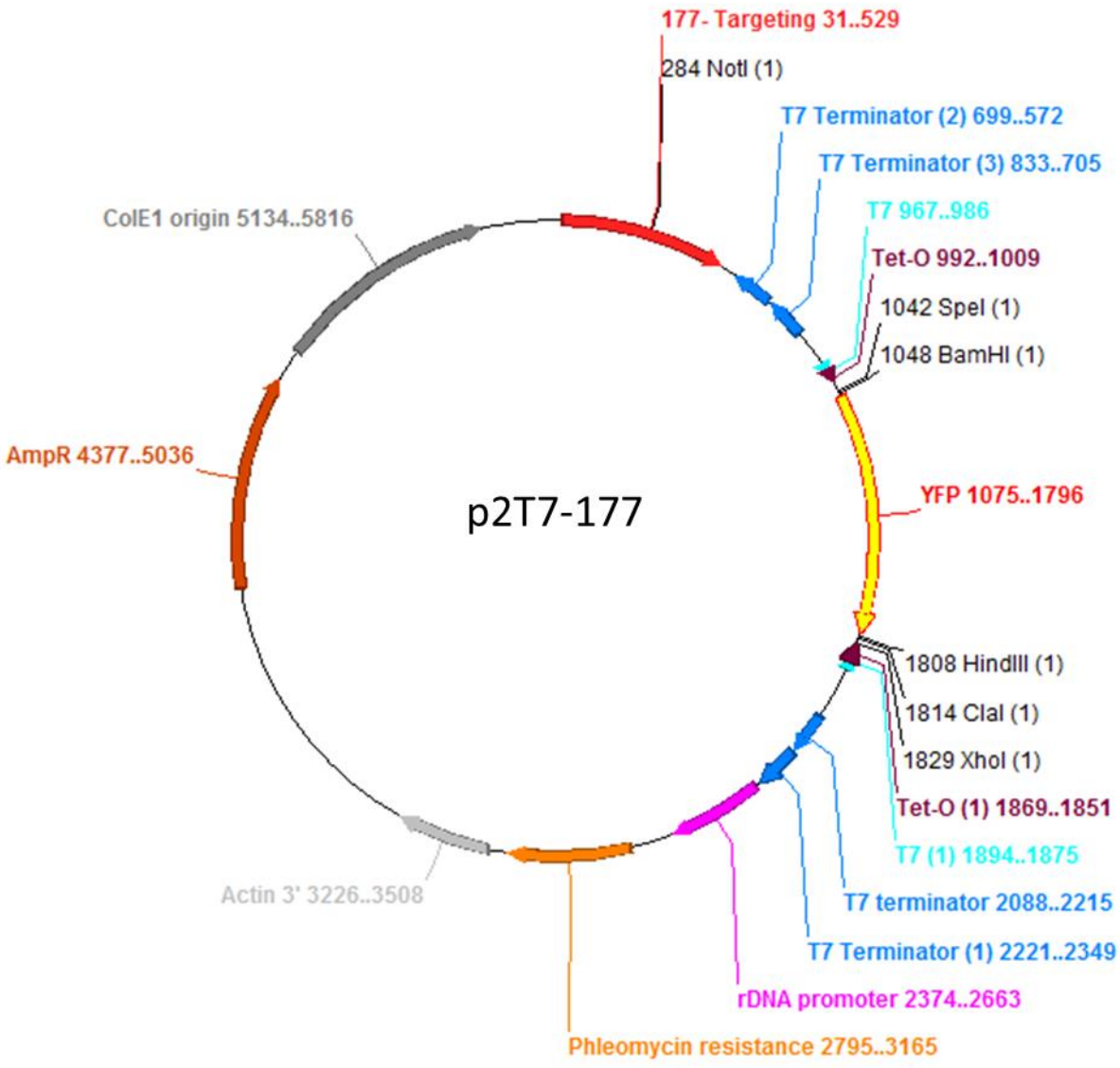


Figure 21: Map of p2T7-177 plasmid for inducible RNAi.

Vector map was generated with A plasmid Editior (ApE). Vector sequence is available from http://www.wicksteadlab.co.uk/vectors.shtml. Vector published (Wickstead et al., 2002).

2.1.6 Primers for endogenous tagging

Table 5: Sequences for endogenous tagging primers

Primers for TbHydin were generously provided by Dr Sam Dean (Oxford University).

Primer Name	Primer sequence (5' to 3')	Gene ID	Destination Vector	Template
N_TbCCDC113_F	AAACACCGTCCGTTACTGTGCTGA GAAGCGTTTTCTGCTAACTCGTTGT GTCAACGACACCCTAAACCCGCAG GAATTTCatgcctttgtctcaagaag	Tb927.7.4100	N/A	pPOTv2
N_TbCCDC113_F	TCCGCCCATAACTGTTGTTCATTGT TTTCTCCAACGAGGAACTCCAAGG TGGAACCTGCTGGCTTCACCTCAGT TGCCATcttgtacagctcgtccatgc	Tb927.7.4100	N/A	pPOTv2
N_TbDRC5_F	TAATATATATTTTTTCGCCTCTTGTT TTTGGTAATTTGTAAGTTGCTACTT TATTATATATTTGGGAAGTAAATAT TTTGatgcctttgtctcaagaag	Tb927.5.2270	N/A	pPOTv2
N_TbDRC5_R	CGAACGTGTTCCTGAGCTCGCACTC GATGGAACTGGCTACGCTGCGCCC CCTTCCTCGGAACAGTTGACGCGG TGTTCATcttgtacagctcgtccatgc	Tb927.5.2270	N/A	pPOTv2
N_TbWDR92_F	ATCTTATTTCATTTACCCCAAGAGC ACTTGCACACAAGCACATACATATT CAAACATACAGTCACTGACTCCAAC TTCCGatgcctttgtctcaagaag	Tb10.61.0160	N/A	pPOTv2
N_TbWDR92_R	TCTATGATCTGCTGCCGTGACATGC GGCGCTGCTGTTCCCGCTGTCGCTC CTTCTCCTCCTCCTCTGTCAAGTAA GACATcttgtacagctcgtccatgc	Tb10.61.0160	N/A	pPOTv2
N_TbCen2_F	TTTGCATTACTATACTCCTCTATTGT AATTTTTTCGACTTGCAGTGCCGTT GATTGATTTCTCCACCTGCCAACAC ACCTatgcctttgtctcaagaag	Tb927.8.1080	N/A	pPOTv2
N_TbCen2_R	CTCAGACTCATCCCCGGAGCTATCG AACGGTTCACACTCTGCGCGACGG GGGAATTACCGCCACGGTTGGTGG TGCTCATcttgtacagctcgtccatgc	Tb927.8.1080	N/A	pPOTv2
N_TbHydin_F	AGGCGACAGCGACAGCGAAAGCG GAGGGAAGTGAAGTGTTTGAGATT TACTGAACAAGGAGTAGAGTAGG AACAACAACCatgcctttgtctcaagaag	Tb927.6.3150	N/A	pPOTv2
N_TbHydin_R	GTCGCTGGAAGTTCAAGTGCCGCA GTTTCTGAGGCGGCTTTTCGTTCCC TCGTTACCCTTGGCGGTTGCTTTAC GGGCATcttgtacagctcgtccatgc	Tb927.6.3150	N/A	pPOTv2
N_TbCCDC11_F	TGTAAACATCAGATCTATATTAAAT TCTTGCGCCATTTAAACTGAGGCG AAAGTTTTTTTTTTGTTTTTTTAAAA AAAAGatgcctttgtctcaagaag	Tb927.5.1230	N/A	pPOTv2
N_TbCCDC11_R	ATACGCTTCTCCACAGCTTGGGCCT GTTCACGAGCGCGCTGCTCCTCCTC	Tb927.5.1230	N/A	pPOTv2

	CCGCTGTCGCCGGATTGCAACGTG AGCCATcttgtacagctcgtccatgc			
N_TbMNS1_F	TGTGACAAGAATTTCGCTATTTCAT CTTCCGTATTCGGTTGCGCACAATT GTAGTGCTTTCTTAGTTTGTGAACA TTACTatgcctttgtctcaagaag	Tb927.6.4520	N/A	pPOTv2
N_TbMNS1_R	TCCCGCGCGAGCTCTGCCTGAACA GCTTTGTAGCGATTCACACGGCGA AGCCGCTCAACACGGGCATCTCCA GCGTCCATcttgtacagctcgtccatgc	Tb927.6.4520	N/A	pPOTv2
TbCCDC19 5' UTR fragment_F	CAGGCGGCCGCCGCCGCATGTTCT GCTTCC	Tb927.8.4580	pEnT5	genomic DNA
TbCCDC19 5' UTR fragment_R	CTGGGATCCGGCGGTAAACGGCTA CAC	Tb927.8.4580	pEnT5	genomic DNA
TbCCDC19 ORF fragment_F	CAGTCTAGAATGATTGGTATGATG GCG	Tb927.8.4580	pEnT5	genomic DNA
TbCCDC19 ORF fragment_R	CTGGCGGCCGCCTGTTCCAACGGC TTGTCA	Tb927.8.4580	pEnT5	genomic DNA
TbCCDC96 5' UTR fragment_F	CAGGCGGCCGCTCCGTACTCATAT CACAAC	Tb927.7.4510	pEnT5	genomic DNA
TbCCDC96 5' UTR fragment_R	CTGGGATCCGGTTACTATACAAATT G	Tb927.7.4510	pEnT5	genomic DNA
TbCCDC96 ORF fragment_F	CAGTCTAGAATGTACGAGTATGAT GAC	Tb927.7.4510	pEnT5	genomic DNA
TbCCDC96 ORF fragment_R	CTGGCGGCCGCGTCATAGCCTGGT GTTGG	Tb927.7.4510	pEnT5	genomic DNA

2.1.7 Primers for exogenous tagging

Table 6: Sequences for exogenous tagging primers

Primer Name	**Primer sequence (5' to 3')**	**Gene ID**	**Destination Vector**	**Template**
TbCCDC19_F	CAGTCTAGAATTGGTATGATGG	Tb927.8.4580	pDex777	genomic DNA
TbCCDC19_R	GATGGATCCCTATCGCCGGGCC	Tb927.8.4580	pDex777	genomic DNA
TbRib72_F	CAGTCTAGAGCTTACCAGCAGTCAC	Tb11.47.0006	pDex777	genomic DNA
TbRib72_R	GATGGATCCTTACTTGGATTCACCTG	Tb11.47.0006	pDex777	genomic DNA

2.1.8 Primers for RNAi

Table 7: Sequences for RNA interference primers

Primer Name	Primer sequence (5' to 3')	Gene ID	Destination Vector	Template
TbCen2_F	CCCAAGCTTATAGCTCCGGGGATGAGTCT	Tb927.8.1080	p2T7-177	genomic DNA
TbCen2_R	CCCGGATCCATCAGCCTCGTCGATCATCT	Tb927.8.1080	p2T7-177	genomic DNA
TbMNS1_F	CAGCTCGAGGGAACCTAGAAGCGCAACTG	Tb927.6.4520	p2T7-177	genomic DNA
TbMNS1_R	GATGGATCCGAAGCAGCTTTTCATCCGTC	Tb927.6.4520	p2T7-177	genomic DNA
TbCCDC11_F	CAGAAGCTTTGAAGCCTTCTGTCGTGCT	Tb927.5.1230	p2T7-177	genomic DNA
TbCCDC11_R	GATGGATCCTTTTGGTTTTCCTTTGCCAC	Tb927.5.1230	p2T7-177	genomic DNA
TbCCDC19_F	CAGCTCGAGATTGGTATGATGG	Tb927.8.4580	p2T7-177	genomic DNA
TbCCDC19_R	GATGGATCCCTATCGCCGGGCC	Tb927.8.4580	p2T7-177	genomic DNA

2.1.9 Restriction enzymes

Table 8: Restriction enzymes

All restriction enzymes were used in the supplied buffer supplemented with BSA if required. Suitability of enzymes for digests was assed using the manufacturer's performance table (https://www.neb.com/tools-and-resources/usage-guidelines/nebuffer-performance-chart-with-restriction-enzymes). Double digests were performed *in silico* using the manufacturer's digest simulator (https://www.neb.com/tools-and-resources/interactive-tools/double-digest-finder).

Enzyme	Concentration (units/ml)	Catalogue number	Supplier
Not1-HF	20,000	R3189	New England Biolabs Ltd.
Xba1	20,000	R0145	New England Biolabs Ltd.
BamHI-HF	20,000	R3136	New England Biolabs Ltd.
HindIII	20,000	R3104	New England Biolabs Ltd.
ClaI	10,000	R0197	New England Biolabs Ltd.
SpeI	10,000	R0133	New England Biolabs Ltd.

2.2 Methods: Bioinformatics

2.2.1 Organisms used for conservation analysis

Extant organisms from across Eukaryota were chosen for an unbiased spread. Where possible, organisms were chosen that had a fully sequenced genome in order to prevent erroneous translation of the genome to a predictive proteome *in silico.*

Table 9: Genomes of organisms used for bioinformatical analysis

40 extant eukaryotic organisms were included in the set of 'reference' genomes used in this study to assess conservation.

Organism	**NCBI Taxid**	**Genome reference**
Acanthamoeba castellani	5755	(Clarke *et al.*, 2013)
Albugo candida	65357	(Links *et al.*, 2011)
Anopheles gambiae	7165	(Gardner *et al.*, 2002)
Apis mellifera	7460	(Honeybee Genome Sequencing, 2006)
Arabidopsis thaliana	3702	(Arabidopsis Genome, 2000)
Batrachochytrium dendrobatidis	109871	(Rosenblum *et al.*, 2013)
Bombyx mori	7091	(Mita *et al.*, 2004)
Caenorhabditis elegans	6239	(Consortium, 1998)
Chlamydomonas reinhardtii	3055	(Merchant *et al.*, 2007)
Ciona intestinalis	7719	(Dehal *et al.*, 2002)
Cryptosporidium parvum	5807	(Abrahamsen *et al.*, 2004)
Danio rerio	7955	(Howe *et al.*, 2013)
Dictyostelium discoideum	44689	(Eichinger *et al.*, 2005)
Drosophila melanogaster	7227	(Adams *et al.*, 2000)
Eimeria maxima	5804	(Blake *et al.*, 2012)
Gallus gallus	9031	(International Chicken Genome Sequencing, 2004)
Giardia lamblia	5741	(Franzen *et al.*, 2009)
Homo sapiens	9606	(Venter *et al.*, 2001)
Leishmania major	5664	(Ivens *et al.*, 2005)
Macaca mulatta	9544	(Rhesus Macaque Genome *et al.*, 2007)
Mus musculus	10090	(Mouse Genome Sequencing *et al.*, 2002)
Naegleria gruberi	5762	(Fritz-Laylin *et al.*, 2010)
Neospora caninum	29176	(Reid *et al.*, 2012)
Oncorhynchus mykiss	8022	(Berthelot *et al.*, 2014)
Paramecium tetraurelia	5888	(Aury *et al.*, 2006)
Phytophthora infestans	4787	(Haas *et al.*, 2009)
Plasmodium falciparum	5833	(Gardner *et al.*, 2002)
Rattus norvegicus	10116	(Gibbs *et al.*, 2004)
Saccharomyces cerevisiae	4932	(Goffeau *et al.*, 1996)

Schizosaccharomyces pombe	4896	(Wood *et al.*, 2002)
Stegodyphus mimosarum	407821	(Sanggaard *et al.*, 2014)
Sus scrofa	9823	(Groenen *et al.*, 2012)
Tetrahymena thermophila	5911	(Eisen *et al.*, 2006)
Thalassiosira pseudonana	35128	(Armbrust *et al.*, 2004)
Toxoplasma gondii	5811	(Kissinger *et al.*, 2003)
Tribolium castaneum	7070	(Tribolium Genome Sequencing *et al.*, 2008)
Trypanosoma brucei	5702	(Berriman *et al.*, 2005)
Volvox carteri	3067	(Prochnik *et al.*, 2010)
Xenopus tropicalis	8364	(Hellsten *et al.*, 2010)
Theileria annulata	5874	(Pain *et al.*, 2005)

2.2.2 Identifying protein orthologs

Reiterative searches using the top sequence result were used to identify orthologs using a basic local alignment search tool-protein (BLAST-P) hosted by the US National Library of Medicine (http://www.ncbi.nlm.nih.gov/Blast). These sequences were annotated as the reciprocal best blast hit (RBH).

For two protein sequences; The identity and similarity values were created using the protein sequence MATCHER pairwise alignment tool on EMBOSS (Rice *et al.*, 2000) (http://www.ebi.ac.uk/Tools/psa/emboss_matcher/). This tool users a BLOSUM62 matrix to calculate identity and similarity values. The MATCHER tool only allows two sequences to be submitted therefore for three or more protein sequences the identity and similarity values were created using the Clustal Omega multiple sequence alignment tool on EBI (http://www.ebi.ac.uk/Tools/msa/clustalo/).

2.2.3 Databases used

A number of online databases have been used during this project. These include

- Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). World Wide Web URL: http://omim.org/
- TriTrypDB, Kinetoplastid genomics resource. (Aslett *et al.*, 2010). World Wide Web URL: http://tritryp.org/

- OrthoMCL, Ortholog groups of protein sequences. (Li *et al.*, 2003). World Wide Web URL: http://www.orthomcl.org/orthomcl/

2.2.4 Identifying protein domains

The conserved domain database (CDD) (Marchler-Bauer *et al.*, 2005; Marchler-Bauer *et al.*, 2011) and Pfam, http://pfam.xfam.org/, (Punta *et al.*, 2012) (Finn *et al.*, 2014). were used to identify domains within proteins of interest via the sequence search tool. The Pfam species distribution tree was also used to identify additional proteins listed with the same domain.

2.2.5 Hidden markov model

Hidden markov models were created using Skylign (Wheeler *et al.*, 2014) (http://skylign.org/). Prior to logo generation, sequences were aligned in BioEdit as normal.

2.2.6 Creating sequence alignments

Protein sequences were aligned in BioEdit (Ibis Biosciences, California, USA) using the embedded ClustalW algorithm (Thompson *et al.*, 2002; Thompson *et al.*, 1994). Alignments were visualised and coloured (for amino acid identity and similarity) used the 'Graphic View' option in BioEdit.

2.2.7 Creating phylogenetic trees

Sequence alignments from BioEdit were saved as FAS files and imported into Seaview (Gouy *et al.*, 2010). Maximum-likelihood trees were assembled and bootstrapped with 100 replicates to give a high confidence value in branch arrangement.

2.2.8 Proteomes analysed

Relevant proteomes were selected from published literature (Table 10). Proteomes were interrogated for the presence/absence of proteins of interest.

Table 10: A reference list of proteomes analysed

Organism	**Reference**	**Focus of proteome**
Chlamydomonas reinhardtii	(Pazour *et al.*, 2005)	Flagellum
	(Li *et al.*, 2004)	Flagella and basal body
	(Stolc *et al.*, 2005)	Flagellum
	(Keller *et al.*, 2005)	Basal body
Ciona intestinalis	(Nakachi *et al.*, 2011)	Sperm
Drosophila melanogaster	(Muller *et al.*, 2010)	Centrosome
	(Avidor-Reiss *et al.*, 2004)	Cilia
	(Dorus *et al.*, 2006)	Sperm
	(Wasbrough *et al.*, 2010)	Sperm
	(Dobbelaere *et al.*, 2008)	Centrosome
Giardia lamblia	(Lauwaet *et al.*, 2011)	Basal body
Homo sapiens	(Sauer *et al.*, 2005)	Mitotic spindle
	(Jakobsen *et al.*, 2011)	Centrosome
	(Andersen *et al.*, 2003)	Centrosome
	(Amaral *et al.*, 2013)	Sperm tail
	(Ostrowski *et al.*, 2002)	Motile cilia
	(Baker *et al.*, 2013)	Sperm head and flagellum
	(Ross *et al.*, 2007)	Differentiating to ciliated epithelia
	(Balestra *et al.*, 2013)	Centriole
Macaca mulatta	(Skerget *et al.*, 2013)	Sperm
Mus musculus	(Ishikawa *et al.*, 2012a)	Immotile cilia
	(Liu *et al.*, 2007)	Immotile cilia
	(Hoh *et al.*, 2012)	Motile cilia
	(McClintock *et al.*, 2008)	ciliogenesis
Naegleria gruberi	(Fritz-Laylin and Cande, 2010)	Basal body and flagella
Oncorhynchus mykiss	(Nynca *et al.*, 2014a)	Sperm
Paramecium tetraurelia	(Yano *et al.*, 2013)	Ciliary membrane
	(Arnaiz *et al.*, 2009)	Cilia
	(Arnaiz *et al.*, 2010)	Reciliation
Phytophthora infestans	(Judelson *et al.*, 2012)	Flagellum
Rattus norvegicus	(Mayer *et al.*, 2009)	Immotile cilia
Sus scrofa	(Narita *et al.*, 2012)	Immotile cilia
Tetrahymena thermophila	(Smith *et al.*, 2005)	Motile cilia
	(Gould *et al.*, 2011)	pellicule
	(Kilburn *et al.*, 2007)	Basal body
Trypanosoma brucei	(Broadhead *et al.*, 2006)	Flagellum
	(Hart *et al.*, 2009)	Flagellum
	(Baron *et al.*, 2007b)	Flagellum

	(Oberholzer *et al.*, 2011)	Flagellum
Xenopus tropicalis	(Hayes *et al.*, 2007)	Epithelium
	(Gache *et al.*, 2010)	Microtubule interactome

2.2.9 Predicting 3D protein structure

The three-dimensional (3D) structure of proteins were modelled using the I-TASSER server (Roy *et al.*, 2010; Zhang, 2008) (http://zhanglab.ccmb.med.umich.edu/I-TASSER/). The amino acid sequence of proteins were uploaded to the I-TASSER server as txt files or directly pasted into the online form. Once processed the predicted models were downloaded as 2D GIF files

2.3 Methods: Molecular biology

2.3.1 Primer design and plasmid construction

Constitutive expression of a YFP::TbCCDC19 fusion protein from one of the endogenous loci was generated using the pEnT5 plasmid (Kelly et al., 2007). Primer sets were manually designed according to guidelines (Kelly et al., 2007) and used to amplify two fragments from gDNA, which were cloned into pEnT5 in a three way ligation to produce a plasmid to tag TbCCDC19 at the N terminus.

Constitutive expression of the remaining endogenous expression fusion proteins (Table 5) was achieved using a PCR based tagging system where the template for the PCR is a plasmid instead of gDNA. There are a few PCR based tagging techniques in *T. brucei* (Arhin *et al.*, 2004; Oberholzer *et al.*, 2006; Shen *et al.*, 2001). A recently developed version called 'PCR only tagging' (POT) was used in this study (Dean *et al.*, 2015). The plasmid template for POT (pPOT) has an accompanying Perl script primer design tool. This was used to generate long PCR primers. To generate gene specific primers, the user enters the TriTryDB accession number and the script provides the long primer pair required for long primer PCR amplification and homologous recombination of the amplicon at the loci of interest.

An inducible expression plasmid, pDex777-YFP (Poon *et al.*, 2012) was used as the backbone to create an exogenous expression construct for TbRib72, TbCCDC19 and

TbCCDC96 (Table 6). Primers were manually designed and the ORFs were amplified by PCR and cloned into the pDex777-YFP vector using the BamHI and XbaI sites.

An inducible RNAi plasmid, p2T7-177 (Wickstead et al., 2002) was used as the backbone to create RNAi constructs for TbCCDC19, TbCCDC11, TbMNS1 and TbCen2. To generate gene-specific primers for TbCen2, TbCCDC11 and TbMNS1, the respective open reading frame (ORF) sequences were obtained from TriTrypDB (Aslett *et al.*, 2010) and entered into RNAit (Redmond et al., 2003). RNAit generated target specific primers (Table 7) that were then used to amplify a small region of the ORF from gDNA. The PCR amplicons were cloned into the p2T7-177 vector using the BamHI and HindIII sites (TbCen2 and TbCCDC11) or BamHI and XhoI sites (TbMNS1) so that the gene specific target sequence was situated between the vector's two inducbible T7 promoters. To generate an RNAi plasmid to ablate TbCCDC19, primers were manually designed and the ORF of TbCCDC19 was amplified and cloned into the p2T7-177 vector using the BamHI and XhoI sites.
All custom primers (Table 5, 5 and 6) were ordered from Invitrogen (Paisley, UK).

2.3.2 Extraction of genomic DNA

Trypanosoma brucei genomic DNA (gDNA) was extracted from procyclic Lister 427 cells (Cross and Manning, 1973). To extract the gDNA, 5 x 10^6 cells were pelleted using a centrifuge at 800 x g for 5 minutes. The media supernatant was discarded and the cell pellet was suspended in 200µl PBS. A DNeasy® kit (69504, QIAGEN, Crawley, UK) was used to extract the gDNA according to the manufacturer's protocol for blood and cultured cells with the deviation of gDNA being eluted with nuclease-free water instead of the supplied EB buffer. Extracted DNA was quantified using a Nanodrop spectrophotometer (Thermo Scientific, UK) which was blanked with nuclease-free water.

2.3.3 Polymerase chain reaction

During this project, different polymerase chain reaction (PCR) reagents and conditions were used for different application, as outlined below.

To generate sequences to clone into p2T7-177, PCR was performed using Phusion® high-fidelity DNA polymerase (New England Biolabs, Hitchin, UK). For PCR 1µl of gDNA was used in a 50µl reaction containing 10µl of 5X Phusion® HF buffer, 10mM dNTPs, 10µm of each primer, 0.5µl Phusion® DNA polymerase with the remaining volume made up of Nuclease-free water. The lid of the thermocycler was heated to 110°C prevent sample evaporation. The parameters were as follows, the template DNA was denatured at 98°C for 30 seconds before 35 cycles of denaturation at 98°C for 8 seconds, annealing of the primers at 60°C for 8 seconds and elongation at 72°C for 40 seconds, then a final elongation at 72°C for 5 minutes.

To generate sequences to clone into the pGem®-T Easy subcloning vector (Promega) PCR was performed using GoTaq® DNA polymerase (M830A, Promega, UK). This produced amplicons with a 3' poly-A tail which could be directly ligated into the pGem®-T Easy subcloning vector without undergoing restriction enzyme digestion.

To generate long primer PCR Fragments for direct transfection into *T. brucei* cells the expand high fidelity PCR system (11732641001, Roche) was used according to published protocol (Dean *et al.*, 2015).

2.3.4 Agarose gel electrophoresis

Analysis of PCR products was achieved using agarose gel electrophoresis. 1% agarose gels were made by melting electrophoresis grade agarose (Fisher Scientific) in 1 x TAE (Tris-acetate-EDTA) buffer (40mM Tris (pH7.6), 20mM acetic acid, 1mM EDTA) using a microwave. When the molten agarose had cooled, ethidium bromide (2µl of a 10µg/ml stock solution) was added which enables visualisation of the DNA using a UV transilluminator (UVP, Cambridge, UK) and image taken using a Uvitec camera (UVItec, Camrbridge, UK). The molten agarose was poured into a sealed gel tray with a multi-well comb. Once the agarose gel had set, the gel was placed in an electrophoresis tank (Scie-Plas Ltd, Cambridge, UK) which was flooded with 1 x TAE buffer until the gel surface was covered. Either a 1Kb DNA ladder (NEB) or 100bp DNA ladder (NEB) was loaded onto the gel with the DNA samples (2µl) and marker ladder were then loaded as appropriate before the gel was run at 75 volts for 1-2 hours to allow separation of the DNA fragments.

2.3.5 Extraction and purification of DNA from an agarose gel

Analysis of plasmid digest products was achieved using agarose gel electrophoresis. Agarose gels were prepared as previously described without the addition of ethidium bromide to the molten agarose. After electrophoresis, gels were briefly stained in ethidium bromide solution (2µl of a 10µg/ml ethidium bromide stock solution in 100ml 1XTAE) after electrophoresis to visualise DNA bands. The required band(s) were cut from the gel using a sterile scalpel. DNA was purified from the agarose gel slice using a QIAquick® gel extraction kit (QIAGEN) according to manufacturer's instructions. Extracted DNA was quantified using a Nanodrop spectrophotometer (Thermo Scientific, UK) which was calibrated with MilliQ ultrapure water (Millipore).

2.3.6 DNA purification

PCR products were purified using a QIAquick® PCR purification kit (28104, QIAGEN) according to manufacturer's instructions.

2.3.7 DNA sequencing

Source BioScience plc (http://www.sourcebioscience.com/) was used for DNA sequencing to verify amplification of the desired sequence. Samples were submitted via an on-site collection box that was emptied daily. The primers used for sequencing were the same custom primers used for the original PCR. An exception was the screening of pGem®-T Easy inserts when Source bioscience supplied stock primers T7F and SP6R to amplify across the multiple cloning site.

2.3.7 Restriction enzyme digest

The DNA to be digested was quantified and the appropriate restriction enzymes were added in excess to ensure efficient digestion. All restriction enzymes were used with their supplied buffer and BSA where specified. See section '2.1.9 Restriction enzymes' for restriction enzymes used in this project. Typically, purified PCR products and plasmids were separately digested for 1.5 hours at 37°C in a water bath to create complementary sticky ends necessary for ligation. Digestion products were either purified using a PCR

purification kit (QIAGEN) or for plasmids, extracted using QIAquick® gel extraction kit (28704, QIAGEN) following agarose gel electrophoresis.

2.3.8 Ligation of PCR products into pGem- T Easy sub cloning vector backbone

PCR products were amplified with GoTaq® to produce a poly-A 3' tail to enable the purified PCR fragment to be ligated into the pGem®-T Easy backbone (Promega) without the need for restriction enzyme digestion. The ligation reaction contained 50ng of vector backbone and approximately 150ng of insert DNA. Purified PCR products that were of a low concentration were concentrated using a DNA120 SpeedVac system (Thermo Scientific). Ligations were performed using T4 DNA ligase (M1801, Promega) and rapid ligation buffer (C6711, Promega). Ligations were incubated at 25°C for 2 hours or at 4°C overnight.

2.3.9 Direct ligation of PCR Products into destination vector backbone

For direct ligations in pEnT5, p2T7 and pDex777 an instant sticky end ligase master mix (M0370G, NEB) was used. Typically, the ligation reaction contained no more than 50ng of vector backbone and a minimum of 150ng of insert DNA. Ligations were incubated at 25°C in a heat block for 15 minutes before incubation on ice prior to transformation.

2.3.10 Production of chemically competent cells

DH5α™ were streaked on an LB agar plate without antibiotic selection. The LB plate was then incubated at 37°C for 16 hours. A colony was selected and transferred to 25ml of SOB media and incubated for 8 hours on a shaking platform at 37°C. After 8 hours, 16ml of the bacterial culture was transferred into a fresh volume of 800ml SOB media which contained extra $MgCl_2$ and incubated for 16 hours on a shaking platform at 37°C. When the optical density of the culture had reached 0.4-0.6 the bacterial culture was incubated on ice for 10 minutes before being centrifuged at 2500 x g at 4°C. After centrifugation, the supernatant was discarded and the cell pellet was suspended in Inoue buffer (55 mM $MnCl_2{\cdot}4H_2O$ 15 mM $CaCl_2{\cdot}2H_2O$ 250 mM KCl 10 mM PIPES (0.5M, pH 6.7)) before a second centrifugation. The supernatant was discarded and the pellets were resuspended in Inoue buffer containing 10% DMSO. Suspension was incubated on ice for 10 minutes

before competent cells were aliquot out and snap frozen in liquid nitrogen. Competent cells were stored at -80^{o}C.

2.3.11 Bacterial transformation

During this project several types of competent cells were used. These were NEB 5-alpha high efficiency competent *E. coli* cells (C2987H, NEB), NEB 5-alpha subcloning efficiency *E. coli* cells (C2988J, NEB), MAX efficiency® DH5α™ cells (18258-012, Invitrogen) or competent cells which I had produced myself (see section 2.3.10 Production of chemically competent cells). The total volume of ligation mixture was chilled on ice and 100µl competent cells were added. Cells were incubated on ice with the ligation mixture for 30 minutes before a 30 second heat shock at 42°C. Cells were then incubated on ice for 2 minutes before the addition of 50µl room temperature SOC media (15544-034, Invitrogen™). Cell suspensions were incubated at 37°C for 1 hour in a shaking incubator before the whole volume was aseptically spread on warm LB agar (BPE1425, Fisher Scientific) plates which contained ampicillin (100µg/ml) (Melford, Chelsworth, UK) for selection. Plates were inverted and incubated at 37°C overnight before any resulting colonies were counted and picked using a sterile pipette tip.

2.3.12 Bacterial culture

Transformed bacteria were grown in suspension culture in preparation for plasmid extraction using a QIAprep® Spin Miniprep kit (27104, QIAGEN). Isolated bacterial colonies were picked from transformation LB agar plates and seeded into 5ml LB broth (BPE1426, Fisher Scientific) containing ampicillin (100µg/ml) and incubated in a shaking incubator at 160 rpm at 37°C overnight.

To grow bacteria in preparation for plasmid extraction using a QIAGEN Plasmid MIDI kit (12143, QIAGEN). 1ml of fresh 5ml bacterial culture may be used to seed a 50ml LB broth culture containing ampicillin (100µg/ml) and incubated in a shaking incubator at 160 rpm at 37°C overnight.

2.3.13 Plasmid DNA extraction

After overnight growth, 1ml of the transformed bacterial culture was spun down and the plasmid DNA was extracted from the bacteria using a QIAprep® Spin Miniprep kit (27104, QIAGEN) according to manufacturer's instructions.

Plasmid DNA was extracted from the 50ml culture using a QIAprep® midiprep kit (QIAGEN) according to modified manufacturer's instructions. A deviation from the manufacturer's instructions was prior to adding the supernatant to the QIAGEN-tip, the supernatant was filtered through a 0.22µm MILLEX® filter unit (SLGS033SS, Millipore) using a 50ml luer slip syringe (12651406, Fisher Scientific) to remove contaminating cell debris.

When using a QIAGEN® plasmid *plus* midi kit (12943, QIAGEN) a Vac-Man® vacuum manifold (A7231, Promega) was used to apply a vacuum to the QIAGEN spin columns.

2.3.14 Plasmid linearisation

Prior to transfection (see 2.4.3 Transfection), to allow non-homologous recombination, it was necessary to linearise pEnT5, p2T7-177 and pDex777 plasmid constructs with the restriction enzyme, NotI-HF (NEB). Per transfection, 10-20µg of plasmid was incubated at 37°C overnight with an excessive concentration of NotI-HF, the supplied buffer and BSA. Successful linearisation of the plasmid was confirmed via agarose gel electrophoresis and the linearised plasmid was purified using a QIAquick® PCR purification kit (28104, QIAGEN) according to manufacturer's instructions.

2.4 Methods: *T. brucei* cell culture

2.4.1 Preparation of procyclic *T. brucei* culture media

Procyclic form *T. brucei* cells were grown in a semi-defined media (SDM-79) (Brun and Schonenberger, 1979) which is now commercially manufactured annually by Gibco™ (a Life Technologies brand) and supplied in powdered form (catalogue reference 074-90916 N). During this project SDM-79 powder lots 939640 and 1152685 were used.

To prepare a batch of SDM-79, one tub of SMD-79 powder (Gibco™) and 10g sodium hydrogencarbonate (236527, Sigma-Aldrich) was added to 4 litres of distilled water and the solution adjusted to 7.3 pH. This SDM-79 solution was supplemented with 37.5mg porcine hemin (51280, SLS, Hessle, UK) and a final concentration of 10% FBS (F9665, Sigma-Aldrich). During this project FBS lots 024M3397, 051M3396, 061M3397, 062M3397, 063M3397, 083M3397 and 030M3399 were used.

The SDM-79 complete media was filtered using a vacuum filter and stericup system (SCGPU11RE, Millipore) and stored at 4^{o}C until use.

2.4.2 Culture of procyclic form *Trypanosoma brucei*

Procyclic form *T. brucei* were maintained at 28°C in SDM-79 culture media. The strain used in this study were SmOx P927 *T. brucei* cells (Poon *et al.*, 2012) which were grown in an asynchronous culture maintained on 1µg ml^{-1} puromycin (Gibco®). Cultures were maintained in sterile CELLSTAR® plastic non-vented flasks (690160 and 658170, Greiner bio-one).

2.4.3 Transfection

Cells were electroporated in a Nucleofector™ (Lonza), using program X-001 and human T-cell Nucleofector™ kit (VPA-1002, Lonza) using the supplied kit cuvettes. Cells were transfected with 10 µg of purified linearised construct DNA (pEnt5, p2T7-177 or pDex777) or purified PCR product (POT). For each transfection 1 x 10^{7} cells were centrifuged at 800 x g for 10 minutes in a '1-14 Microfuge' (10014, Sigma). After the media supernatant was discarded, the cell pellets were suspended in 100µl complete Amaxa solution (18.2µl of Amaxa supplement and 81.8µl Amaxa solution, Lonza). The DNA to be transfected was then added to the cell suspension, mixed and the total volume transferred to a transfection cuvette. The cuvette was placed in the Nucleofector™ and programme X-001 was used. Transfected cells were transferred to 10ml of warm complete SDM-79 and incubated at 28°C with no drugs. After 16 hours of recovery growth the appropriate selection drugs were added to culture flasks and cultures passaged as appropriate.

2.4.4 Production of conditioned media

To prepare a batch of conditioned media, 50mls of *T. brucei* procyclic cells which were not being maintained on any drugs were cultured to a density of 1-2 x 10^7 cells/ml. The culture was centrifuged for 10 minutes to pellet the cells. The media supernatant was transferred to a sterile 50ml tube and the cell pellet was disposed of. The centrifugation step was repeated and the conditioned media supernatant was filtered through a 0.22µm MILLEX® filter unit (Millipore) to remove any remaining cells. From this filtered conditioned media, 10mls was taken and incubated at 28°C overnight to check for cell growth. If procyclic cells were present, the media was centrifuged and filtered again. If no procyclic cells grew, then the conditioned media was used for cloning out a cell line.

2.4.5 Producing a clonal population from a transfected cell line

Clonal cell lines were achieved by diluting non-clonal cells in conditioned media. For each cell line to be cloned out 18mls of conditioned media was required. Required selection drugs were added to the conditioned media.

Cells to be cloned out were diluted to a final concentration of 5 cells ml^{-1} in 20mls of conditioned SDM-79. This diluted culture was dispensed in a 96 well plate (Greiner Bio-One) with 200µl per well. Plates were sealed with parafilm M® (FIL1020, SLS) or electrical tape to prevent contamination. Plates were incubated at 28°C and checked regularly for cell growth. If a well in the 96 well plate was positive for procyclic cell growth, then these cells were considered clonal and the clonal culture was scaled up through a series of passages involving a 24 well plate, a 6 well plate and finally a 10ml flask.

2.4.6 Preparing *T. brucei* cells for cryopreservation

For cryopreservation, 10mls of *T. brucei* procyclic cell culture were centrifuged and the cell pellet was resuspended in 500µl of supernatant and 500µl freezing mix (14% glycerol in SDM-79). The 1ml cell suspension was transferred to a cryovial (CRY2210, SLS) and a coloured cryovial lid insert (CRY2268, SLS) was fitted in accordance with Table 11. The cryovial was inserted into a screw top BioJar (490, Air Sea Containers Ltd) for transport to

the -80°C freezer where cryovials were transferred to a CoolCell®SV2. Cryovials were kept at -80°C for 48 hours before being transferred to liquid nitrogen for long term storage.

Table 11: Cryovial lid insert colour code

Cryovial lid insert colour	Cell line description
Green	Parental cell line (e.g SmOx P927, 29:13, 427)
Yellow	Cell line has been stably transfected with an exogenous expression vector (e.g pDex777)
Red	Cell line has been stably transfected with an RNAi vector (e.g p2T7-177)
Blue	Cell line has been stably transfected with an endogenous tag (e.g pEnT5 or POT)
White	Cell line has been stably transfected with an endogenous tag **and** an RNAi vector

2.4.7 Revival of cryopreserved cell lines

For each cryovial to be revived, 10mls of SDM-79 was decanted into a culture flask and incubated at 28°C. The cryovial was recovered from the -80°C freezer or liquid nitrogen storage in a BioJar filled with ice. As the contents of the cryovial thawed the cells were transferred into the culture flask containing warm media. Cells were left to recover overnight before appropriate selection drugs were added to the media.

2.4.8 Plasmid induction

Monoclonal cell lines to be induced were removed from selection drugs 24 hours prior to induction. A 20ml culture of cells at a density of 1 x 10^6 cells ml-1 was spilt into two 10ml cultures to ensure identical cell concentration between the induced and non-induced flasks. One of the 10ml flasks was induced and the other was not for control purposes. Induction was achieved using doxycycline (Sigma) at a final concentration 1µg/ml.

2.4.9 Saw tooth growth curve

Data to plot saw tooth growth curves of procyclic *Trypanosoma brucei* cell lines were collected by seeding cell lines at 1 x 10^6 cells ml^{-1}, incubating flasks for 24 hours at 28°C

before counting cell density and the passaging them back to 1 x 10^6 cells ml^{-1}. Cells were cultured in the presence or absence of doxycycline (to achieve an induced and non-induced flask, respectively). This was repeated for 'X' days.

The SmOx parental cell line was used to create cell lines containing inducible vectors (see section 2.1.5 Plasmid for RNA interference) which were induced by the addition of doxycycline to the growth media. The parental SmOx cell line was grown in the presence and absence of doxycycline to ensure the drug did not affect the parental cell line in the absence of a vector.

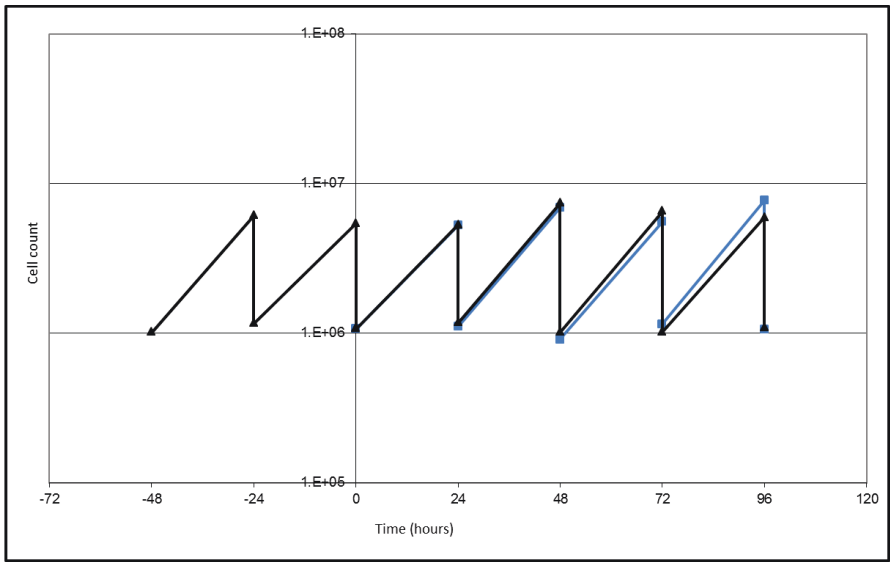


Figure 22: Parental cell line growth

The parental cell line used in this thesis, SmOx P927 was grown in the presence (blue) or absence (black) of doxycycline. There was no effect on growth of the cells caused by the presence of doxycycline in the media.

2.4.10 Continuous growth curve

Data to plot continuous growth curves of procyclic *Trypanosoma brucei* cell lines were collected by seeding cell lines at 1 x 10^5 cells ml^{-1} Flasks were incubated at 28°C and the cell density counted every 24 hours for 72 hours.

2.4.11 Cell cycle analysis

To detect changes from normal cell cycle progression, populations of cells from different time points were analysed and classified according to the number of DNA containing

organelles each cell contained. Cells were recorded with their K (kinetoplast) and N (nucleus) number e.g. 1K1N, 2K1N, 2K2N.

2.4.12 Motility assay

Cultures to be analysed (10µl) were deposited into a chamber of a C-chip haemocytometer (DHC-N01, LabTech). The haemocytometer was mounted on an AE31 inverted microscope (Motic Deutschland GmbH, Wetzlar, Germany). Sequential still images were captured at a rate of 2 frames per second on a Jenoptik ProgRes® camera (Indigo Scientific, UK) for 1 minute (15 sets per cell line). Sequential images were analysed in ImageJ (NIH) using a macro kindly provided by Richard Wheeler, Oxford University (Wheeler, 2012) . A copy of the macros is stored in the digital appendix (CD; inside front cover) as appendix 1.

2.5 Methods: Protein analysis

2.5.1 Preparation of whole cell protein samples

Total protein samples were prepared by isolating 2.5 x 10^7 whole cells from a mid-log asynchronous culture. Cells were pelleted and the supernatant was discarded. The pellet was washed in PBS containing protease inhibitors (p8340, Sigma) before suspension of the pellet in 100µl hot Laemmli buffer (2% SDS, 10% glycerol, 100 mM DTT, 50 mM Tris-HCl (pH 6.8), 0.2% bromophenol blue). The cell suspension was incubated at 100^0C for 5 minutes. Protein samples were stored at -80^0C.

2.5.2 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

For all protein work the mini-PROTEAN® electrophoresis system (165-8001, Bio-Rad) was used. Protein samples were resolved by electrophoresis in an acrylamide gel. To prepare an acrylamide gel a resolving gel and a stacking gel were used. The percentage of acrylamide in the resolving gel was varied, depending on the size of the protein of interest. For resolving gel component volumes see Table 12, the volumes listed are for one gel. Table 13 shows the appropriate volumes for a single 5% stacking gel.

The resolving gels were poured into the casting frame and 2mls of isopropanol were added to ensure a level top. The resolving gels were left to polymerise for a minimum of 30 minutes at room temperature. Once the resolving gel had fully polymerised the isopropanol was rinsed off with distilled water and the stacking gel was poured on top. The well comb was inserted into the stacking gel prior to polymerisation. After polymerisation of the gel, the gel(s) were inserted into the tank and flooded with running buffer (3.08g Tris, 18.8g glycine, 10ml 10% SDS per litre of H_2O). Protein samples, 20µl (5 x 10^6 cells per lane) were loaded into the wells of the polyacrylamide gel. A ColorPlus™ pre-stained protein ladder (p77115, NEB) was also loaded into one or two lanes for band size identification and gel orientation. Mini gels were electrophoresed for 1 hour at 180 volts or until the dye front ran off the bottom of the gel, which ever occurred first.

Table 12: Resolving gel components

Component	**Volume per 8% gel**	**Volume per 12% gel**
Water	3.45ml	2.45ml
30% acrylamide	2ml	3ml
1.5 molar Tris (pH 8.8)	1.9ml	1.9ml
10% SDS	75µl	75µl
10% APS	75µl	75µl
TEMED	4.5µl	3µl

Table 13: Stacking gel components

Component	**Volume per gel**
Water	2.7ml
30% acrylamide	670µl
1.5 molar Tris (pH6.8)	500µl
10% SDS	40µl
10% APS	40µl
TEMED	4µl

2.4.3 Coomassie brilliant blue staining

Polyacrylamide gels were stained to confirm the presence of protein (before protein transfer) or to confirm the absence of protein (after protein transfer). The acrylamide gels were immersed in Coomassie brilliant blue staining solution (0.2% coomassie brilliant blue (B0149, Sigma-Aldrich), 10% acetic acid and 50% methanol in distilled water). Gels were

typically stained for 4 hours before being destained overnight in destain solution (10% acetic acid and 40% methanol in distilled water).

2.5.4 Protein transfer and western blotting

Protein was transferred from the polyacrylamide gel to Immobilon PVDF membrane (IPVH00010, Millipore) via the wet transfer method using a mini trans-blot® module (170-3930, Bio-Rad). Prior to transfer the PVDF membrane was incubated in methanol for 5 minutes and then rinsed in distilled water. The cassette module was loaded as depicted in Figure 23 before inserting into the trans-blot® module. The trans-blot® module was filled with transfer buffer (192mM glycine, 25mM Tris, 20% methanol in distilled water) and a magnetic flea was submerged in the module tank. The tank was positioned on a magnetic stirrer and the transfer was run for 1 hour at 100 volts.

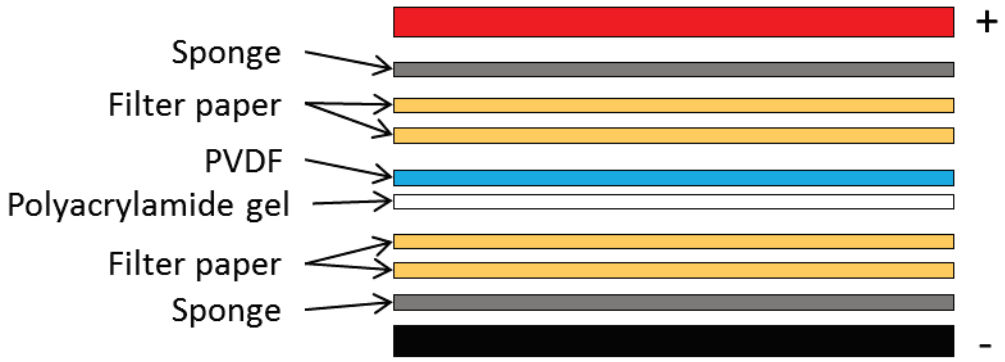


Figure 23: Wet protein transfer

The wet protein transfer cassette module was loaded in the following order black to red; sponge, 2x filter paper (170-3965, Bio-Rad), polyacrylamide gel, PVDF membrane, 2x filter paper, sponge.

Successful transfer was confirmed in two ways;

1) The pre-stained protein marker ladder had transferred from the polyacrylamide gel to the PVDF membrane

2) A coomassie brilliant blue stain of the polyacrylamide gel to confirm protein transfer (see section 2.4.3 Coomassie brilliant blue staining)

The PVDF membrane was then rinsed in Tris buffered saline (TBS) buffer (50 mM Tris-Cl, 150 mM NaCl; pH7.4) before being incubated in WB blocking buffer (5% milk powder in TBS) for one hour. PVDF membranes were incubated with primary antibody overnight at

4^{o}C before being rinsed in blocking buffer twice for 5 minutes and then incubated with the secondary antibody for 1-2 hours. Membranes were rinsed with TBS before antibody detection using the ECL substrate kit (NEL103001EA; Perkin Elmer).

2.5.5 Ponceau staining

After the PVDF membrane had been probed with antibodies and signal had detected the membrane was stained with Ponceau S staining solution (p7170, Sigma) to confirm the presence of protein on the membrane. The membrane was then washed in distilled water and allowed to dry.

2.6 Methods: Light microscopy

2.6.1 Light microscopy

Cells in mid-log culture were used to prepare samples for light microscopy. Slides were prepared using whole cells or extracted cytoskeletons.

- Whole cells were washed in PBS and settled onto SuperFrost® glass slides (MNJ-200, Fisher Scientific Ltd.) before fixation in in methanol at -20 ^{0}C for 30 minutes.
- To prepare cytoskeleton samples, the whole cells were settled onto the SuperFrost® glass slides and then incubated with 1% NP-40 in PEME buffer (100mM PIPES, pH6.9, 2mM EGTA, 1mM $MgSO_4$, 0.1mM EDTA) for 2 minutes to extract the membranes. The cytoskeletons were fixed in methanol at -20 ^{0}C for 30 minutes.

After fixation in methanol, the slides were washed in PBS for 5 minutes (repeated three times) before immunolabelling or staining with DAPI (4',6-diamidino-2-phenylindole).

An exception to the method outlined above was for DNA-containing organelle quantification (see 2.4.11 Cell cycle analysis). For this experiment, cells were settled onto slides as described and then fixed using 2% paraformaldehyde (R1018, Agar Scientific) for 10 minutes before washing the slides in PBS and staining with DAPI.

2.6.2 Immunofluorescence

As a control, all primary and secondary antibodies used in this study were applied to cells individually and screened for auto fluorescence.

Prior to antibody labelling, cells were blocked for 1 hour with 1% BSA in PBS. Cells were labelled with a primary antibody for 1 hour at room temperature (approximately 20^{o}C) before being washed in PBS 3 x 1 minute. Then cells were probed with a secondary antibody for 1 hour before being washed in PBS 3 x 1 minute. Cells were mounted in Vectashield containing 4,6-diamidino-2-phenylindole (DAPI) (H-1200, Vector Laboratories) to visualise nuclear and mitochondrial DNA a coverslip mounted and secured with clear nail varnish. Two different light microscopes were used during this project. Images were captured on a DM5500 B epifluorescence microscope (Leica microsystems) with an Orca cooled CCD camera (Hamamatsu) [for pDex777_YFP::TbCCDC19 experiments only] or on an Axio Imager Z2 microscope (Carl Zeiss Ltd) with an Orca-R^2 CCD camera (Hamamatsu) using Zen software (Carl Zeiss Ltd) [for all other experiments]. The micrographs were processed using ImageJ software (NIH).

2.6.3 Quantification of fluorescence intensity

Data was collected as Z-stacks to eliminate focal plane bias and the sum signal from 10 optical sections was normalised against the background. To measure the fluorescence intensity of basal body antigen labelling a circular region of interest (ROI) of 7 x 7 pixels (7px^2) was used. This set ROI was used for all mature basal bodies, pro basal bodies and background measurements. The mean grey value of the background was subtracted from the mean grey value of the basal body ROI to give the 'true' mean of the basal body fluorescence. The true mean of the ROI was multiplied by the area to give a value of average signal in arbitrary units per micron squared (a.u./µm^2).

2.6.4 Measurement of flagellum length

Flagella were traced using the semi-automated NeuronJ ImageJ plugin (Meijering *et al.*, 2004) and measured using the integral ImageJ measurement function. Measurements were recorded and processed in Microsoft Excel. Flagella were traced using the YFP signal (YFP::TbCen2, non-induced). In cases where flagella were being measured after the induction of RNAi an inverted phase image was used to give a pseudo-fluorescence effect (Figure 24; B, pink line). This inverted phase method was applied for TbCCDC11, TbCCDC19 and TbMNS1 analysis.

2.6.5 Measurement of cell body dimensions

A mask of the cell body was made in ImageJ. Briefly, a greyscale image of fluorescence signal (single channel) was opened in ImageJ and the image background subtracted using a measurement from background pixels. The thresholding tool was used to colour the cell body white and the background pixels black. The white cell body was selected as the ROI and this cell mask was used to calculate the cells 2D surface area and 2D cell perimeter.

Figure 24 shows an example 1K1N1F cell, to give context to the methodology used for flagellum length and cell morphometric analysis. The posterior end of the cell was measured between the posterior and the basal body (Figure 24; A). For cell body perimeter and two-dimensional (2D) surface area measurements the KMX-1 immunofluorescence image was used to outline the cell body (Figure 24; C, yellow outline) which was then used to generate a binary mask of the cell body (Figure 24; D). This mask was used to calculate the perimeter and 2D surface area of the cell body which excluded the flagellum protruding beyond the anterior end of the cell body.

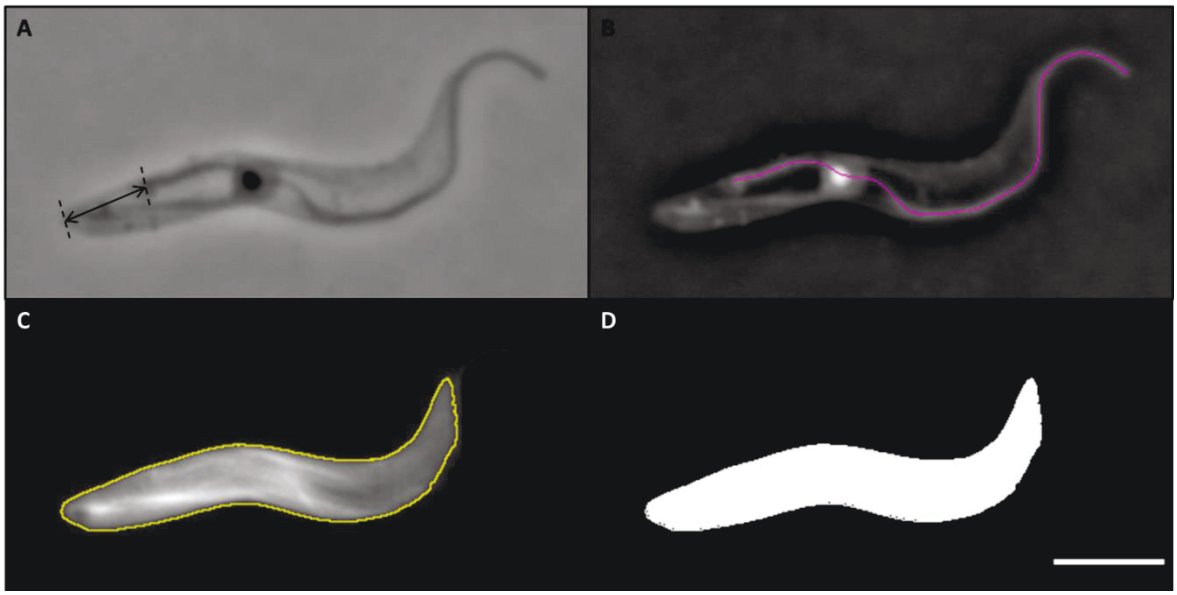


Figure 24: Cell Morphometrics

The length of the posterior end of the cell was measured from phase contrast images (A), flagellum length was measured using the NeuronJ plugin on inverted phase contrast images (B). KMX-1 immunofluorescence images were used in greyscale to create a mask of the cell body (yellow outline; C) which was used to create a binary image (D) from which cell perimeter and 2D surface area measurements were calculated. Scale bar = 5µm.

2.7 Methods: Electron microscopy

2.7.1 Scanning electron microscopy

For scanning electron microscopy (SEM) cells were fixed in suspension using a final concentration of 2.5% glutaraldehyde (TAAB, UK) in the culture media. The cells were left to fix for 5 minutes before centrifugation. The cell pellet was then washed with PBS and distilled water before cell suspension was pipetted onto Thermanox™ coverslips (M181, TAAB) and left to settle for approximately 5 minutes. Once settled, cells were dehydrated in an ascending ethanol series from 30% to 90%, 5 minutes each incubation. The cells were then incubated in 100% ethanol for 5 minutes – repeated 3 times. Once the water content of the cells had been replaced with ethanol the coverslips were critical point dried in an E3000 (Polaron, UK). The coverslips were then mounted onto 15mm SEM stubs (G3313, Agar Scientific) using sticky carbon tabs (G3347N, Agar Scientific) and coated for 30 seconds, which deposited ~17nm of gold onto the sample surface using an Agar auto sputter coater (B7431, Agar Scientific). The sample stub was inserted into a Hitachi S-3400 scanning electron microscope and the sample chamber was put under vacuum. Sample was visualised using the 'fast 1' setting, after focussing the beam a high resolution image was captured using the 'Slow 3' setting. Scanning electron micrographs were captured at a working distance of 7-9mm using secondary electrons and an accelerating voltage of 5kV. The data were processed using ImageJ software (NIH).

2.7.2 Transmission electron microscopy

Cells were harvested at mid-log phase from an asynchronous culture. Cells were fixed directly in suspension using a final concentration of 2.5% electron microscopy grade glutaraldehyde (G002, TAAB) in the culture media. After 3-5 minutes of gluteraldehyde fixation, the suspension was centrifuged for 5 minutes at 1.8 x 1000 RPM in CL2 centrifuge (Thermo Scientific). The supernatant was discarded into an aldehyde specific waste bottle and the fixed cell pellet was re-suspended in a primary fixative (2.5% glutaraldehyde, 2% paraformaldehyde and 0.1% tannic acid (T0125, Sigma-Aldrich) in 0.1M phosphate buffer (pH 7.0)). The fixed cell suspension was transferred to a 1.5ml tube. All centrifugation steps from this point were performed using a Microfuge B

(338721, Beckman) as the tube racks were fixed to spin at an angle of 90 degrees and this allowed the formation of a flat pellet.

The cell suspension in primary fixative was centrifuged for 3 minutes. The supernatant was discarded and fresh primary fixative was applied to the cell pellet. The pellet was incubated in primary fixative for 2 hours at room temperature (20^{o}C). The pellets were washed with 0.1M phosphate buffer (pH 7.0) and post-fixed in 1% osmium tetroxide (O002, TAAB) in 0.1M phosphate buffer (pH 7.0) for 1 hour at room temperature. The samples were rinsed and stained *en bloc* (in the dark) for 1 hour in 2% uranyl acetate (U011, TAAB) and then dehydrated in an ascending acetone series from 30% to 90% acetone, 15 minutes each. The samples were then transferred to 100% acetone for 30 minutes and this was repeated 3 times. To embed the sample in Agar 100 resin (R1140, Agar Scientific), the following steps were used;

- 3 parts 100% acetone to 1 part resin for 3 hours (3:1)
- 1 part 100% acetone to 1 part resin for 3 hours (1:1)
- 1 part 100% acetone to 3 parts resin for 3 hours (1:3)
- 100% resin overnight
- Fresh 100% resin for 3 hours

The resin was polymerised at 60^{0}C for 24 hours or 100^{0}C for 1 hour in a resin oven, inside a fume hood. The resin blocks were excised from their containers and the blocks were trimmed with a razor blade. For sectioning, the resin blocks were secured in a microtome chuck, mounted in a PowerTome (RMC) and trimmed with a glass knife. Once the cutting edge had been established, a diamond knife was used to obtain 70 nm sections which were floated onto water and collected onto copper grids. Micrographs were captured on a Hitachi H-7650, operated at 100kV using an AMT 2kx2k CCD camera (Advanced Microscopy Technologies, Suffolk, UK). Images were captured at 25,000 to 40,000x original magnification. The data were processed using ImageJ software (NIH).

2.7.3 Axoneme analysis

Transmission electron micrographs of axoneme cross sections were elliptically corrected and rotated using published ImageJ macros (Gadelha *et al.*, 2006). Elliptical correction was required as axonemes may not be sectioned in exact cross section and sometimes appeared ovular instead of circular. The macro is available to download from http://www.wicksteadlab.co.uk/scripts.shtml (EllipseCorrection1.0). For all elliptical corrections presented in this thesis, axonemes were arranged to be in a proximal to distal view (with the microtubule doublets arranged in a clockwise direction). The A tubule was selected and marked as the point of reference on each axonemal doublet. The A tubule of doublet 1 was selected first (directly above central pair) and then the A tubules were marked sequentially in a clockwise direction 1 through to 9.

2.7.4 Serial block face SEM

Samples were prepared as for TEM (see 2.2.38). Resin blocks were trimmed and mounted onto stubs (Gatan, UK) using superglue. For the PCF wildtype sample, the resin block was loaded into a Quanta 250 FEG (FEI, Netherlands) with a fitted Gatan 3view system. Images were recorded at an accelerating voltage of 3.4 kV and at a pressure of 50.6 pascals.

The resin block for the TbCen2 sample was loaded into a Merlin compact field emission SEM (Zeiss) fitted with a 3View® serial block-face imaging system (Gatan). After alignments were made, serial images of the block face were recorded at an accelerating voltage of 5 kV and a pressure of 45 pascals for the 'TbCen2_72 hours' sample.

Pixel size was 7.3nm (TbCen2_72 hours) or 10.99nm (wildtype) and slice thickness was 100nm for both samples. Images were recorded using Digital Micrograph (Gatan).

2.7.5 SBF-SEM data analysis

The data was originally recorded in digital micrograph as dm3 or dm4 files, which were then batch converted to tiff files. The sequential tiff files were opened in ImageJ (NIH) and assembled into a stack. Individual whole cells were selected and isolated using the sub-

stack and crop functions of ImageJ. The cropped data stacks were then imported into Amira 5.4.2 (FEI) for segmentation and rendering.

Segmentation of the cell body and flagellum was performed roughly using the manual brush tool and then refined using the automated thresholding feature of Amira. Segmentation of internal cell organelles was performed using the manual brush tool. Each cellular component was saved as a separate 'material' within Amira. Surface rendering was performed using the 'SurfaceGen' and 'SurfaceView' modules of Amira.

Measurements of the distances between organelle were conducted on the rendered surface model using the 3D measuring tool. Organelle volumes were generated using the "material statistics" module of Amira. Videos were created using the 'DemoDirector' and 'MovieMaker' modules of Amira and exported in MPEG format.

2.8 Methods: Other techniques

2.8.1 Statistical analysis

Microsoft Office Excel and GraphPad (La Jolla, California, USA, www.graphpad.com) were used to perform statistical analysis on the data. The P-value of unpaired *t* tests (Equation 1) and paired *t* tests (Equation 2) were calculated with a 99% confidence interval. Correlation coefficients (Equation 3) were compared using a confidence interval of 99% (P-value = 0.01).

$$t = \frac{\overline{X}_a - \overline{X}_b}{\overline{SE}_d}$$

Equation 1: Formula for calculating an unpaired *t* test.

The mean value of the two independent/unpaired samples (X_a and X_b) is divided by the standard error of the differences (SE_d). Figure from (pg148 (Harris *et al.*, 2005).

$$t = \frac{\bar{d}}{\overline{SE}_d}$$

Equation 2: Formula for calculating a paired *t* test.

The mean difference between each paired value (*d*) is divided by the standard error of the differences (SE_d).

The calculation was performed via the GraphPad website interface.

Figure from (pg147 (Harris *et al.*, 2005).

$$R = \frac{\sum(x - \bar{x}) \quad (y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

Equation 3: Formula for calculating the correlation coefficient.

Formular to calculate the Pearson correlation coefficient (R). Where X and Y represent the values from the two sets of data (e.g. length and speed). Figure from (pg159 (Harris *et al.*, 2005).

3. A screen to identify ciliopathy proteins in *T. brucei*

3.1 Introduction

In humans, diseases resulting from the absence or malfunction of cilia are collectively known as ciliopathies. There are a number of model organisms currently used to investigate the cause and phenotypes of these diseases. Presented here is an initial screen that was performed *in-silico* with a focus on *Trypanosoma brucei* as a suitable model organism for studying ciliopathies.

A preliminary bioinformatics study was carried out to search for novel proteins of the flagellum/cilium. Identified candidates were predicted to have orthologs in the human genome, the dysregulation of which was associated with or predicted to be involved in at least one ciliopathy condition. The bioinformatics analysis and preliminary localisation for selected candidates is presented in this chapter.

3.2 Aims

The research aims of this work are:

- To identify potential flagellum proteins in *Trypanosoma brucei* using an *in silico* bioinformatics screen
- To evaluate potential flagellum proteins through bioinformatics and assess conservation
- To determine the localisation of identified proteins through an endogenous YFP tag system and confirm if localisation is at the flagellum of *T. brucei*

3.3 Results

3.3.1 Identifying putative ciliopathy proteins

This project began with a bioinformatics analysis of published proteomic and bioinformatics studies to generate a list of uncharacterised proteins which had a putative flagellum-associated function and were conserved in *H. sapiens*. Therefore, the uncharacterised proteins were likely ciliopathy proteins conserved in *T. brucei*. The two key studies used for this analysis were the *T. brucei* flagellar proteome (TbFP) (Broadhead *et al.*, 2006) and a *T. brucei* focussed subtractive bioinformatics study for components of motile flagella (TbCMF) (Baron *et al.*, 2007b).

The list of proteins comprising the TbFP (Broadhead *et al.*, 2006) were generated by analysis of flagella harvested from procyclic (insect) forms of *T. brucei.* This provided 331 cytoskeletal proteins because any membranous components would be discarded in the supernatant fraction after detergent and salt extraction (Broadhead *et al.*, 2006). Of the 331 proteins in the TbFP, 208 were determined as trypanosomatid specific and therefore were not of interest as they would not be ciliopathy candidates.

Baron *et al.,* (2007) used bioinformatics to subtract the genomes of organisms which do not have motile cilia (*Caenorhabditis elegans* and *Arabidopsis thaliana*) from genomes of organisms that do possess motile cilia (*Trypanosoma cruzi, T. brucei, Leishmania major, Chlamydomonas reinhardtii, Homo sapiens, Mus musculus, Ciona intestinalis,* and *Drosophila melanogaster*) and therefore the remaining genes (n=50), hypothetically, have a role in motile flagella assembly or flagellum maintenance (Baron *et al.*, 2007b). The list of these 50 genes is referred to as the '*T. brucei* components of motile flagella' (TbCMF), consists of 50 genes, 30 of which had not been previously linked to flagellum form or function.

In this work the factors I considered to select proteins of interest included;

- Presence of the protein in the TbFP
- Presence of the gene in the TbCMF
- Published literature
- The online Mendelian inheritance of man (OMIM) database of genetic diseases
- Conservation between *T. brucei* and *H. sapiens*

These criteria were used to select candidate genes that had been linked to the flagellum by their presence in TbFP and/or TbCMF and whose human ortholog had been linked to a human disease – indicated by annotation in OMIM TbCMF (Baron *et al.,* 2007b; table S2) or TbFP (Broadhead *et al.,* 2006; supplementary figure 2a).

The final list of 26 candidate proteins contained 9 proteins that were found in the TbCMF and 23 proteins that were in the TbFP, 8 proteins were listed in both publications (Figure 25).

In the initial experiments, the localisation of candidates 1, 3, 6 and 7, showed localisation to the flagellum. In this chapter I outline preliminary experiments showing the bioinformatics and localisation of candidates 3, 6, 7 and 12. Candidate 1 is characterised in more depth in Chapter 4 and 5.

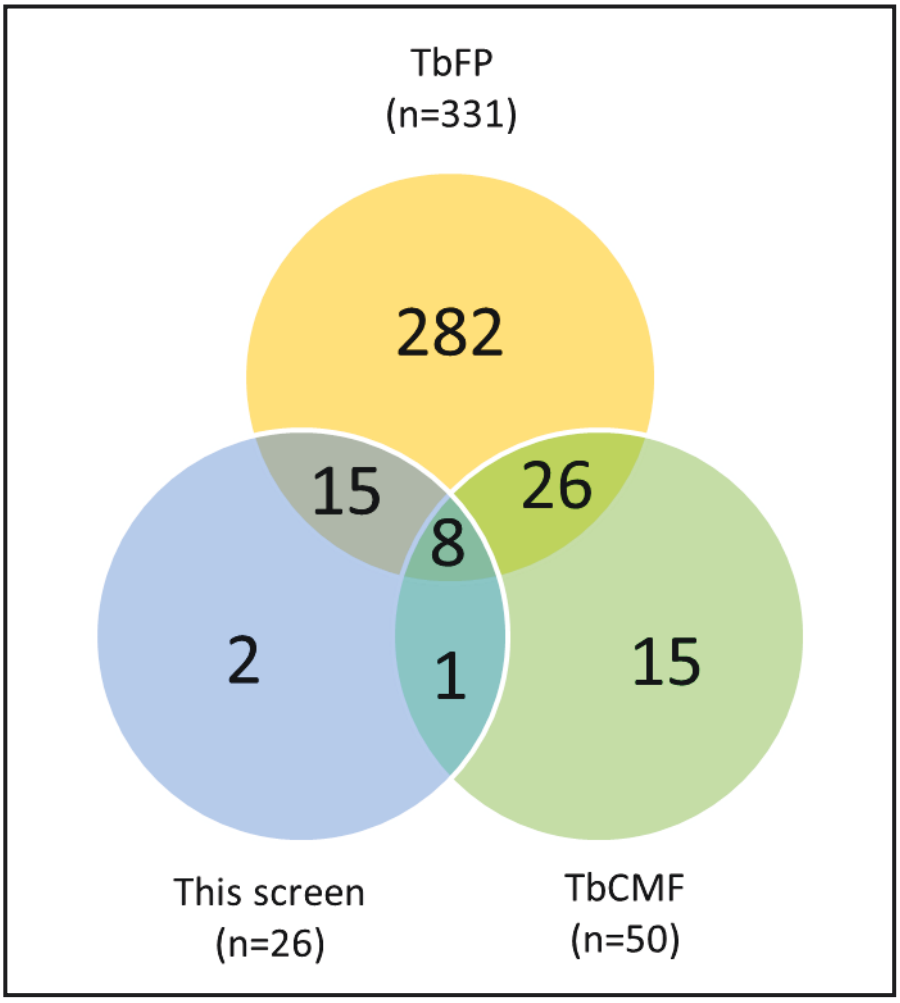


Figure 25: Occurrence of candidate proteins in previous studies

The 26 proteins identified from the preliminary bioinformatics screen. 8 were listed in both the TbFP and the TbCMF, 15 only occurred in the TbFP and 1 only occurred in the TbCMF. Two additional candidates were selected based on a review of the literature and conservation to give a total of 26 candidates.

Table 14: List of candidate proteins

Summary of the 26 candidates that have been selected for this project. A requirement for selection was the conservation between H. *sapiens* and *T. brucei*. Presence in the TbFP or TbCMF is indicated by 'Y' in the corresponding row, absence is indicated by an 'N'. Ciliopathy phenotypes derived from OMIM are listed; Bardet Biedl syndrome (BBS), cone and rod dystrophy (CORD), juvenile myoclonic epilepsy (JME), Leber congenital amaurosis (LCA), macular dystrophy (MD), polycystic kidney disease (PKD), primary ciliary dyskinesia (PCD), retinitis pigmentosa (RP), spinocerebella ataxia (SCA) and Ushers syndrome (US).

Candidate	TriTryp Accession	Conserved in *T. brucei*	Conserved in *H. sapiens*	In TbFP	In TbCMF	Confirmed or Putative link to ciliopathic conditions	Ciliopathic Clinical Symptom (OMIM)
1	Tb927.8.4580	Y	Y	Y	N	Putative	CORD, RP
2	Tb10.26.0760	Y	Y	Y	N	Confirmed	PCD
3	Tb927.7.4510	Y	Y	Y	N	Putative	RP
4	Tb927.7.6280	Y	Y	Y	N	Putative	MD
5	Tb927.5.2950	Y	Y	Y	Y	Confirmed	JME
6	Tb11.47.0006	Y	Y	Y	Y	Confirmed	JME
7	Tb927.5.2270	Y	Y	Y	Y	Putative	SCA
8	Tb10.61.2870	Y	Y	Y	N	Confirmed	RP
9	Tb11.02.4620	Y	Y	Y	N	Putative	MD
10	Tb927.10.9570	Y	Y	Y	N	Putative	PKD
11	Tb11.01.6740	Y	Y	Y	N	Putative	JME
12	Tb10.61.0160	Y	Y	N	Y	Putative	RP
13	Tb11.02.5550	Y	Y	Y	Y	Putative	CORD, RP
14	Tb09.211.2560	Y	Y	Y	Y	Confirmed	PCD
15	Tb927.4.1720	Y	Y	Y	Y	Confirmed	PCD
16	Tb927.3.1990	Y	Y	Y	N	Putative	LCA
17	Tb927.3.5140	Y	Y	Y	N	Putative	Unknown
18	Tb927.4.5380	Y	Y	Y	N	Putative	JME
19	Tb927.4.3130	Y	Y	Y	Y	Putative	Male Infertility
20	Tb927.4.1740	Y	Y	Y	Y	Putative	Deafness
21	Tb927.5.1900	Y	Y	Y	N	Putative	Deafness
22	Tb927.5.500	Y	Y	Y	N	Putative	BBS
23	Tb927.5.520	Y	Y	Y	N	Putative	PCD
24	Tb927.5.2530	Y	Y	Y	N	Putative	JME
25	Tb927.8.1680	Y	Y	N	N	Confirmed	US
26	Tb11.02.2520	Y	Y	N	N	Confirmed	US

3.3.2 Bioinformatics analysis of candidate 3: Tb927.7.4510

The initial bioinformatics screen to detect potential cilia/flagella proteins that were conserved between *H. sapiens* and *T. brucei,* identified Tb927.7.4510 as a candidate (Table 14, candidate 3). Using the basic local alignment search tool of proteins (BLASTp) on NCBI, Tb927.7.4510 was identified as an ortholog of the human protein coiled-coil domain-containing protein 96 (CCDC96), therefore Tb927.7.4510 shall be referred to as TbCCDC96. Orthologs of CCDC96 were highly conserved across eukaryotes and are known as flagellum associated protein 184 (FAP184) in the algae *C. reinhardtii* (Albee et al., 2013) and the oomycte, *Phytophthora infestans* (Judelson et al., 2012). Orthologs of CCDC96 were not found in organisms that do not build a motile flagellum; *Caenorhabditis elegans* and *Arabidopsis thaliana*. *C. elegans* is a nematode that assembles immotile, sensory cilia and *A. thaliana* is a higher land plant that has diverged from flagellated organisms and does not possess orthologs of any flagellum proteins identified to date (Hodges *et al.*, 2010; Hodges *et al.*, 2011) except α and β tubulin.

The gene encoding for CCDC96 is mapped to 4p16 in the human genome. The OMIM database was consulted to identify any potential ciliopathies associated with mutations in genes at that locus. Several diseases were listed. However the one most relevant to this project was retinitis pigmentosa 40 (#613801). Retinitis pigmentosa (RP) is a recognised ciliopathy that affects the connecting cilium of the photoreceptor cells and can cause sight loss (for reviews see (Adams *et al.*, 2007; Mockel *et al.*, 2011)). This was a piece of evidence that indicated CCDC96 could be a protein involved in the assembly or maintenance of cilia/flagella.

An alignment of the CCDC96 protein sequences from *T. brucei, H. sapiens* and *C. reinhardtii* (Figure 26; A) revealed a high level of amino acid conservation between the sequences. TbCCDC96 and HsCCDC96 were 26.9% identical and 49.3% similar at the amino acid level. The CCDC96 sequences of *H. sapiens* and *C. reinhardtii* (FAP184) were the most similar of the three with 37.1% identity and 59.2% similarity. The CrFAP184 and TbCCDC96 protein sequences shared 31.3% identity and 58.4% similarity. The protein sequence alignment (Figure 26; A) showed that most of the sequence conservation occurs near the C terminus, approximately at position 400-500 on the alignment.

Analysis of CCDC96 orthologs revealed that they contained a single conserved domain of unknown function (DUF); DUF 4201 (Pfam family 13870) (Figure 26; B).

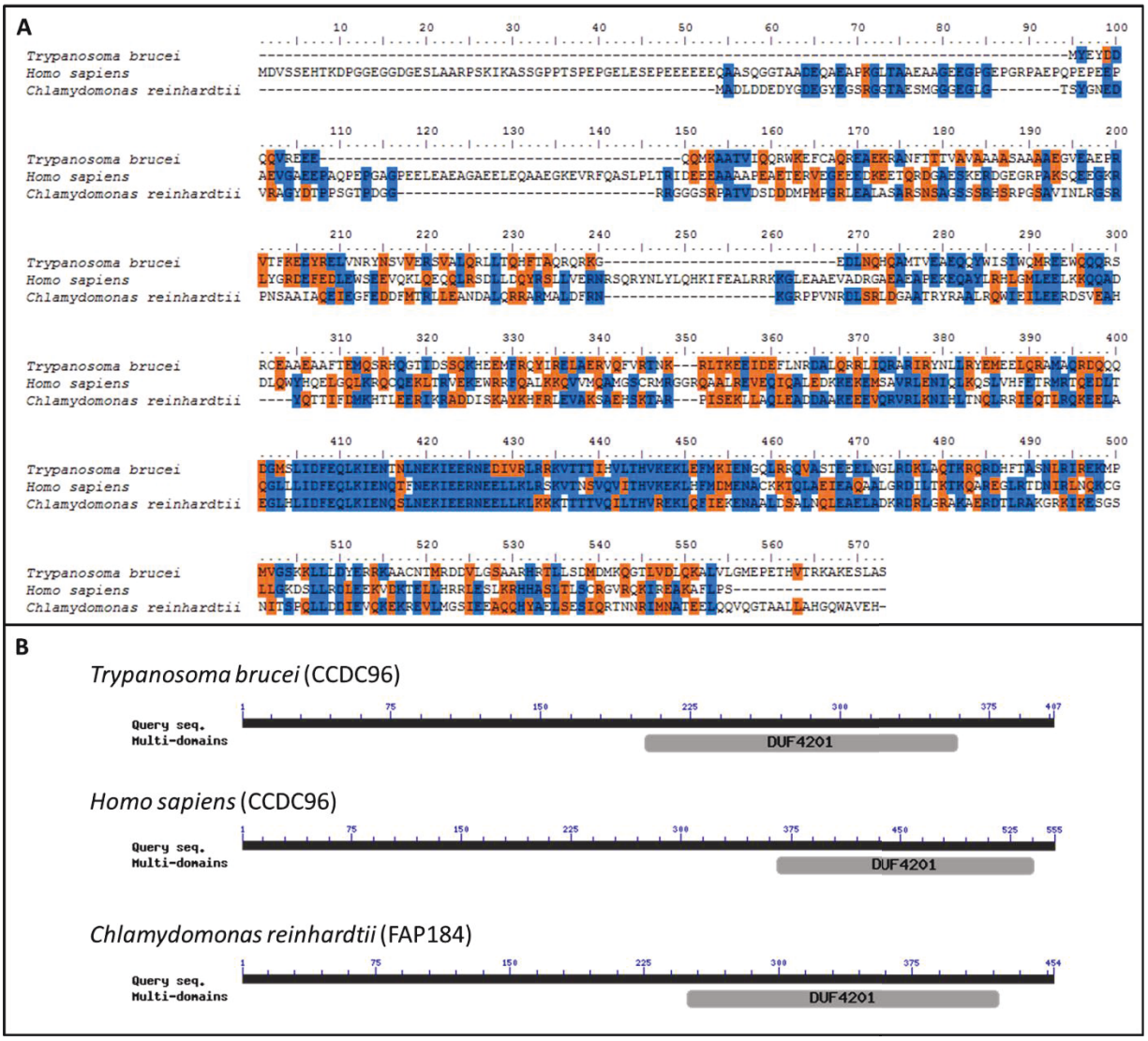


Figure 26: Assessment of CCDC96 orthologs

(A) Alignment of the orthologous protein sequences from *Trypanosoma brucei* (XP_846074), *Homo sapiens* (NP_699207) and *Chlamydomonas reinhardtii* (XP_001697427). Identical amino acids are blue and similar amino acids are orange. (B) Schematic of Tb927.7.4510 protein, HsCCDC96 and CrFAP184 showing the position of the conserved DUF4201 within the protein (NCBI). Protein diagrams are not to scale.

Further investigation revealed that only two proteins, CCDC96 and coiled-coil domain-containing protein 113 (CCDC113), contained a DUF4201 domain. CCDC113 was also a highly conserved protein with orthologs in *T. brucei* (Tb927.7.4100; TbCCDC113) and *C. reinhardtii* (FAP263). TbCCDC113 was not on the list of 26 originally identified candidates (Table 14) but was subsequently included due to interest in the DUF4201.

The gene encoding for CCDC113 is mapped to 16q21 in the human genome. The OMIM database was consulted to identify any potential ciliopathies associated with mutations in genes at that locus. Several diseases were listed, including retinitis pigmentosa 45 (#613767). RP is also linked to the area of the human genome that the CCDC96 gene was mapped to. This small family of conserved DUF4201 domain-containing proteins were strong candidates for potential ciliopathy proteins.

TbCCDC113 was not detected in the original TbFP performed on procyclic *T. brucei* (Broadhead *et al.*, 2006) but was detected in a proteomic screen of flagella from bloodstream form *T. brucei* (Dataset released online before publication (Smith and Price, 2013)). The flagellum proteome of the bloodstream form detected 184 proteins (Smith and Price, 2013) of which, 63 were not found in the previous flagellum proteome of the procyclic form (Broadhead *et al.*, 2006). When the protein sequences of selected CCDC113 orthologs were aligned (Figure 27; A) a high degree of sequence similarity (orange) and identity (blue) was seen. The DUF4201 of CCDC113 orthologs (Figure 27; B) is similarly positioned within the CCDC113 protein as it is in CCDC96 orthologs. Again, the sequences of *H. sapiens* and *C. reinhardtii* (FAP263) were the most similar of the three CCDC113 proteins with 35.5% identity and 59.4% similarity. The CrFAP263 and TbCCDC113 protein sequences shared 30.5% identity and 54% similarity. TbCCDC113 and HsCCDC113 were 27.8% identical and 50.2% similar at the amino acid level.

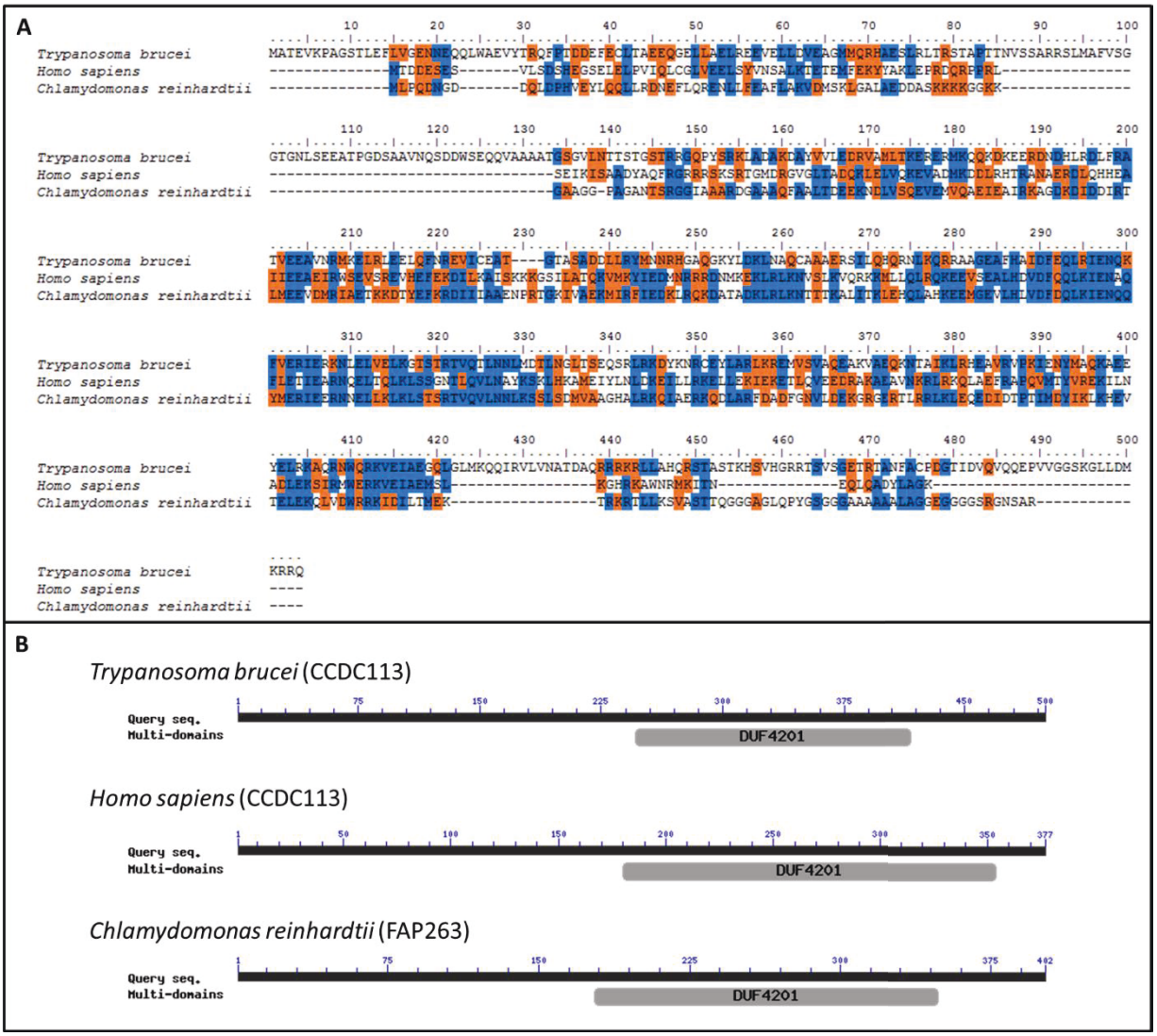


Figure 27: Assessment of CCDC113 orthologs

(A) Alignment of the orthologous protein sequences from *Trypanosoma brucei* (XP_846033), *Homo sapiens* (NP_054876) and *Chlamydomonas reinhardtii* (XP_001703742). Identical amino acids are blue and similar amino acids are orange. (B) Schematic of Tb927.7.4100 protein, HsCCDC113 and CrFAP263 showing the position of the conserved DUF4201 within the protein (NCBI). Protein diagrams are not to scale.

It is likely that DUF4201 has a function restricted to flagellum proteins due to the exclusive conservation of CCDC96 and CCDC113 in flagellated organisms. Although these two proteins were highly conserved in flagellated organisms, their detection in cilia/flagellar proteomes is inconsistent (Table 15). Neither CCDC96 or CCDC113 were detected in proteomic studies that focused on the centriole/ basal body (Keller *et al.*, 2005; Kilburn *et al.*, 2007), the centrosome (Andersen *et al.*, 2003; Jakobsen *et al.*, 2011) or mitotic spindle (Sauer *et al.*, 2005). CCDC96 and CCDC113 were also not identified in the proteomes of immotile cilia types (Ishikawa *et al.*, 2012b; Liu *et al.*, 2007; Mayer *et al.*, 2009; Narita *et al.*, 2012). As expected, CCDC96 and CCDC113 were listed as proteins found in motile cilia/flagella (Fritz-Laylin and Cande, 2010; Hoh *et al.*, 2012; Judelson *et al.*, 2012; Pazour *et al.*, 2005; Stolc *et al.*, 2005) but not ubiquitously. CCDC96 and CCDC113 were not detected in one proteome of human sperm flagella (Amaral *et al.*, 2013) or motile cilia (Ostrowski *et al.*, 2002) but other CCDCs proteins were, so presumably the sample preparation was not disruptive for coiled-coil proteins. CCDC113 was detected in the sperm proteomes of *C. intestinalis* (Nakachi et al., 2011), *H. sapiens* (Baker *et al.*, 2013) and *M. mulatta* (Skerget *et al.*, 2013) but CCDC96 was not – despite being conserved in those organisms. This could mean that CCDC113 is more abundant in the axoneme than CCDC96 or peptides from CCDC96 were incorrectly assigned to CCDC113 because of the conserved DUF4201.

Table 15: Presence of DUF4201 proteins in relevant proteomes

Relevant proteomes were analysed for the presence of TPH domain-containing proteins. Detection of the protein is indicated by a green box and if the protein was not detected a red box. If the protein was not conserved (NC) in the organism that is indicated by a grey box (see key).

Organism	Focus of Study	CCDC96	CCDC113	Reference
Chlamydomonas reinhardtii	Flagella	Yes	Yes	(Pazour *et al.* , 2005)
	Flagella	Yes	Yes	(Stolc *et al.* , 2005)
	Basal body	No	No	(Keller *et al.* , 2005)
Ciona intestinalis	Sperm	No	Yes	(Nakachi *et al.* , 2011)
Homo sapiens	Mitotic spindle	No	No	(Sauer *et al.* , 2005)
	Centrosome	No	No	(Jakobsen *et al.* , 2011)
	Centrosome	No	No	(Andersen *et al.* , 2003)
	Sperm tail	No	No	(Amaral *et al.* , 2013)
	Motile cilia	No	No	(Ostrowski *et al.* , 2002)
	Sperm head and tail	No	Yes	(Baker *et al.* , 2013)
	Ciliated cell differentiation	No	No	(Ross *et al.* , 2007)
	Centriole	No	No	(Balestra *et al.* , 2013)
Macaca mulatta	Sperm	No	Yes	(Skerget *et al.* , 2013)
Mus musculus	Immotile cilia	No	No	(Ishikawa *et al.* , 2012)
	Immotile cilia	No	No	(Liu *et al.* , 2007)
	Motile cilia	Yes	Yes	(Hoh *et al.* , 2012)
	ciliogensis	No	Yes	(McClintock *et al.* , 2008)
Naegleria gruberi	Centriole and flagella	Yes	Yes	(Fritz-Laylin and Cande, 2010)
Phytophthora infestans	Flagellum	Yes	Yes	(Judelson *et al.* , 2012)
Rattus norvegicus	Immotile cilia	No	No	(Mayer *et al.* , 2009)
Sus scrofa	Immotile cilia	No	No	(Narita *et al.* , 2012)
Tetrahymena thermophila	Motile cilia	No	No	(Smith *et al.* , 2005)
	Basal body	No	No	(Kilburn *et al.* , 2007)
Trypanosoma brucei	Flagellum (TbFP)	Yes	No	(Broadhead *et al.* , 2006)
	Flagellum (TbCMF)	No	No	(Baron *et al.* , 2007)
Xenopus tropicalis	Ciliated epithelium	No	No	(Hayes *et al.* , 2007)

Yes	Conserved in that organism and detected in the proteome
No	Conserved in that organism but not detected in the proteome
NC	Not conserved in that organism

3.3.3 Localisation of DUF4201 domain-containing proteins in *T. brucei*

To study the localisation of TbCCDC96 and TbCCDC113, stable cell lines were generated that constitutively expressed a YFP fusion protein of TbCCDC96 or TbCCDC113 from one endogenous locus. See section 2.3.1 Primer design and plasmid construction. The TbCCDC96 protein was tagged with YFP at the N terminus using the pEnT5 vector system (Kelly *et al.*, 2007) to replace one endogenous locus (Figure 28). Tagging of TbCCDC113 was achieved at the N terminus of the protein using a long primer PCR based method for gene tagging known as pPOT (Dean *et al.*, 2015) (Figure 29).

YFP::TbCCDC96 localised to the flagellum in 1K1N cells (Figure 28; A and E). This stage of the cell cycle was identified by the presence of a single kinetoplast (K, closed white arrowhead) and a single nucleus (N, open white arrowhead). The 1K1N cell had a single flagellum and single set of basal bodies (BB, Figure 28; D). The next stage in the cell cycle is known as a 2K1N cell (Figure 28; B). At this stage the cell only has one nucleus but the kinetoplast has divided into two. The cell now has 2 flagella, known as the new flagellum (NF) and old flagellum (OF) (Figure 28; B and C). The new flagellum grows alongside the old flagellum (Briggs *et al.*, 2004b) and is always positioned more posterior to the old flagellum (Robinson *et al.*, 1995). Before cytokinesis occurs, the cell is a 2K2N cell (Figure 28; C), which has two kinetoplasts (closed white arrowheads) and two nuclei (open white arrowheads). The YFP::TbCCDC96 signal was present in the NF as it extended in the 2K1N (Figure 28; B) and the 2K2N (Figure 28; C) cells. The area of the 1K1N cell (Figure 28; A) marked out by the dashed line is displayed in the inset panels (Figure 28; D –G). The phase contrast panel (Figure 28; D) shows the cell had two basal bodies (BB, black open headed arrows) and a single flagellum.

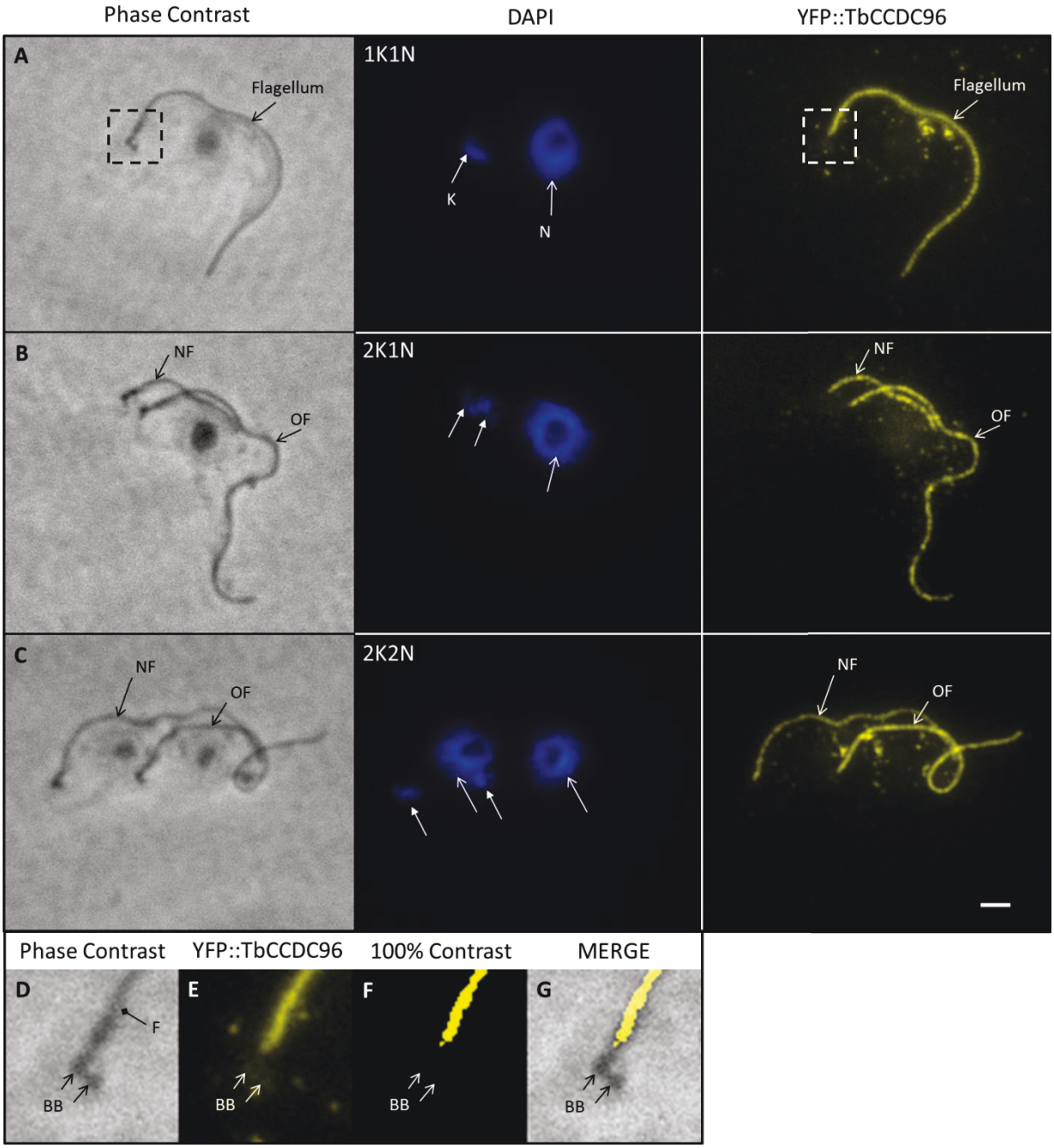


Figure 28: YFP::TbCCDC96 localises to the flagellum throughout the cell cycle

TbCCDC96 was tagged at the N terminus with YFP (YFP::TbCCDC96; yellow) and expressed from the endogenous locus in procyclic *T. brucei*. Cells were detergent extracted fixed and labelled with DAPI to stain the DNA. The YFP::TbCCDC96 signal can be seen localised to the flagellum in 1K1N (A), 2K1N (B) and 2K2N (C) cells. Insets (D - G) show the proximal end of the flagellum from the 1K1N cell (A). (D) Phase contrast image of the basal bodies (BB, black open head arrows) and the proximal end of the flagellum. (E) Native YFP::TbCCDC96 signal, which is localised to the flagellum and not to the basal bodies (BB, open head arrows). (F) Pixels of YFP::TbCCDC96 signal are made absolute by setting contrast to 100%. (G) Merge of phase contrast panel and absolute YFP::TbCCDC96 signal. NF; new flagellum, OF; old flagellum, BB; basal bodies. Scale bar = 2µm.

In summary, fluorescence microscopy of detergent extracted cytoskeletons showed that YFP::TbCCDC96 localised to the flagellum in procyclic *T. brucei* (Figure 28). Therefore, TbCCDC96 is confirmed as a cytoskeletal protein in *T. brucei*. YFP::TbCCDC113 also localised to the flagellum in 1K1N cells (Figure 29; A) and 2K2N cells (Figure 29; B). The YFP::TbCCDC113 signal was present in the NF (Figure 29; B) of the 2K2N cell, which demonstrates that it is incorporated during flagellum assembly. In summary, fluorescence microscopy of whole cells showed that YFP::TbCCDC113 localised to the flagellum in procyclic *T. brucei* (Figure 29).

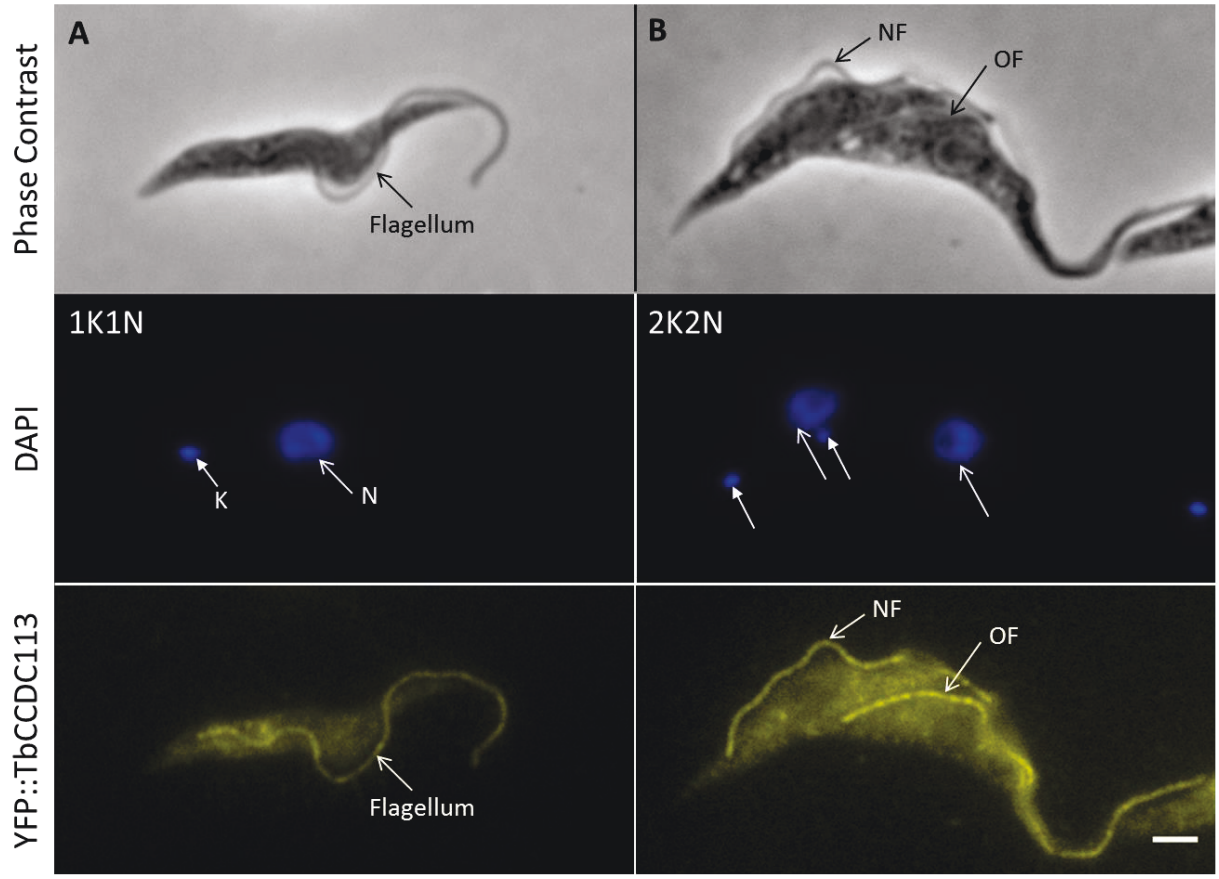


Figure 29: Localisation of YFP::TbCCDC113 to the flagellum

TbCCDC113 was tagged at the N terminus with YFP (YFP::TbCCDC113; yellow) and expressed from the endogenous locus in procyclic *T. brucei*. Cells were fixed and labelled with DAPI to stain the DNA. The YFP::TbCCDC113 signal can be seen localised to the flagellum in 1K1N (A) and 2K2N (B) cells. NF; new flagellum, OF; old flagellum. Scale bar = 2µm.

3.3.4 Bioinformatics analysis of Rib72 proteins in *T. brucei*

The main focus of this section is the Rib72 family of proteins. In this work, Tb927.11.1430 (Table 14, candidate 6) was identified as an ortholog of Rib72 using BLASTPp to find the RBH. However, during RBH analysis it became clear that there were additional proteins that were highly similar. There were three orthologs of Rib72 in *T. brucei* but they are not paralogs. The two additional Rib72 orthologs were Tb927.5.2950 (Table 14, candidate 5) and Tb927.10.7690 (not a candidate). All three TbRib72 orthologs were detected in the TbCMF (Baron *et al.*, 2007b) and in the TbFP (Broadhead *et al.*, 2006). A common feature of the TbRib72 orthologs was the inclusion of 3 x DUF1126 domains (PF06565) within the proteins. Conservation of orthologs is summarised in Table 16. The TbCMF also lists TbCMF34 (Tb927.3.1040/Tb927.3.1060) as an ortholog of Rib72 but my RBH analysis disputes as the protein only shares homolog with Rib72/EFHC1 in the DUF1126 region and not over the whole protein. Another piece of evidence that the paralogs Tb927.3.1040/Tb927.3.1060 are not Rib72 orthologs is that the accession annotation in TriTrypDB is 'cAMP response protein, putative (CARP4)'. Another protein in *T. brucei* that contains a DUF1126 domain in Tb927.4.1720 (TbCMF40 and Table 14; candidate 15), which is also not a Rib72 ortholog and is described as a 'putative nucleoside diphosphate kinase' in TriTrypDB (Aslett *et al.*, 2010).

The focus of this section is Tb927.11.1430 (Table 14, candidate 6), which by RHB Tb927.11.1430 was identified as an ortholog of HsEFHC1_1 (isoform 1). An alignment of the protein sequences (Figure 30) revealed a 31.5% identity at the amino acid level and 49.3% similarity.

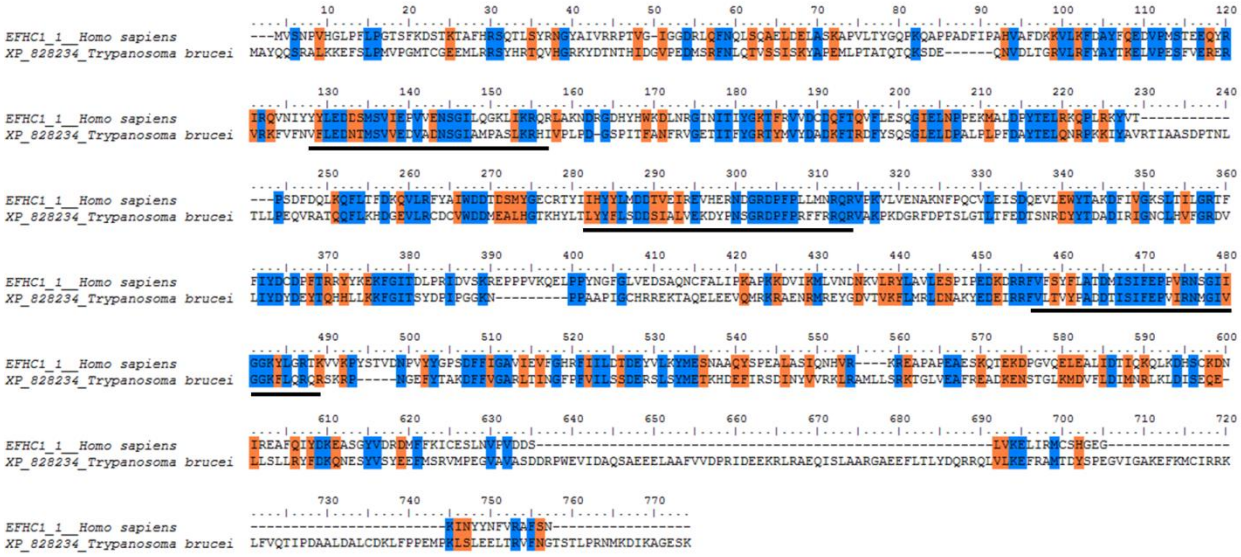


Figure 30: Alignment of TbRIb72 and HsEFHC1_1

Alignment of the orthologous protein sequences Tb927.11.1430 (TbRib72) and HsEFHC1_1. Identical amino acids are blue and similar amino acids are orange. Black lines indicate the position of the three DUF1126 domains.

3.3.5 Localisation of a Rib72 protein in *T. brucei*

Only work on one of the three TbRib72 orthologs is presented here. The pDex777 plasmid (Poon *et al.*, 2012) was used to ectopically express a copy of TbRib72 (Tb927.11.1430) fused to YFP at the N terminus. See section 2.3.1 Primer design and plasmid construction. The pDex777 vector is an inducible system for ectopic protein expression; it is induced by the addition of doxycycline to the culture media. The linearised pDex777_YFP::TbRib72 construct inserted into a transcriptionally silent region of the *T. brucei* genome known as the 177bp repeats of the mini chromosomes. Therefore when the cell line is in a non-induced state the transfected construct is silent at its insertion site and no fusion protein visible. After the addition of doxycycline, the pDex777_YFP::TbRib72 construct was induced therefore the YFP::TbRib72 fusion protein was expressed. When cells were fixed and examined by light microscopy 72 hours post induction of pDex777_YFP::TbRib72, the YFP fusion protein localised to the flagellum (Figure 31).

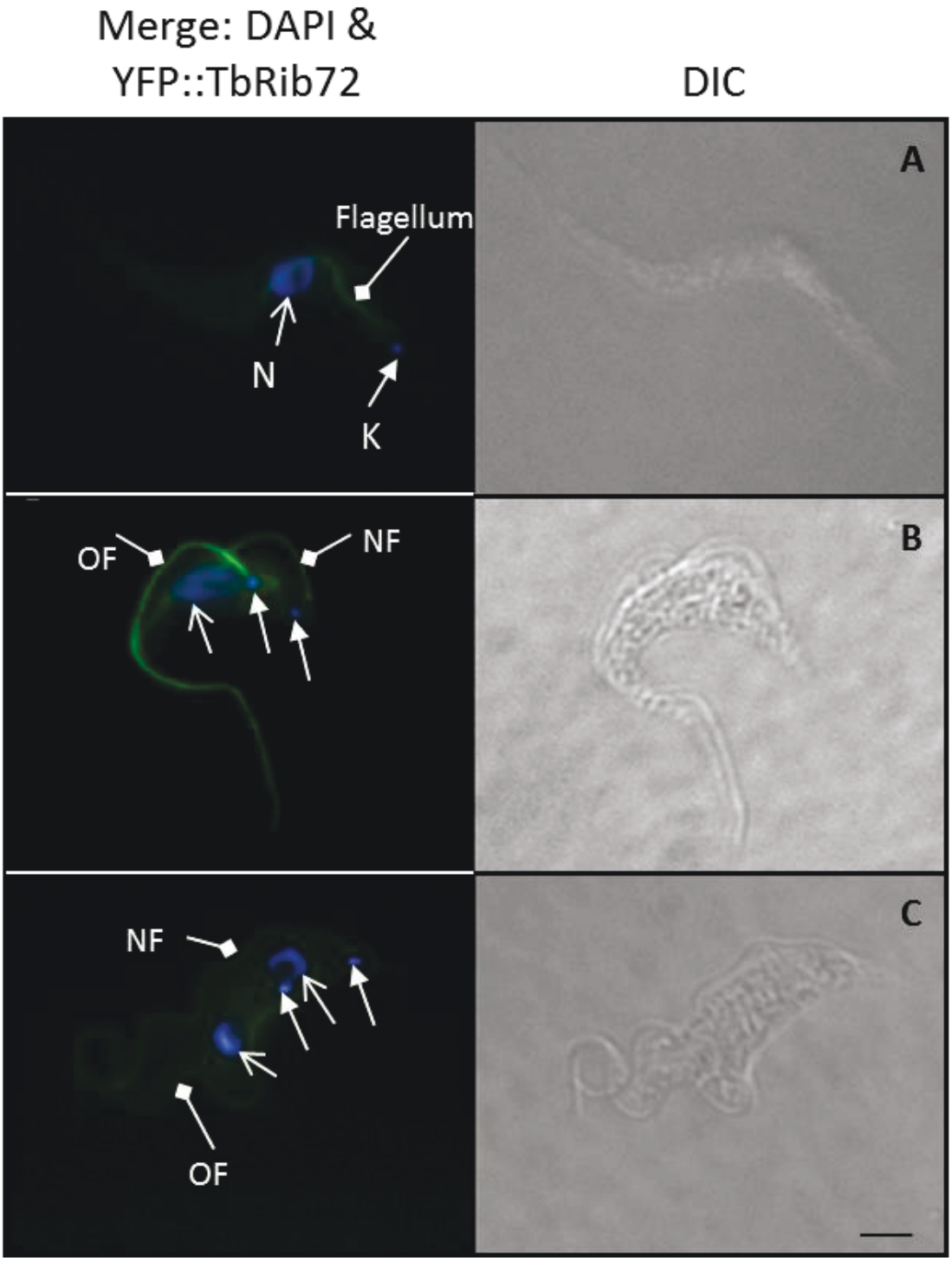


Figure 31: Exogenous expression of YFP::TbRib72 localises to the flagellum

TbRib722 was ectopically expressed with YFP fused to the N terminus (YFP::TbRib72, green). Cells were fixed and labelled with DAPI to stain DNA. The YFP::TbRib72 signal was localised to the flagellum in 1K1N (A), 2K1N (B) and 2K2N (C) cells. NF; new flagellum, OF; old flagellum. Images were captured on a Zeiss axioplan. Scale bar = 2µm.

3.3.6 Bioinformatics analysis of TbDRC5

Here I show that the *Trypanosoma brucei* protein, Tb927.5.2270, is an ortholog of human T-complex-associated testes expressed (TCTE1). Tb927.5.2270 was identified in the original list of 26 proteins (Table 14, candidate 7), it is in the TbCMF (as TbCMF44) (Baron *et al.*, 2007b) and in the TbFP (Broadhead *et al.*, 2006). The *Chlamydomonas reinhardtii* ortholog of TCTE1 is DRC5/FAP155, which was detected in the CrFP (Pazour *et al.*, 2005).

Orthologous DRC5 sequences were established through a reciprocal best blast hit (RBH) approach. An alignment of the DRC5 protein sequences from *T. brucei*, *H. sapiens* and *C. reinhardtii* show a high level of conservation (Figure 32; A). The DRC5 sequence of *T. brucei* shares 34.9% amino acid identity with the *H. sapiens* sequence and 55.3% similarity. *C. reinhardtii* FAP155 is 32.1% identical and 53.2% similar at the amino acid level to *H. sapiens.* Protein sequences were entered into the search function of the conserved domain database (CDD) (Marchler-Bauer *et al.*, 2011), which revealed no specific domains were present. However, all DRC5 sequences contained a C terminal area of leucine rich repeats (Figure 32; B).

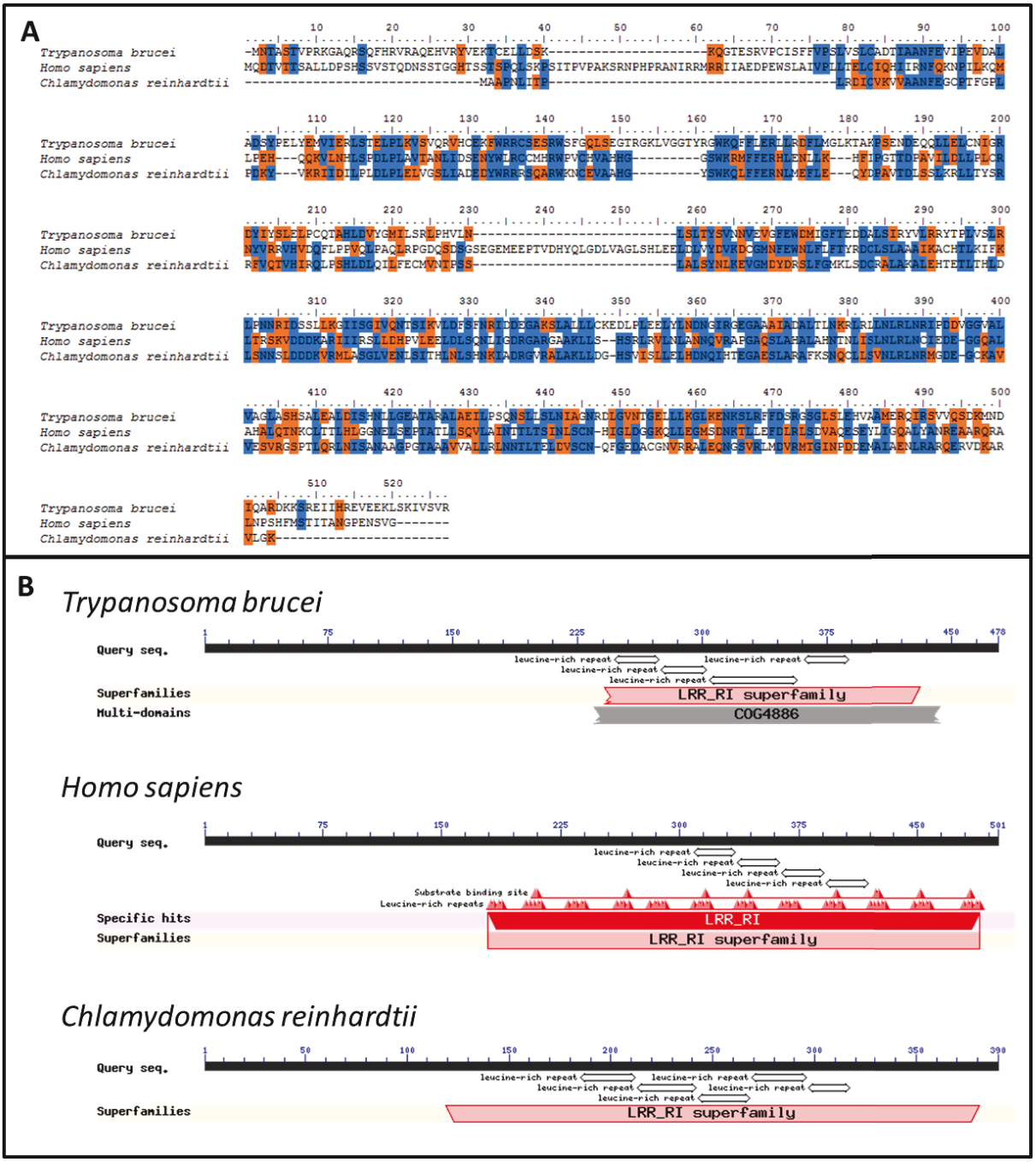


Figure 32: Assessment of DRC5 orthologs

(A) Alignment of the orthologous protein sequences from *Trypanosoma brucei* (XP_844901), *Homo sapiens* (NP_872345) and *Chlamydomonas reinhardtii* (XP_001689742). Identical amino acids are blue and similar amino acids are orange. (B) Schematic of Tb927.5.2270 protein, HsTCTE1 and CrFAP155 showing the position of conserved domains/repeats within the protein (NCBI). Protein diagrams are not to scale.

3.3.7 Localisation of TbDRC5

TbDRC5 was localised using an endogenous YFP tag at the N terminus, which was achieved using the pPOT system (Dean *et al.*, 2015) to target Tb927.5.2270. See section 2.3.1 Primer design and plasmid construction. Homologous recombination of the PCR product modified one allele so expression of YFP::TbDRC5 was constitutively active from one endogenous locus.

YFP::TbDRC5 localised to the flagellum of *T. brucei* 1K1N cells (Figure 33; A). The YFP::TbDRC5 signal was visible on the new flagellum and old flagellum of cells with 2 flagella (Figure 33; B, 2K1N and C, 2K2N). This implies that YFP::TbDRC5 was stably incorporated into the flagellum as it was assembled and does not have cell cycle variable expression. The area of the 1K1N cell (Figure 33; A) marked out by the dotted line is displayed in the inset panels (Figure 33; D –G). The phase contrast panel (Figure 33; D) shows the proximal end of the flagellum and basal bodies (BB). The native YFP::TbDRC5 of the same inset area is shown (Figure 33; E). The image of the native YFP::DRC5 signal was processed so that the contrast is absolute (a pixel is yellow or black) (Figure 33; F). When the absolute YFP image is merged with the phase contrast image (Figure 33; G) it is apparent that the YFP::DRC5 signal is localised to the axoneme and not to the basal bodies.

In summary, fluorescence microscopy of detergent extracted cytoskeletons showed that YFP::TbDRC5 localised to the flagellum of procyclic *T. brucei* (Figure 33). Therefore TbDRC5 is confirmed as a cytoskeletal protein in *T. brucei*. This was the first time that any member of the TCTE1 family has been localised at an endogenous expression level to a cilium or flagellum.

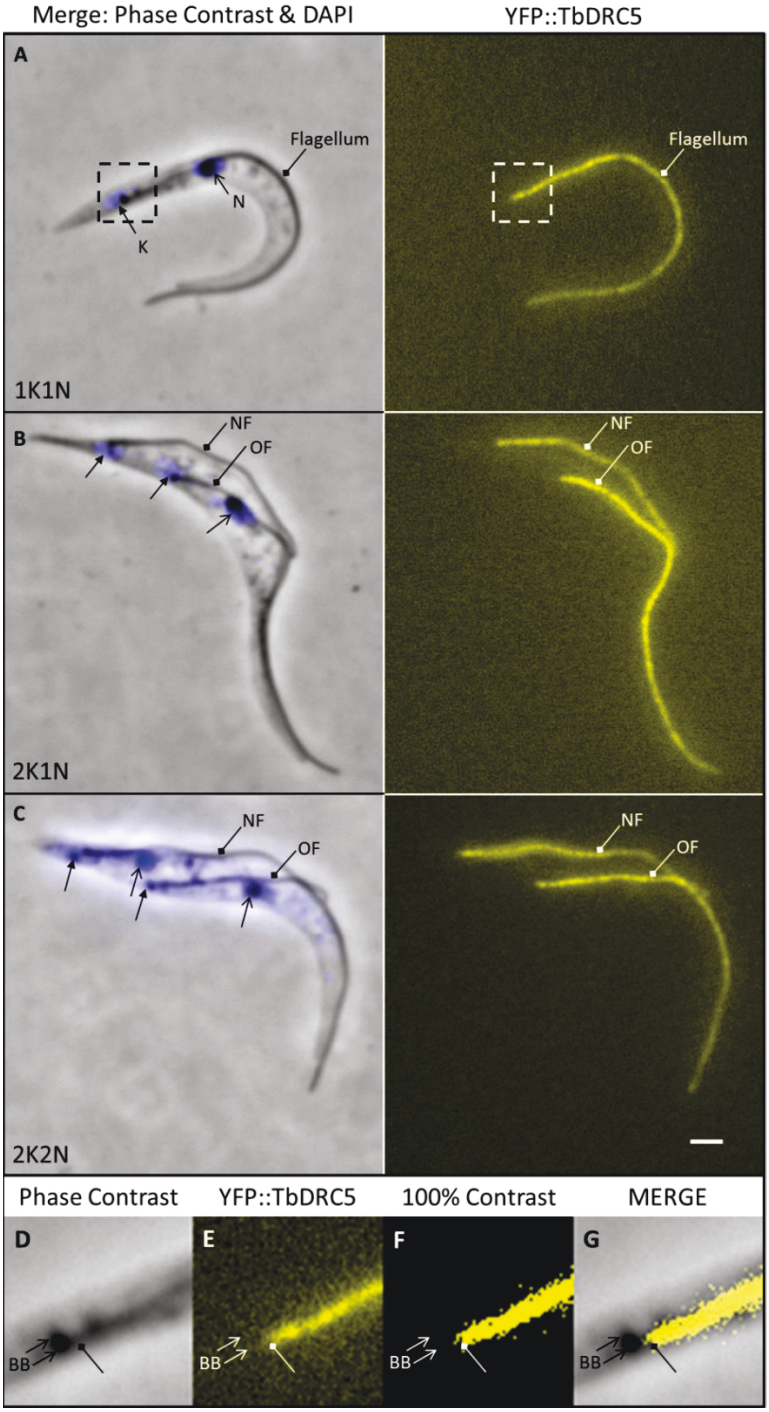


Figure 33: YFP::TbDRC5 localises to the flagellum throughout the cell cycle

TbDRC5 was tagged at the N terminus with YFP (YFP::TbDRC5; yellow) and expressed from the endogenous locus in procyclic *T. brucei*. Cells were detergent extracted fixed and labelled with DAPI to stain the DNA. The YFP::TbDRC5 signal can be seen localised to the flagellum in 1K1N (A), 2K1N (B) and 2K2N (C) cells. Insets (D - G) show the proximal end of the flagellum from the 1K1N cell (A). (D) Phase contrast image of the proximal end of the flagellum and basal bodies (BB, black open head arrows). (E) Native YFP::TbDRC5 signal, which is localised to the flagellum (diamond head arrow) but not the basal bodies (BB, open head arrows). (F) Pixels of YFP::TbDRC5 signal are made absolute by setting contrast to 100%. (G) Phase contrast and absolute YFP::TbDRC5 signal are overlaid. K; kinetoplast, N; nucleus, NF; new flagellum, OF; old flagellum, BB; basal bodies. Scale bar = 2µm.

3.3.8 Bioinformatics analysis of candidate 12: Tb10.61.0160

From the original list of 26 genes, Tb10.61.0160, (Table 14, candidate 12), is in the TbCMF (as TbCMF1) (Baron *et al.*, 2007b) but not in the TbFP (Broadhead *et al.*, 2006). Reciprocal BLASTp analysis was performed, using the Tb10.61.0160 protein sequence as the seed sequence. Orthologous sequences were established through reciprocal best blast hit (RBH). This revealed that the protein sequence was orthologous to the human protein WD-repeat containing protein 92 (WDR92) and contained a WD40 domain (PF00400). The Tb10.61.0160 protein shall be referred to as TbWDR92. An alignment of the WDR92 protein sequences from *T. brucei*, *H. sapiens* and *C. reinhardtii* show a high level of identity (Figure 34; A, blue residues) and all sequences contained WD40 repeat domains (Figure 34; B). The amino acid sequence of HsWDR92 was 58.1% identical to CrWDR92 and 43.9% identical to TbWDR92. In humans, WDR92 is also known as Monad, which is most highly expressed in the testis (Saeki *et al.*, 2006). Monad is also reported as being involved in regulating the apoptosis pathway (Itsuki *et al.*, 2008) but has not been localised at a subcellular level. WDR92 was detected in proteomic analysis of zebrafish gonads along with two other WD40 domain-containing proteins, WDR13 and WDR146 (Groh *et al.*, 2011).

The gene encoding HsWDR92 is mapped to 2p14 within the human genome. In the TbCMF, the human ortholog of Tb10.61.0160 was implicated in retinitis pigmentosa (RP), a potential ciliopathy (Baron *et al.*, 2007b); Table S2). However when I searched the OMIM database no link to RP was found at 2p14 but RP28 was annotated at 2p15 (#606068). I propose that the locus of RP28 was refined in the time period between the TbCMF and this study.

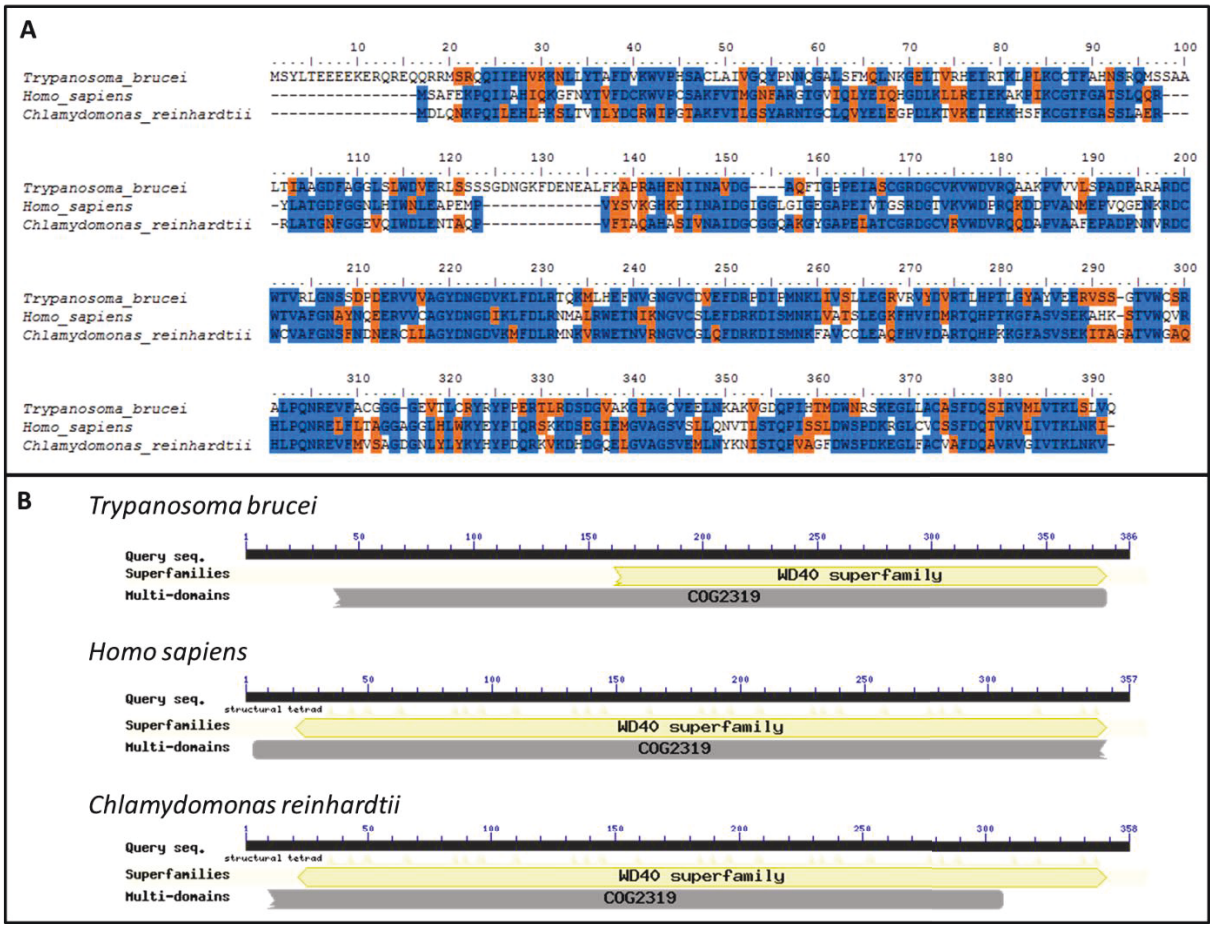


Figure 34: Assessment of WDR92 orthologs

(A) Alignment of the orthologous protein sequences from *Trypanosoma brucei* (XP_828080), *Homo sapiens* (NP_612467) and *Chlamydomonas reinhardtii* (XP_001695914). Identical amino acids are blue and similar amino acids are orange. (B) Schematic of Tb10.61.0160 protein, HsWDR92 and CrWDR92 showing the position of the WD40 repeats (superfamily) within the proteins (NCBI). Protein diagrams are not to scale.

3.3.9 Localisation of candidate 12: Tb10.61.0160

To study the localisation of TbWDR92, a long primer PCR based method for gene tagging known as pPOT (Dean *et al.*, 2015), was used to generate a stable procyclic cell line that constitutively expressed a N terminus YFP fusion protein, YFP::TbWDR92 from one endogenous locus. See section 2.3.1 Primer design and plasmid construction. This showed that YFP::TbWDR92 did not localise to the flagellum in *T. brucei* procyclic cells (Figure 35). In light of this negative flagellum localisation result, the occurrence of WDR92 orthologs in cilia/flagella proteomes was assessed. Tb10.61.0160 was not detected in a proteomic study of the *T. brucei* flagellum (Broadhead *et al.*, 2006). WDR92 orthologs from rat (NP_001121051) and pig (XP_003125123) were not present in their respective sensory cilia proteomes (Mayer *et al.*, 2009; Narita *et al.*, 2012). The human ortholog was not detected in a motile cilia proteome (Ostrowski *et al.*, 2002) or proteomic analysis of centrosomes (Andersen *et al.*, 2003; Jakobsen *et al.*, 2011). The *C. reinhardtii* ortholog was also not in the CrFP (Pazour *et al.*, 2005). Deeper bioinformatics analysis revealed that the proteomes of the amoebae *Dictyostelium discoideum* and *Acanthamoeba castellanii* contained divergent orthologs of WDR92. These orthologs were not detected by RBH when the *T. brucei* protein sequence was used as the seed sequences. However when the WDR92 sequence from *Naegleria fowlerei*, an ameboflagellate, was entered into a BLASTp search the result list contained orthologs from amoebae. Neither *D. discoideum* nor *A. castellanii* build flagella during their lifecycles, which is evidence that WDR92 is not a flagellum protein. In summary, Tb10.61.0160 has been established through bioinformatics as an ortholog of WDR92. Despite TbWDR92 (Tb10.61.0160) being listed in the TbFP, it was found that it does not localise to the flagellum in *T. brucei*. The conservation of WDR92 orthologs in organisms that do not assemble motile flagella indicated that WDR92 is not a flagellum protein.

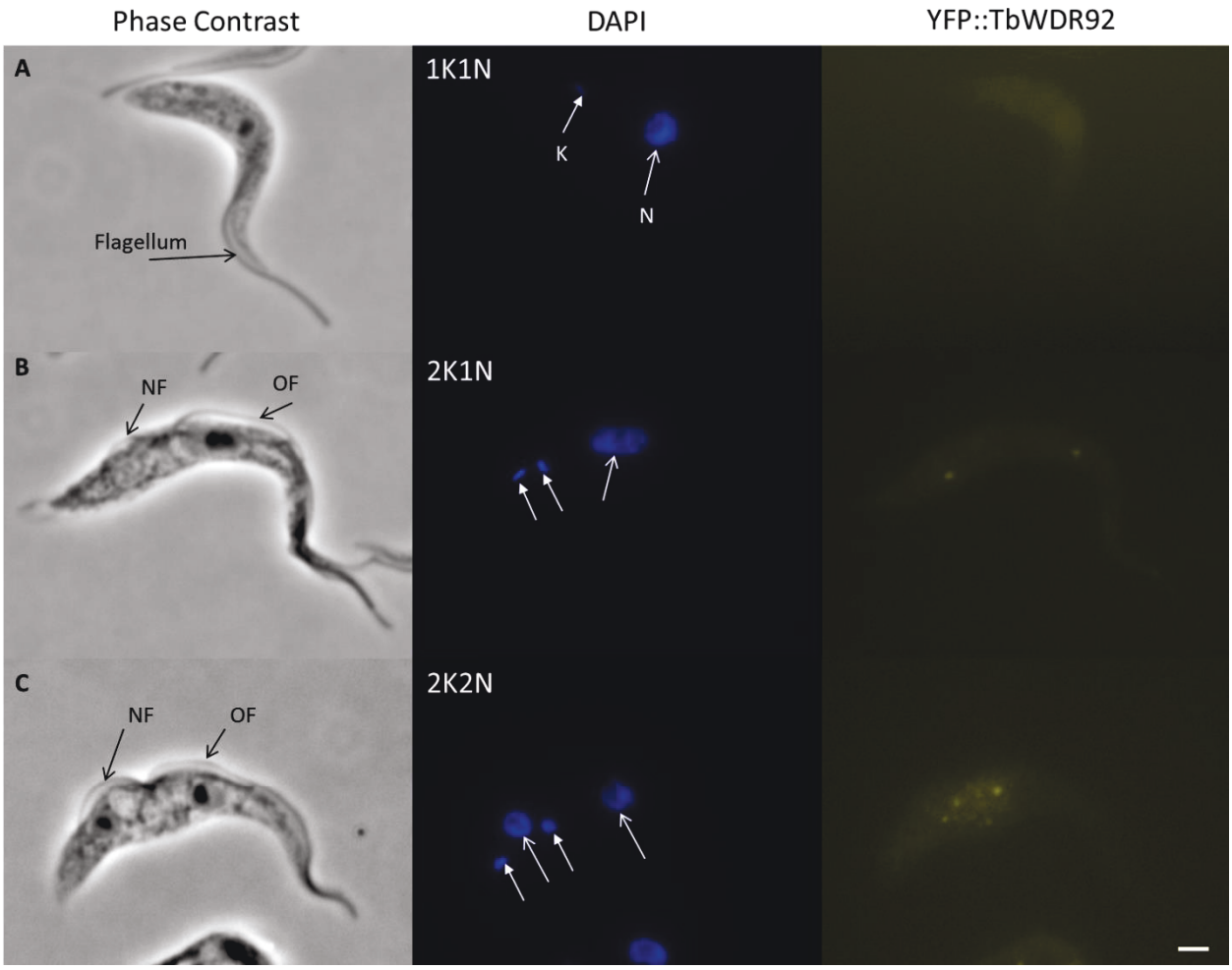


Figure 35: Localisation of YFP::TbWDR92

TbWDR92 was tagged at the N terminus with YFP (YFP::TbWDR92; yellow) and expressed from the endogenous locus in procyclic *T. brucei*. Cells were fixed and labelled with DAPI to stain DNA. The YFP::TbWDR92 signal was not localised to the flagellum in 1K1N (A), 2K1N (B) and 2K2N (C) cells. NF; new flagellum, OF; old flagellum. Scale bar = 2µm

3.4 Discussion

The work outlined in this chapter has shown how a list of 26 potential flagellum proteins was decided. Bioinformatics analysis of orthology has been presented for 4 of these candidates plus an additional protein, Tb927.7.4100/TbCCDC113. Localisation data for these 5 proteins has also been presented. The impact of these results in the context of literature is discussed below.

3.4.1 Localisation of CCDC96 and CCDC113

As shown in this work, when TbCCDC96 and TbCCDC113 were expressed under their endogenous promoter in *T. brucei*, they both localised to the flagellum. This is the first time that an ortholog of CCDC96 or CCDC113 has been shown to localise to a cilium/flagellum. This result contradicts work in two human cell lines; HeLa and RPE1 (Firat-Karalar *et al.*, 2014) where overexpression of both proteins resulted in localisation to the centrosome and not to the induced cilium of serum starved cells. CCDC96 and CCDC113 are both conserved in *H. sapiens* but neither protein was detected during an extensive proteomic interrogation of the centrosome (Jakobsen *et al.*, 2011). CCDC96 and CCDC113 orthologs have not been detected in any basal body/centriole or centrosome proteomes published (Andersen *et al.*, 2003; Balestra *et al.*, 2013; Jakobsen *et al.*, 2011; Keller *et al.*, 2005) see Table 15 for summary of all proteomes.

Recently work on another coiled-coil protein was published that reported a difference in localisation dependant on cell type (Narasimhan *et al.*, 2014). In Zebrafish, CCDC11 localised to the centriole of cilia from ciliated epithelium in Kupffer's vesicle but in the kidney of the zebrafish, CCDC11 localised to the cilia axonemes (Narasimhan *et al.*, 2014). CCDC113 and CCDC96 could be the second and third examples of a centriole/ cilia protein that has a different function and localisation according to cell type. It is also important to consider that the localisation of HsCCDC96 and HsCCDC113 to the centriole was achieved using an overexpression system and not by endogenous expression or immunostaining (Firat-Karalar *et al.*, 2014).

The flagellum localisation of both CCDC96 and CCDC113 in *T. brucei* is interesting because recently analysis of bovine sperm identified both CCDC96 and CCDC113 as components of

the centriole (Firat-Karalar *et al.*, 2014) and not of the sperm flagellum. Although this is different from the localisation presented in this work, it is in line with CCDC96 and CCDC113 not being detected in the human ciliome, their localisation may be restricted to the centriole in a Chordata specific role.

3.4.2 Rib72 orthologs

Within doublet and triplet microtubules, in addition to α and β tubulin heterodimer protofilaments, there are protofilaments comprising of specialised ribbon proteins (Figure 2). The additional ribbon proteins are a component of the outer doublet and triplet microtubules in the axoneme and basal body (Steffen and Linck, 1988). Although the exact position within the microtubule is debated (Linck *et al.*, 2014).

The first ribbon proteins identified are called tektins (Linck *et al.*, 1985; Steffen and Linck, 1988). There are three tektin proteins in *H. sapiens*; A, B and C. Tektins are not conserved in *T. brucei* (Berriman *et al.*, 2005) but *C. reindardtii* has a single tektin protein (Yanagisawa and Kamiya, 2004), which was detected in the basal body proteome (Keller *et al.*, 2005) and the flagella proteome (Pazour *et al.*, 2005). Another ribbon protein is Rib43a (42.6KDa), which was first characterised in *C. reinhardtii* (Norrander *et al.*, 2000). Rib43a is conserved in *T. brucei* and although it has not been directly localised to the flagellum, presumably is a flagellum protein due to its detection in the TbFP (Broadhead *et al.*, 2006) and TbCMF (TbCMF19; Baron *et al*., 2007).

Rib72 was first characterised in *C. reinhardtii*, which has Rib72 and Rib72-related proteins (Ikeda *et al.*, 2003; Keller *et al.*, 2005). Two axonemal ribbon proteins from the sea urchin *Strongylocentrotus purpuratus* were later realised to be orthologs of Rib72, despite having different molecular weights (Hinchcliffe and Linck, 1998). Rib72 proteins are also conserved in *M. musculus, H. sapiens* and *D. melanogaster* and are known as EF-hand containing proteins 1 and 2 (EFHC1 and EFHC2) (Gu *et al.*, 2005; Ikeda *et al.*, 2005; Rossetto *et al.*, 2011; Suzuki *et al.*, 2004).. It was thought that *C. elegans,* which only forms immotile cilia, did not possess an ortholog of Rib72 (Li *et al.*, 2004), but it does (Linck *et al.*, 2014), although the *C. elegans* ortholog of Rib72 has two DUF1126 domains instead of three.

The Rib72 family of proteins is highly conserved in eukaryotes and appears ubiquitous in organisms that form a motile flagellum. Mouse EFHC1 has previously been localised to the flagellum of sperm and tracheal cilia (Ikeda *et al.*, 2005). CrRib72 is localised to both flagellar by light microscopy and by immuno-electron microscopy CrRib72 was found to be a component of protofilament ribbons (Ikeda *et al.*, 2003). Post deflagellation, Rib72 mRNA levels increased 5-fold (Ikeda *et al.*, 2003) and therefore CrRib72 is required for flagellum assembly in *C. reinhardtii.*

Table 16: Conservation of ribbon proteins in eukaryotes.

Table to summarise conservation of Rib72 and Rib43 proteins in eukaryotes. Predicted molecular weight listed in NCBI is shown in brackets. Cr, *Chlamydomonas reinhardtii*; Sp, *Strongylocentrotus purpuratus*; Mm, *Mus musculus*; Hs, *Homo sapiens*; Tb, *Trypanosoma brucei*; Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*. Table adapted from (Linck *et al.*, 2014). ♠ (Ikeda *et al.*, 2003), ‡ (Keller *et al.*, 2005), □ (Norrander *et al.*, 2000), ▪ (Hinchcliffe and Linck, 1998), ○ (Linck *et al.*, 2014), ⊁ (Ikeda *et al.*, 2005), † (Gu *et al.*, 2005), ¤ (Suzuki *et al.*, 2004), ● (Rossetto *et al.*, 2011).

Protein class	**Organism**						
	Cr	**Sp**	**Mm**	**Hs**	**Tb**	**Ce**	**Dm**
Rib72	CrRib72 ♠ (71.9KDa)	SpRib74 ▪ (73.9 kDa)	EFHC1 ⊁ (74.9KDa)	EFHC1-1 ¤ (73.9KDa)	Tb927.11.1430 (86.7KDa)	Y49A10A (52.6KDa)	Defhc1.1 ● (90.8KDa)
				EFHC1-2 ¤ (71.9KDa)	Tb927.5.2950 (88.1KDa)		Defhc1.2 ● (88KDa)
		SpRib85.5 ▪ (85.5 kDa)	EFHC2 (87.5KDa)	EFHC2 † (87.3KDa)	Tb927.10.7690 (82.7KDa)		
	CrRib72-related ‡ (78.7KDa)						
Rib43	CrRib43a □ (42.6 kDa)	SpRib45 ○ (44.9 kDa)	Rib43a-like (45.1 kDa)	Rib43a-like (45.3 kDa)	Tb927.8.4640 (46.8KDa)		CG7264 (45 kDa)

CrRib72 has been localised to the flagella by light and immune-electron microscopy (Ikeda *et al.*, 2003). In humans the gene for EFHC1 has been mapped to 6p12 (Suzuki *et al.*, 2002). In mice EFHC1 localises to tracheal cilia, sperm flagellar (Ikeda *et al.*, 2005) and the mitotic spindle (de Nijs *et al.*, 2006). Mutations in the gene encoding the EFHC1 protein have been reported to cause a form of epilepsy known as juvenile myoclonic epilepsy (JME) in humans (Medina *et al.*, 2008; Suzuki *et al.*, 2004; Suzuki *et al.*, 2002). JME is an inheritable form of epilepsy thought to be caused by abnormal regulation of apoptosis

(Suzuki *et al.*, 2004) although some evidence suggests that EFHC1 may not be directly involved (Bai *et al.*, 2009). The second EF-hand containing protein is EFHC2, which has been mapped to the locus Xp11 in the human genome (Gu *et al.*, 2005). Despite high sequence identity (Figure 31) and shared arrangement of DUF1126 domains with EFHC1, EFHC2 has not yet been localised to cilia/flagella. When the aligned EFHC1 and EFHC2 sequences are assembled into a phylogenetic tree the two distinct clades are clear (Figure 106).

There is an antibody that recognises one of the TbRib72 orthologs, although it is not published which one it recognises (or indeed if it recognises all three) but it does label the axoneme and not the basal bodies (Nett *et al.*, 2009). It is not clear if the three Rib72 orthologs have neofunctionalised.

In the TbCMF, knockdown by RNAi was performed on the 50 identified proteins. CMF mutant phenotypes were visually scored, by low magnification light microscopy, on a scale of 1 to 4 based on the degree of cell clumping in culture. Class 1 mutants were not affected by protein ablation and class 4 mutants were 'severely' affected with large clumps of cells in the culture. RNAi of the three TbRib72 orthologs (TbCMF 2, 3 and 4) revealed that Tb927.11.1430 (TbCMF2) and Tb927.5.2950 (TbCMF3) were both class 2 phenotypes with 'mild' clumping. Tb927.10.7690 (TbCMF4) had a class 3 phenotype, which is 'moderate to large clumps' of cells in the culture. Analysis of cell motility for TbCMF2 and TbCMF3 found that post knockdown 22% of cells in the TbCMF2 culture were immotile and 17% in the TbCMF3 (Baron *et al.*, 2007b). Induced cell lines were not viewed by high magnification light microscopy so it is not clear if cells were able to assemble a flagellum. TbCMF3 also showed 'mild sedimentation' in a sedimentation assay although the length of time the RNAi against TbCMF3 was induced for before the assay was performed is not stated (Baron *et al.*, 2007b). Only class 4 mutants were examined by electron microscopy so it is not known if knockdown of any TbRib72 orthologs have an effect on axoneme ultrastructure. In the TbCMF study only TbCMF 9, 19, 40 and 46 were localised at a subcellular level using a GFP fusion approach (Baron *et al.*, 2007b). Therefore this work is the first time Tb927.11.1430 (TbCMF2) has been localised to the flagellum in *T. brucei.*

3.4.3 The dynein regulatory complex in *T. brucei*

It has been shown that Tb927.5.2270 (TbDRC5) is an ortholog of *H. sapiens* TCTE1. TCTE1 was first identified during linkage mapping of mouse chromosome 17 (Sarvetnick *et al.*, 1989). TCTE1 was thought to be specific to vertebrate sperm but bioinformatics studies have identified orthologs conserved in other eukaryotes (Baron *et al.*, 2007b). The *C. reinhardtii* ortholog of TCTE1, FAP155, was present in the CrFP (Pazour *et al.*, 2005). FAP155 has been identified as a component of the nexin-dynein regulatory complex (N-DRC) (Lin *et al.*, 2011) and renamed CrDRC5.

Work in *C. reinhardtii* suggests that the *sup-pf-4* mutant phenotype (Piperno *et al.*, 1994) is caused by loss of functional CrDRC5, a conclusion reached by examining spots on 2-dimensional SDS-PAGE protein gels (Lin *et al.*, 2011). Using 1 and 2 dimensional protein gel electrophoresis it is established that the *C. reinhardtii* mutant, sup-pf-4 (Piperno *et al.*, 1994) is a result of a mutation generated in the gene encoding FAP155 but no ultra-structural analysis of the mutant strain flagella has been published to date. More recent work in *C reinhardtii* has identified FAP155 as part of the nexin-dynein regulatory complex (Lin *et al.*, 2011) and therefore renamed DRC5 but no specific function has been formally assigned. However, no rescue experiment has been published to formally prove this is the case. The *sup-pf-4* mutant exhibits a normal flagellum beat but 'slightly slower' motility (Lin *et al.*, 2011). Presumably, the mild phenotype is related to the fact that DRC5 is at the distal lobe of the N-DRC and if a more proximal component such as DRC2 or DRC4 were absent, the phenotype would be more severe.

Although work has been carried out on DRC4 (trypanin) (Hutchings *et al.*, 2002; Ralston and Hill, 2006) and DRC2 (Kabututu *et al.*, 2010) in *T. brucei*, a lot of information on the N-DRC of *T. brucei* is based upon the *C. reinhardtii* model (Figure 36; A). In *C. reinhardtii* DRC6 (FAP169) recruits DRC5 (FAP155) (Figure 36; A, arrows). DRC5 is conserved within *T. brucei* (Figure 36; B, arrows) but conservation of DRC6 is restricted to *C. reinhardtii* and *Volvox carteri.* Therefore DRC5 is the most distal component of the dynein regulatory complex in the axoneme of *T. brucei.* This work is the first time that an ortholog of DRC5 has been localised to the flagellum by light microscopy. TbDRC5 was confirmed as a cytoskeletal, flagellum protein (Figure 33) but this work does not prove that TbDRC5 is a

component of the dynein regulatory complex in T. *brucei.* Knockdown of a more proximal DRC component(s) such as DRC3, DRC4 or DRC7 (Figure 36; B, circle) would be necessary to test if DRC5 is still recruited.

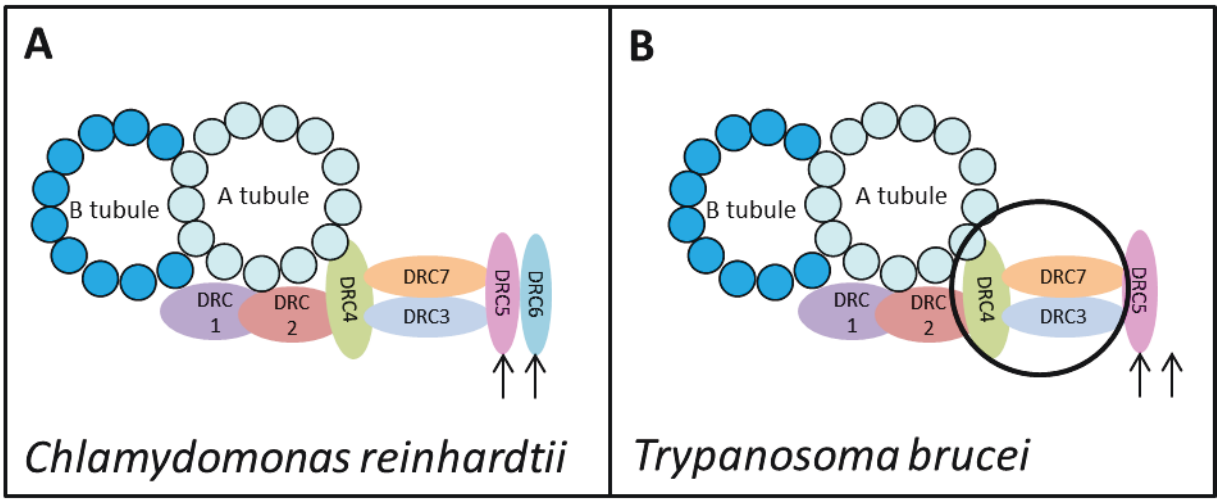


Figure 36: A model for DRC5 function

(A) Cartoon representation of the *C. reinhardtii* N-DRC based on published model (Lin *et al.*, 2011). (B) Cartoon model of what the T. brucei N-DRC may look like based on the *C. reinhardtii* model and known conservation of components. Only one outer doublet microtubule is portrayed with the N-DRC. IDA, ODA and radial spokes are not included.

Using T. *brucei* as a model organism to study the function of TCTE1/FAP155/DRC5 is clinical relevant as TCTE1 has been putatively linked to several ciliopathies. The TCTE1 gene (Gene ID 202500, NCBI) is mapped to 6p21 of the human genome. When 6p21 was examined in the OMIM database, several ciliopathy diseases were putatively linked to that area. This included polycystic kidney disease (#263200), cone and rod dystrophy (#602093), primary ciliary dyskinesia (#612650) and spinocerebellar ataxia (#271250). Polycystic kidney disease (PKD) is caused by malfunction of cilia in the kidneys, which alters the dynamics of liquid flow and ultimately leads to the formation of large cysts, creating enlarged and distorted kidneys and in some cases affects the liver. An inherited disease causing cysts on patients kidneys has been in the medical literature for many years (Adams *et al.*, 1974) but only relatively recently has the cause been linked to cilia (Haycraft *et al.*, 2001; Pazour *et al.*, 2000). Cone and rod dystrophy (CORD) is in the class of retinal ciliopathies (for review see (Wheway *et al.*, 2014)). The connecting cilium in photoreceptor cells (cone and rod cells) is essential for transport of rhodopsin and opsin

to and from the main cell body to the outer segment via IFT. Dysfunction of the connecting cilium affects the high turnover transport necessary for sight. Primary ciliary dyskinesia (PCD) is a ciliopathy condition where the motile cilia of the patient are missing ultrastructural features such as dynein arms or radial spokes. An absence of dynein arms or radial spokes renders the cilia immotile or unsynchronised and therefore unable to function properly. The form of PCD linked to 6p21 is known to be caused by a mutation in the gene that encodes the protein radial spoke head 9 (Castleman *et al.*, 2009). Spinocerebellar ataxia (SCA) is not a ciliopathy independently but the form of SCA caused by a mutation in the region of 6p21 leads to a type of SCA with associated deafness and blindness, which is indicative of a ciliopathy condition. Other forms of SCA are now also being connected to cilia malfunction (Goetz *et al.*, 2012).

3.4.4 TbWDR92 is not a component of motile flagella

The reason for including Tb10.61.0160 in the original list of 26 candidates (Table 14) was its inclusion in the TbCMF (Baron *et al.*, 2007b). The TbCMF was an *in silico* screen to identify potential components of motile flagella. The organisms used as negative controls in the TbCMF study were the land plant *Arabidopsis thaliana* and the nematode *Caenorhabditis elegans*, which do not have motile cilia or flagella. *A. thaliana* and *C. elegans* do not contain orthologs of WDR92. Therefore, due to the elected control organisms used in the TbCMF, WDR92 was incorrectly annotated as a potential component of motile flagella. For example, if *A. castellanii* had been selected as a negative control, WDR92 would not have been annotated as a component of motile flagella due to the fact that *A. castellanii* does possess an ortholog of WDR92. However, immunolabelling with a specific antibody against WDR92 would be necessary to confirm a non-flagellum localisation.

In this work I have shown, through bioinformatics, that WDR92 orthologs are highly conserved in eukaryotes, including those that do not assemble a motile flagellum. Unsurprisingly, in light of this conservation pattern, the *T. brucei* ortholog of WDR92 does not localise to the flagellum.

3.4.5 A brief review of other candidate proteins

WDR16 is known to cause hydrocephalus in the model organism *Danio rerio* (Hirschner et al., 2007). WDR16 is conserved in *T. brucei* and is also in my screen (Candidate 13; Table 1) (TbCMF15). TbWDR16 is known to strongly localise to the mature basal body and immature basal body with weak localisation to the flagellum (Farr and Gull, personal communication in (Hodges et al., 2010)). Knockdown of WDR16 in *D. rerio* did not affect cilia motility (Hirschner et al., 2007) but ultrastructure analysis of any differences to the flagellum or basal body was not performed.

3.4.6 Conclusions

In conclusion, the strategy adopted in this screen to identify uncharacterised flagellum proteins was successful. Any cases where the protein did not localise to the flagellum (candidate 12) have been rationally explained. This chapter has presented the bioinformatics and preliminary localisation data of 4 flagellum proteins, which were previously not localised within *T. brucei.* The function of these proteins was not addressed during the work performed in this preliminary screen.

4. Identification of TPH domain-containing proteins

4.1 Introduction

In humans, diseases resulting from the absence of cilia or malfunctioning cilia are collectively known as ciliopathies. There are a number of model organisms currently used to investigate the phenotypes and causes of these diseases. The work in this chapter has been performed *in silico* with a focus on *Trypanosoma brucei* as a suitable model organism for studying ciliopathies. *T. brucei* is a well characterised experimental model for cilia/flagella structure and function (Gull *et al.*, 1990; Sherwin and Gull, 1989).

This work provides the first evidence that there is a family of proteins that share a common TPH domain. From the work in this chapter I outline that the protein family consists of 7 subfamilies and investigate the degree of conservation across eukaryotes.
This chapter will focus on a set of proteins that I am proposing as a family. Candidate 1 (Table 14; accession Tb927.8.4580) is one member of this protein family. I performed bioinformatics analysis of Tb927.8.4580 to identify orthologs in eukaryotes and to identify similar non-orthologous proteins that are members of the larger protein family.

4.2 Aims

The research aims of this work are:

- To identify proteins in *T. brucei* that contain a TPH domain
- To investigate the conservation of TPH domain-containing proteins in eukaryotes

4.3 Results

4.3.1 Tb927.8.4580 is highly conserved.

The predicted protein sequence for candidate 1 (Table 14; Tb927.8.4580) was retrieved from the trypanosome genome database, www.TriTrypDB.org (Aslett *et al.*, 2010) and entered into a basic local alignment search tool (BLAST; NCBI) for proteins, BLASTp. The top result from the *H. sapiens* proteome was NP_036469, which was annotated as coiled-coil domain-containing protein 19/ Nasopharyngeal epithelium specific protein 1 (CCDC19/NESG1). CCDC19/NESG1 is not uniformly expressed in nasopharyngeal carcinomas of different grades (Liu *et al.*, 2011a). It is hypothesised that expression levels of CCDC19/NESG1 could be used to predict the prognosis of nasopharyngeal carcinoma patients (Liu *et al.*, 2011b). CCDC19/NESG1 is not localised at a sub-cellular level in any organism. HsCCDC19 (NP_036469) is mapped to position 1q22 in the human genome. When this locus was examined in the OMIM database there were several diseases linked to that area of chromosome one, including cone and rod dystrophy and retinitis pigmentosa, which are both known ciliopathy phenotypes (for review see (Tobin and Beales, 2009).

A reciprocal BLASTp within *T. brucei* using HsCCDC19 as the search query returned Tb927.8.4580 as the top hit. Therefore Tb927.8.4580 and HsCCDC19 are orthologous as determined by reciprocal best blast hit (RBH). The *H. sapiens* and *T. brucei* CCDC19 protein sequences were 29.4% identical and 56.2% similar. This can be demonstrated by aligning the sequences (Figure 37). Within the alignment, the amino acids coloured blue indicate that they are identical between the *H. sapiens* and *T. brucei* sequences, amino acids in orange are similar, which means that the substituted amino acids have the same properties and would therefore likely form the same protein tertiary structure. Due to sequence orthology Tb927.8.4580 shall now be referred to as TbCCDC19.

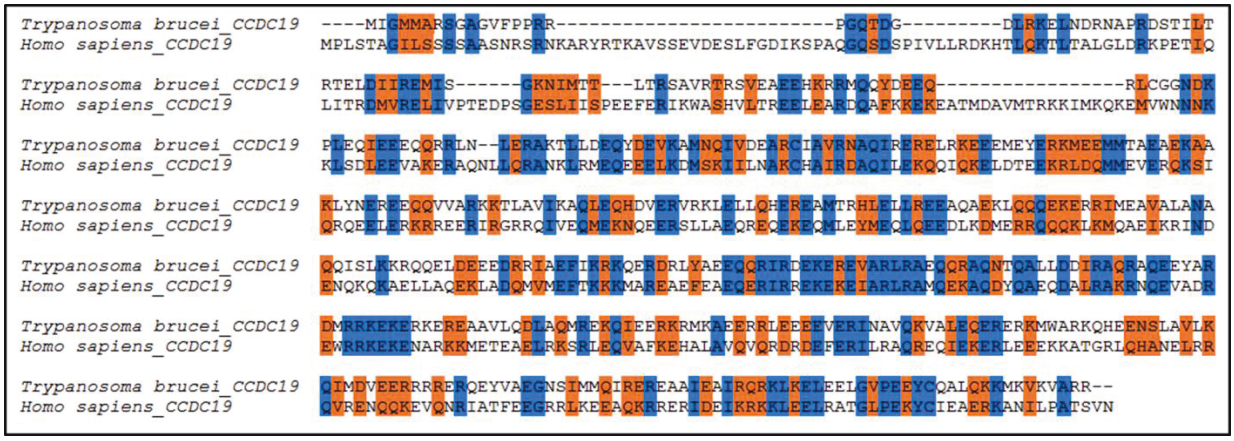


Figure 37: Bioinformatics of TbCCDC19

A) Alignment of the *T. brucei* (top row) and *H. sapiens* (bottom row) sequences for CCDC19 orthologs. Identical amino acids (29.4%) are highlighted in blue and similar amino acids (56.2%) are highlighted in orange.

Once it was established that CCDC19 was conserved between *H. sapiens* and *T. brucei*, orthologs were searched for across eukaryotes. For this study 39 eukaryotic organisms were selected (see section 2.2.1 Organisms used for conservation analysis) based on their known cilia and centriole morphology (Appendix 9), phylogenetic relationships to other organisms and the amount of genomic and proteomic information available.

RBH analysis within the pig, *Sus scrofa*, revealed two isoforms of CCDC19; isoform X1 XP_005663267 and isoform X2 XP_005663268, which are highly identical at the amino acid level with variation at the extremes of the N and C terminus (Appendix 3). The monkey, *M. mulatta,* also had two isoforms of CCDC19, isoform 1 XP_001115183 and isoform 2 XP_002801869, where the only difference is an extension of the N terminus of isoform 1 (or a partial deletion of the N terminus in isoform 2) (Appendix 3).

Sequences of identified orthologs were aligned using the ClustalW algorithm (Thompson *et al.*, 1994) (see section '2.2.6 Creating sequence alignments') in the BioEdit interface. The alignment contained 33 CCDC19 sequences from 31 organisms and showed that the sequences are highly conserved (Figure 38) with a divergent N terminus.

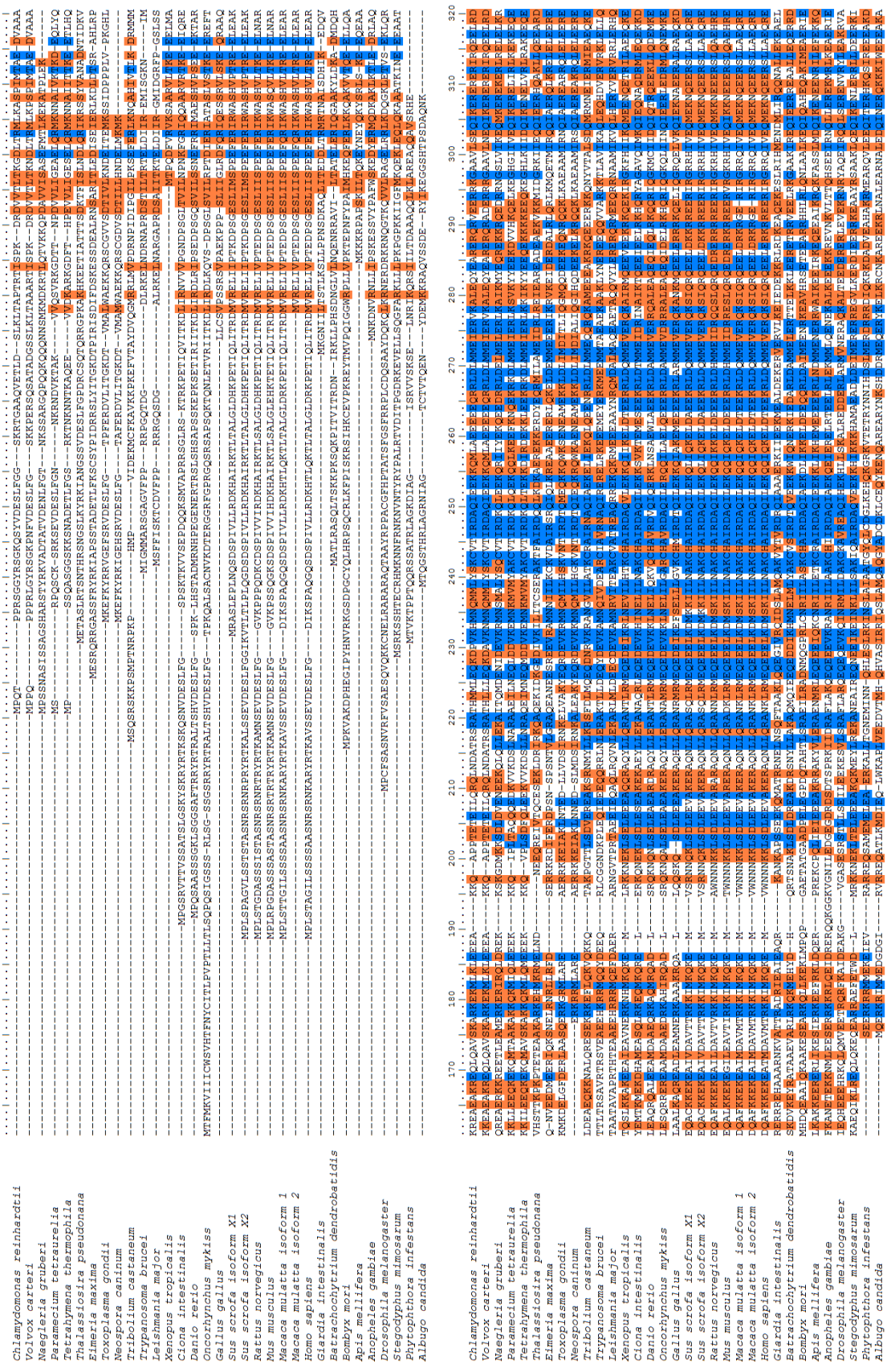


Figure 38 (part A.Figure legend on page 145)

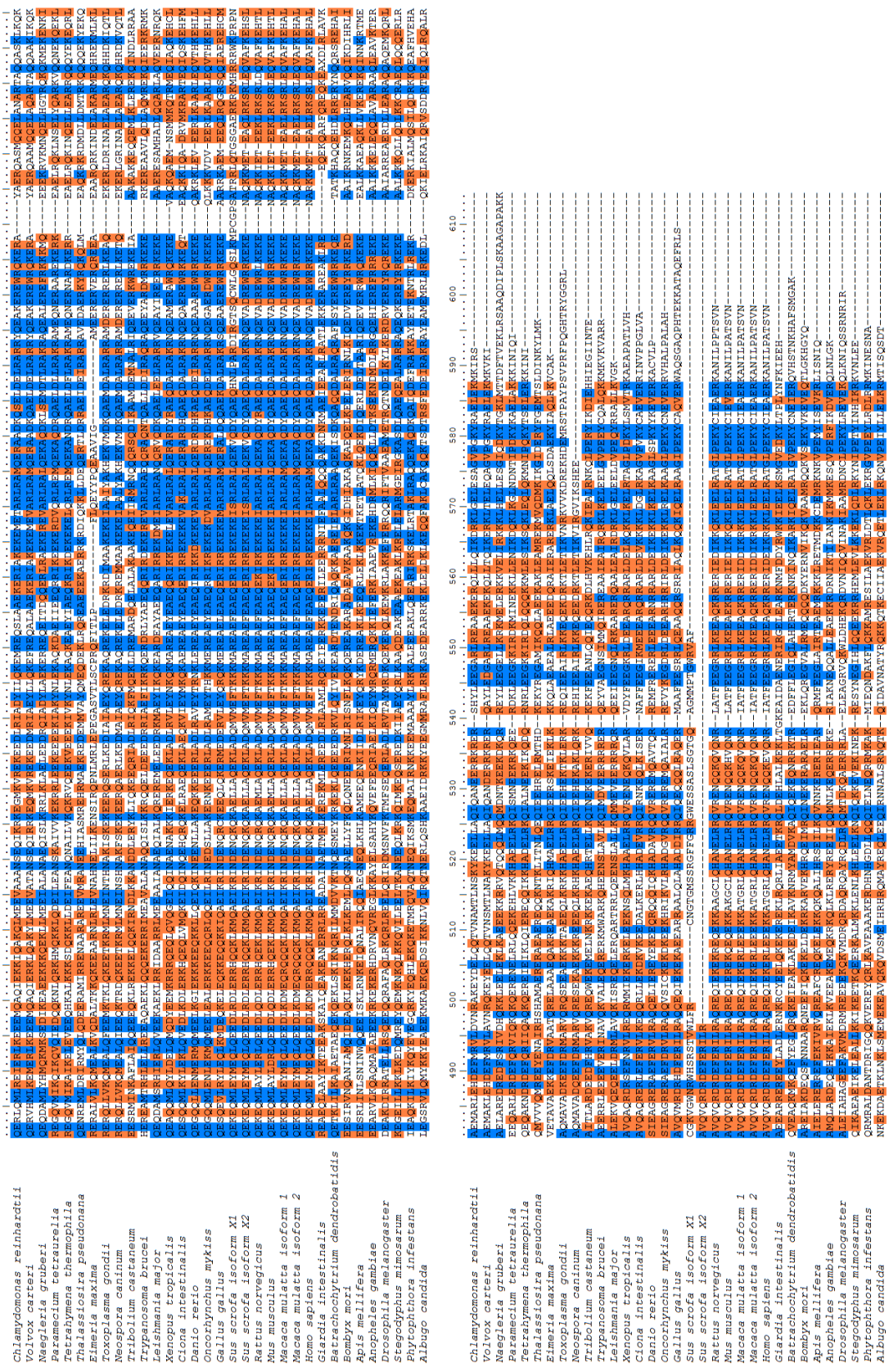


Figure 38: Alignment of eukaryotic CCDC19 orthologs.

Orthologous CCDC19 sequences were taken from across eukaryotes (n=33) to assess the level of amino acid conservation between species. Identical amino acids are coloured blue and similar amino acids are coloured orange.

The aligned sequences were assembled into a phylogenetic tree using SeaView (Gouy *et al.*, 2010) (Figure 39), which demonstrated the relationships between sequences and organisms. The phylogenetic tree showed that the two oomycete sequences, *Albugo laibachii* and *Phythopthora infestans* group together in a distinct clade, with 37.5% sequence identity between the two. The ciliates and apicomplexa clades branch from a common node, which was not surprising as this demonstrated the common lineage of these organisms under the super-phylum of aveolata (Cavalier-Smith, 1993; Stechmann and Cavalier-Smith, 2003). Sequences from other closely related species grouped as expected such as the sequences from chordata or insecta. Within the chordata clade the sequences for *H*. *sapiens* and *M. mulatta* grouped together with 98.2% identity (99.3% similarity) between the HsCCDC19 and MmCCDC19-1 protein sequences and 98.1% identity (99.4% similarity) between the HsCCDC19 and MmCCDC19-2 protein sequences.

No CCDC19 orthologs were detected using RBH (see section 2.2.2 Identifying protein orthologs') in *C. elegans* or *P. falciparum.* The nematode *C. elegans* possesses immotile sensory cilia but does not assemble motile cilia. *P. falciparum* does build a motile flagellum but it is internal to the cell within the cytoplasm and lacks IFT. CCDC19 was also not conserved in the yeasts *Schizosaccharomyces pombe* or *Saccharomyces cerevisiae*, which do not have microtubule centrioles, although CCDC19 was not ubiquitously absent from the fungi kingdom. A CCDC19 ortholog was detected by RBH in *Batrachochytrium dendrobatidis*, which is a fungus that has a flagellated zoospore stage (Longcore *et al.*, 1999).

In conclusion, CCDC19 is conserved in eukaryotes with a motile flagellum. The only exception is *P. falciparum,* which does not have a CCDC19 ortholog.

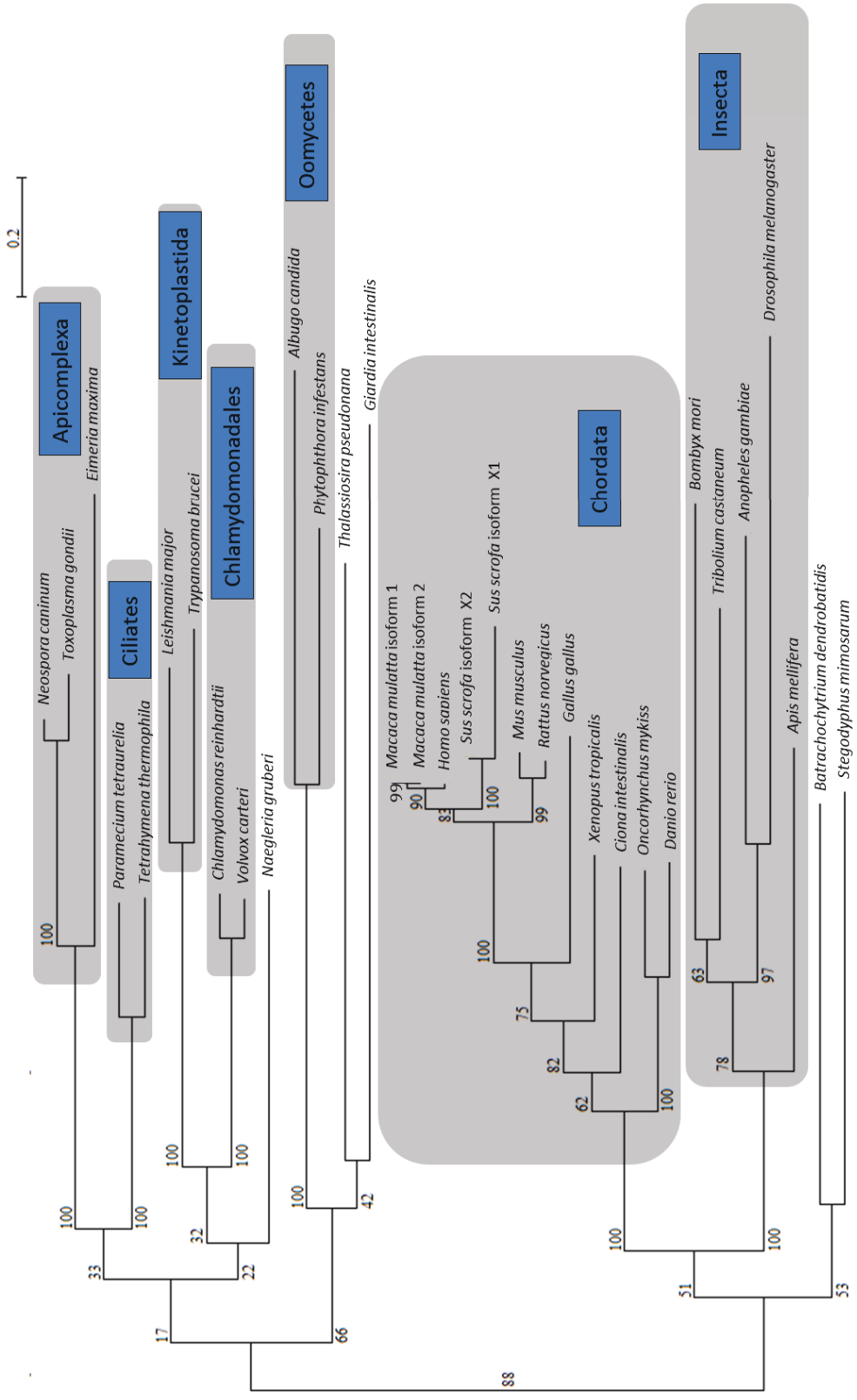


Figure 39: Phylogenetic tree for CCDC19

Maximum likelihood phylogenetic tree of aligned CCDC19 sequences from across eukaryotes. Aligned sequences were bootstrapped with 100 replicates, bootstrap values are displayed at the clade nodes. The scale bar indicates an evolutionary distance of 0.2 amino acid substitutions per position.

4.3.2 TbCCDC19 belongs to a protein family with a shared trichohyalin-plectin homology (TPH) domain

TbCCDC19 (Tb927.8.4580) was first identified in my bioinformatics screen (3.3.1 Identifying putative ciliopathy proteins) as 'candidate 1' (Table 14). To better predict a functional role for TbCCDC19 the protein family databases, Pfam and CDD were used to look at potential domains contained within the protein sequence. A domain that was detected in CCDC19 orthologs was the trichohyalin-plectin (TPH) domain (Pfam family accession: PF13868) (Figure 40; A). The TPH domain receives its name due to orthology to both mouse plectin 1 and human trichohyalin (Nishizawa *et al.*, 2005). Plectin and trichohyalin are both proteins that link cytoskeletal components such as microtubules and intermediate filaments within the cell cytoskeleton (Wiche, 1998).

It was mentioned in section 4.1 that this chapter would look at a family of proteins with a common domain, the TPH domain. TbCCDC19 is the first member of the TPH domain-containing protein family I identified in *T. brucei*. The species tree feature of Pfam was used to identify two additional *T. brucei* proteins that also contained a TPH domain (Figure 40; B and C). The two proteins were Tb927.5.1230 and Tb927.6.4520, which were found to be orthologs of coiled-coil domain-containing protein 11 (CCDC11) and meiosis nuclear specific protein 1 (MNS1), respectively. Protein sequences of HsCCDC11 and Tb927.5.1230 were 21.3% identical and 47.5% similar at the amino acid level; HsMNS1 and Tb927.6.4520 were 21.8% identical and 47.2% similar to each other at the amino acid level.

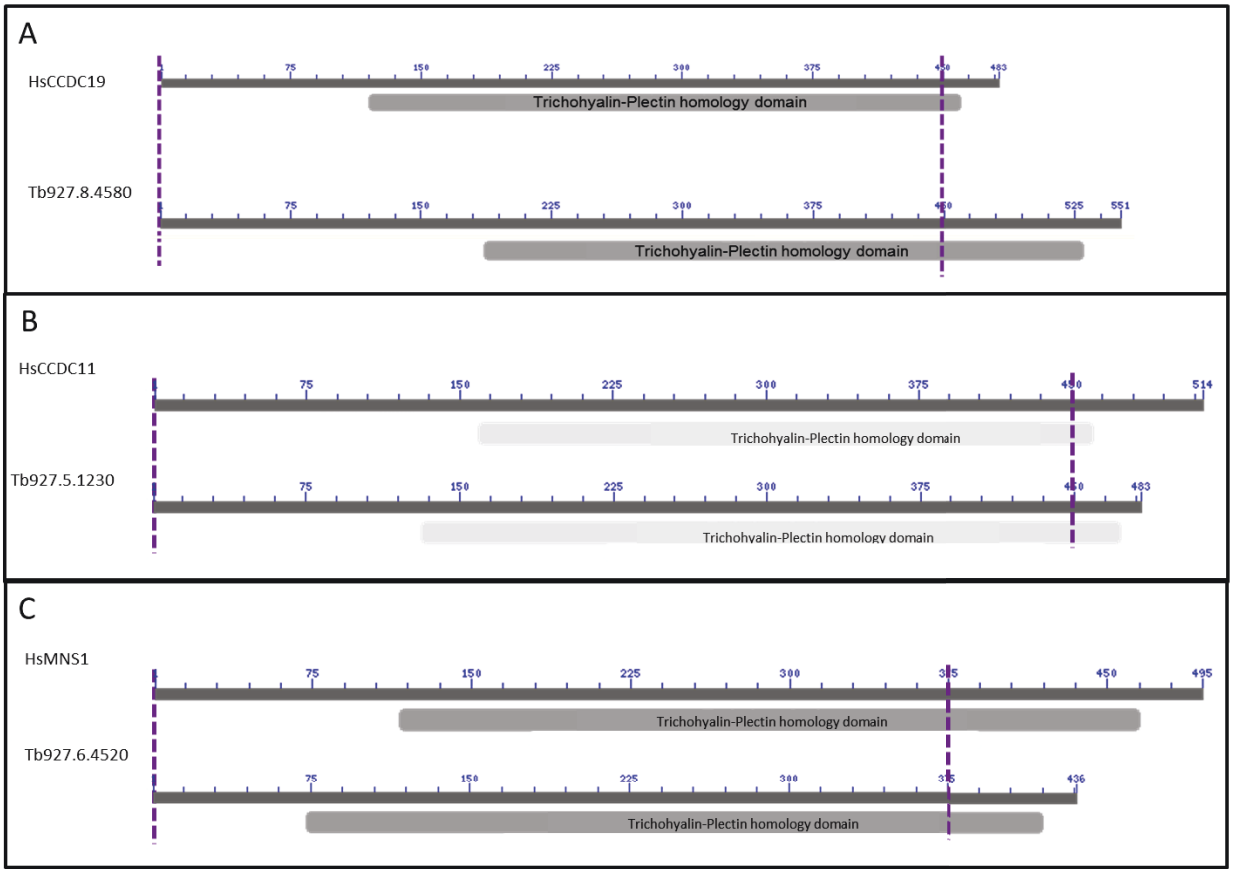


Figure 40: *T. brucei* TPH domain-containing proteins

Cartoons of (A) CCDC19, (B) CCDC11 and (C) MNS1 orthologs from *H. sapiens* and *T. brucei* showing the position of the TPH domains. Cartoons adapted from NCBI.

4.3.3 CCDC11 is highly conserved

The protein sequence of *T. brucei* Tb927.5.1230 was orthologous to HsCCDC11 in the human proteome by RBH analysis, with 21.3% sequence identity. I will therefore rename Tb927.5.1230 as TbCCDC11. To assess conservation of CCDC11 proteins I mined my chosen set of 39 eukaryotic organisms (see Table 9) for orthologs, I detected 24 CCDC11 orthologs using a RBH approach. Orthologous sequences were aligned (Figure 41) and assembled into a maximum likelihood phylogenetic tree (Figure 42). This showed that the sequences from organisms in the same phylum or class grouped together with no long branch outliers.

CCDC11 orthologs were not present in the 7 organisms that do not assemble a motile flagellum (see page 162). In the 32 flagellated organisms, bioinformatics analysis showed that CCDC11 was highly conserved but was not identified in *B. dendrobatidis, D. melanogaster, G. gallus, G. lamblia, N. gruberi, P. falciparum, S. mimosarum and T. pseudonana.*

In conclusion, CCDC11 is highly but not unanimously conserved among flagellated eukaryotes.

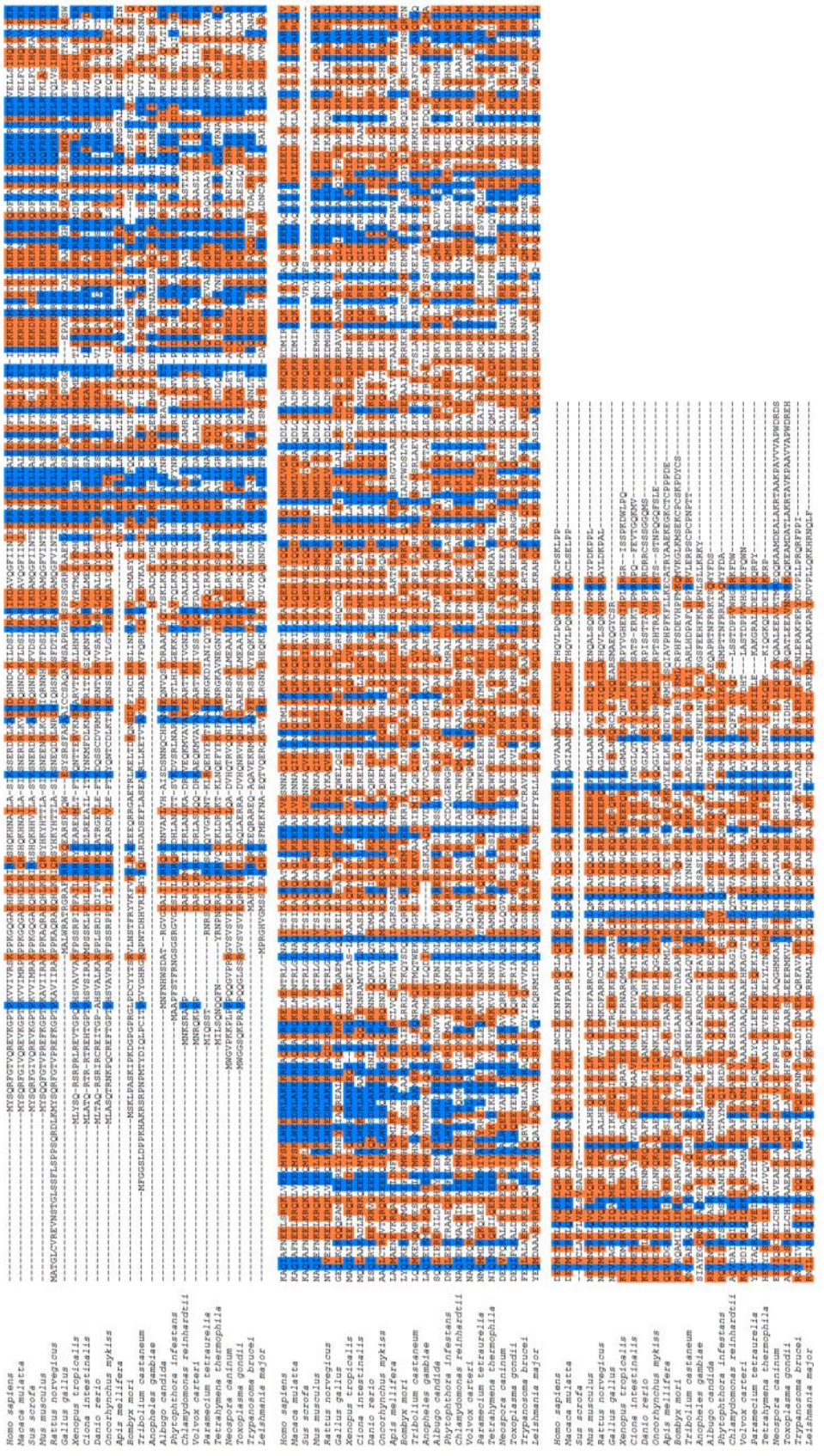


Figure 41: Alignment of eukaryotic CCDC11 orthologs.

Orthologous CCDC11 sequences were taken from across eukaryotes to assess the level of amino acid conservation between species. Identical amino acids are coloured blue and similar amino acids are coloured orange.

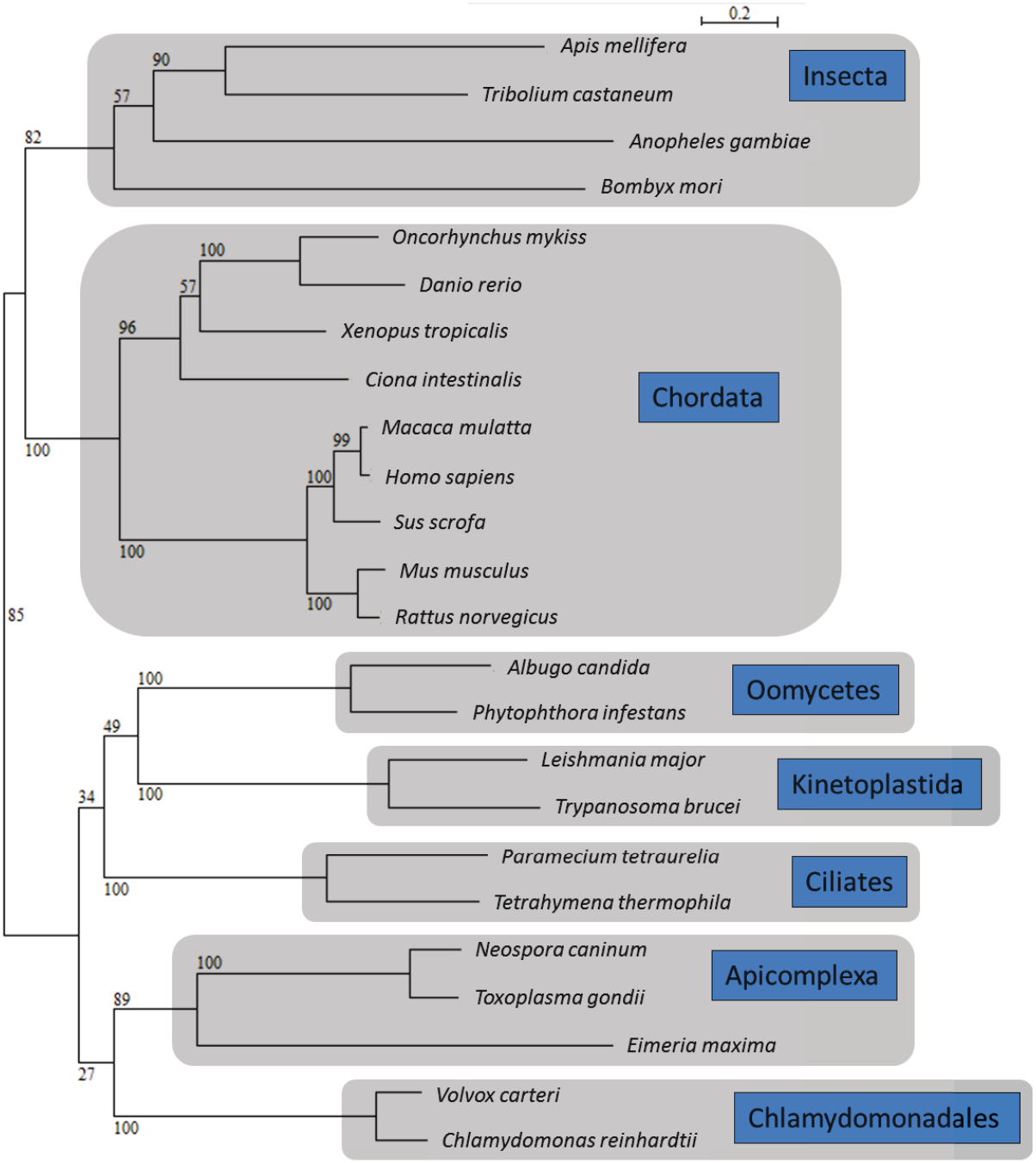


Figure 42: Phylogenetic tree for CCDC11

Maximum likelihood phylogenetic tree of aligned CCDC11 sequences from across eukaryotes. Aligned sequences were bootstrapped with 100 replicates, bootstrap values are displayed at the clade nodes. The scale bar indicates an evolutionary distance of 0.2 amino acid substitutions per position.

4.3.4 MNS1 is highly conserved

As mentioned in section 4.3.2, Tb927.6.4520 was found to be orthologous to the coiled-coil protein, meiosis nuclear specific protein 1 (MNS1) by RBH analysis. The protein sequences of Tb927.6.4520 and HsMNS1 shared 21.8% sequence identity. I therefore renamed Tb927.6.4520 as TbMNS1. Orthologs of MNS1 were searched for in my reference set of organisms (n=39, see section 3.3.1). Conserved orthologs were detected in 29 of the reference organisms by RBH. Sequences were aligned (Figure 43) and assembled into a phylogenetic tree (
Figure 44).

Discounting the 7 organisms that do not assemble a motile flagellum (negative controls), the remaining 3 organisms that lack an MNS1 ortholog are *B. dendrobatidis, D. melanogaster,* and *G. lamblia.* Despite all three of these organisms having a flagellated stage/ cell type in their life cycle. The only TPH domain-containing protein conserved in B. *dendrobatidis* and *G. lamblia* is CCDC19. *D. melanogaster* also has a CCDC19 ortholog detectable by RBH but not a bonafide MNS1 ortholog. This is investigated further in section ’4.3.13 Conservation of TPH domain-containing proteins in arthropods’.

In conclusion, MNS1 is highly but not universally conserved among the 32 flagellated eukaryotes analysed in this bioinformatics.

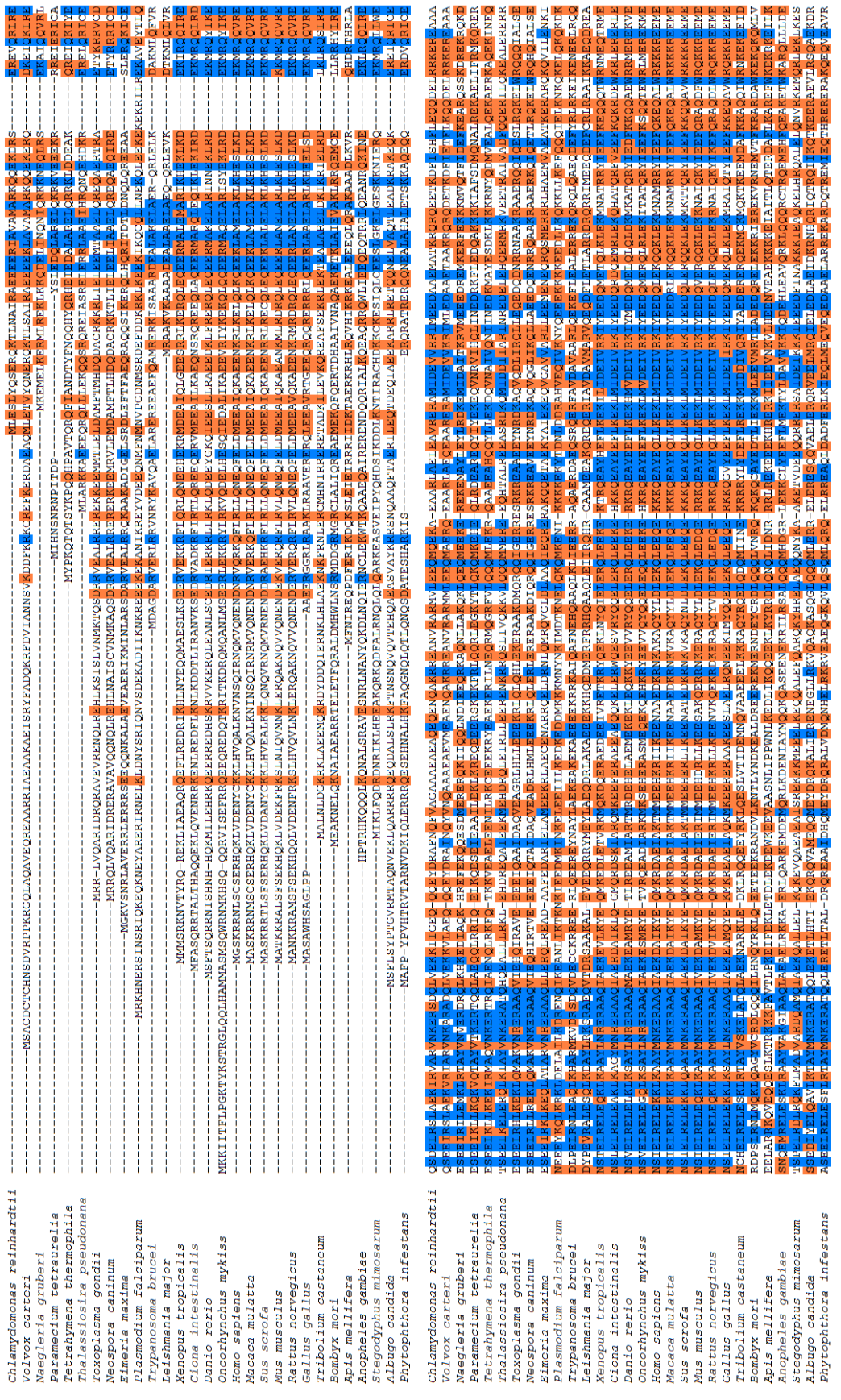


Figure 43 (part A. Figure legend on page 155)

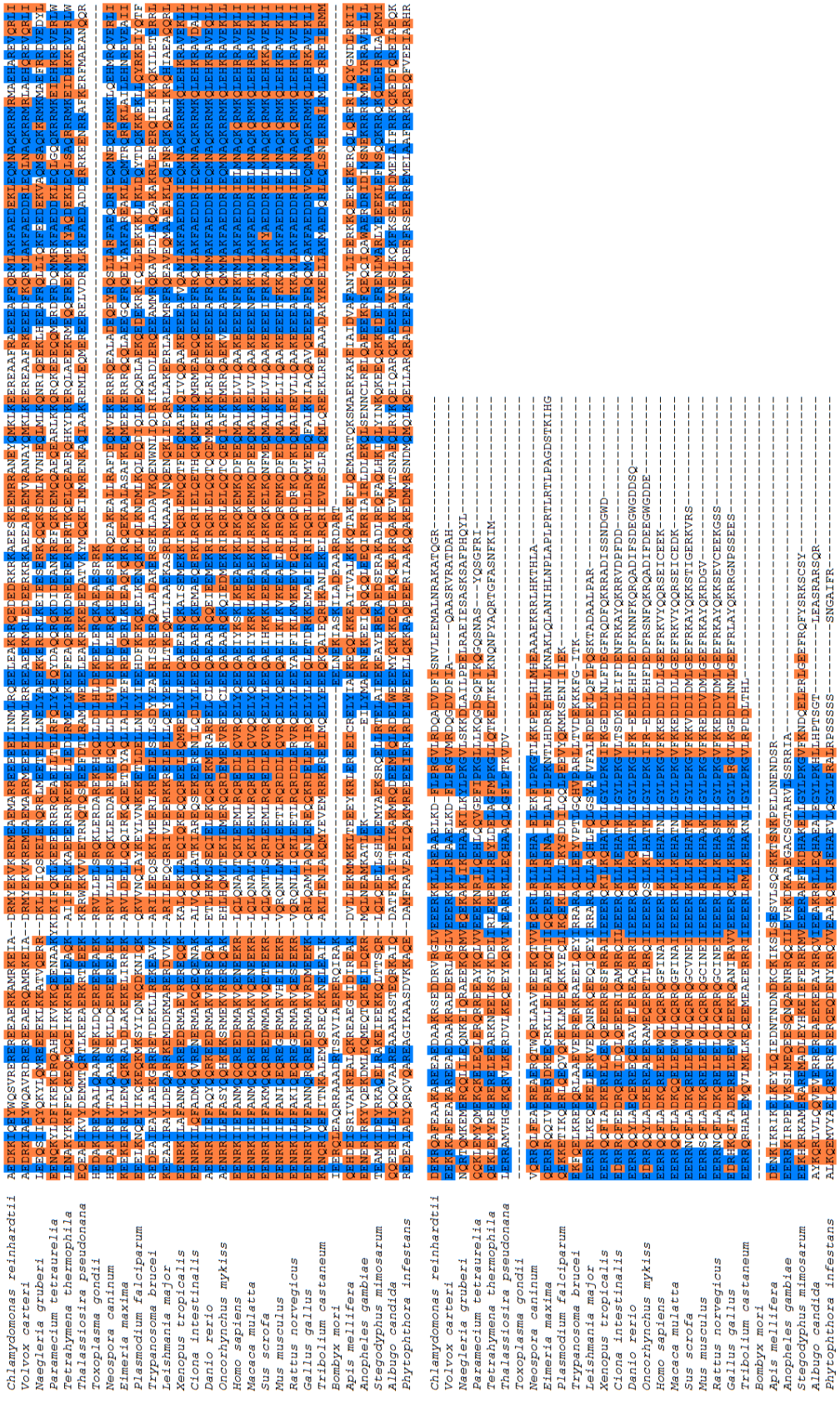


Figure 43: MNS1 alignment

Orthologous MNS1 sequences were taken from across eukaryotes to assess the level of amino acid conservation between species. Identical amino acids are coloured blue and similar amino acids are coloured orange.

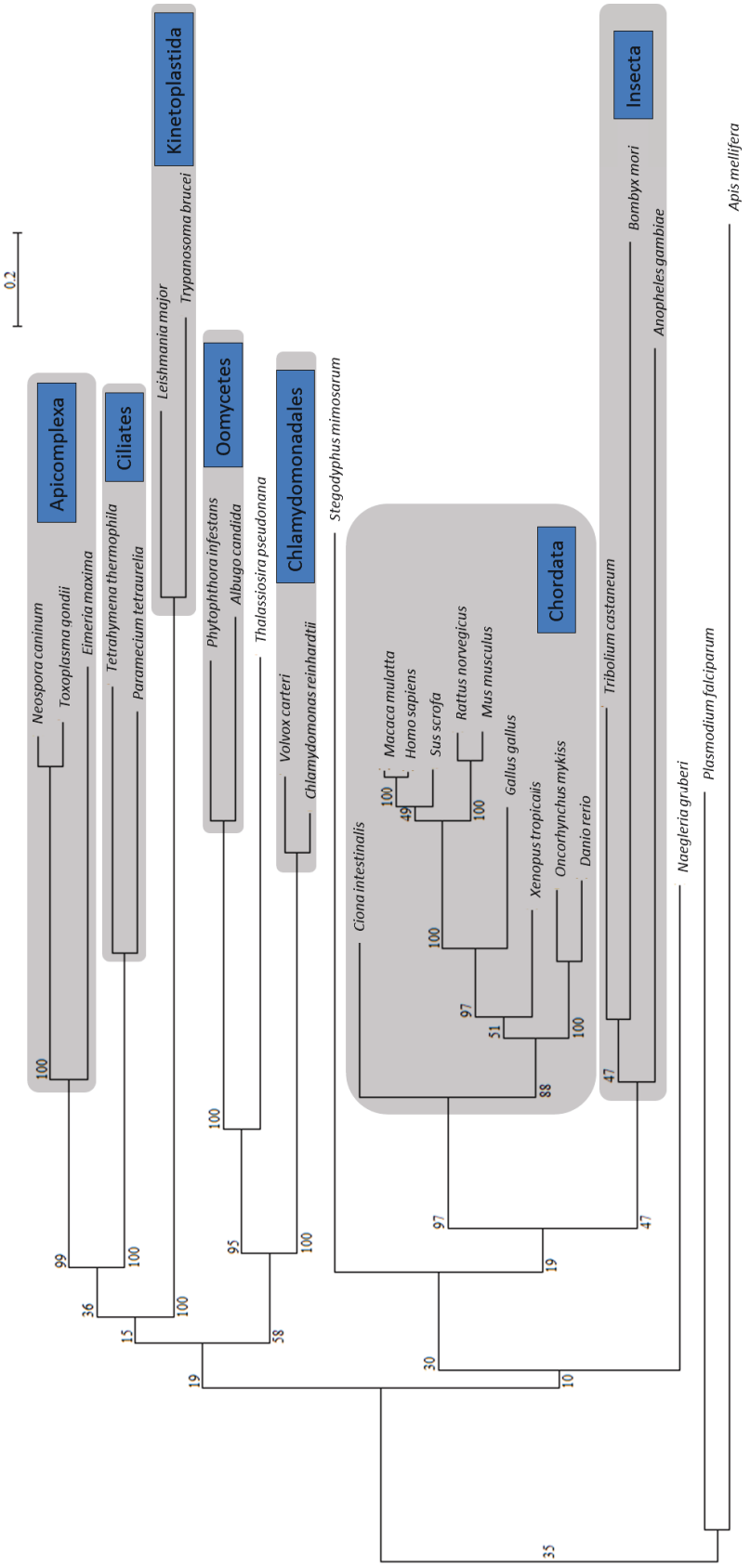


Figure 44: Phylogenetic tree for MNS1

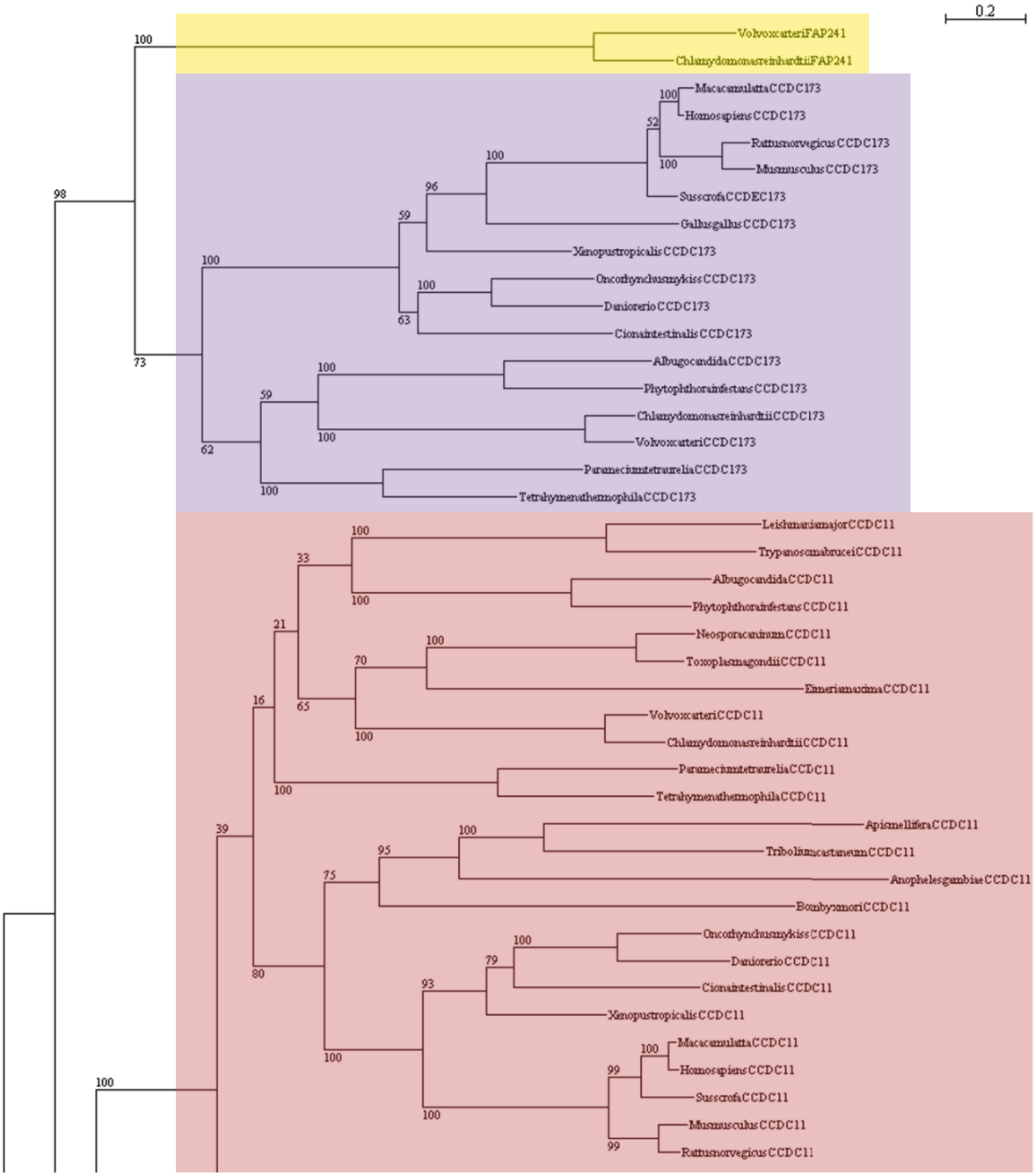
Maximum likelihood phylogenetic tree of aligned MNS1 sequences from across eukaryotes. Aligned sequences were bootstrapped with 100 replicates, bootstrap values are displayed at the clade nodes. The scale bar indicates an evolutionary distance of 0.2 amino acid substitutions per position.

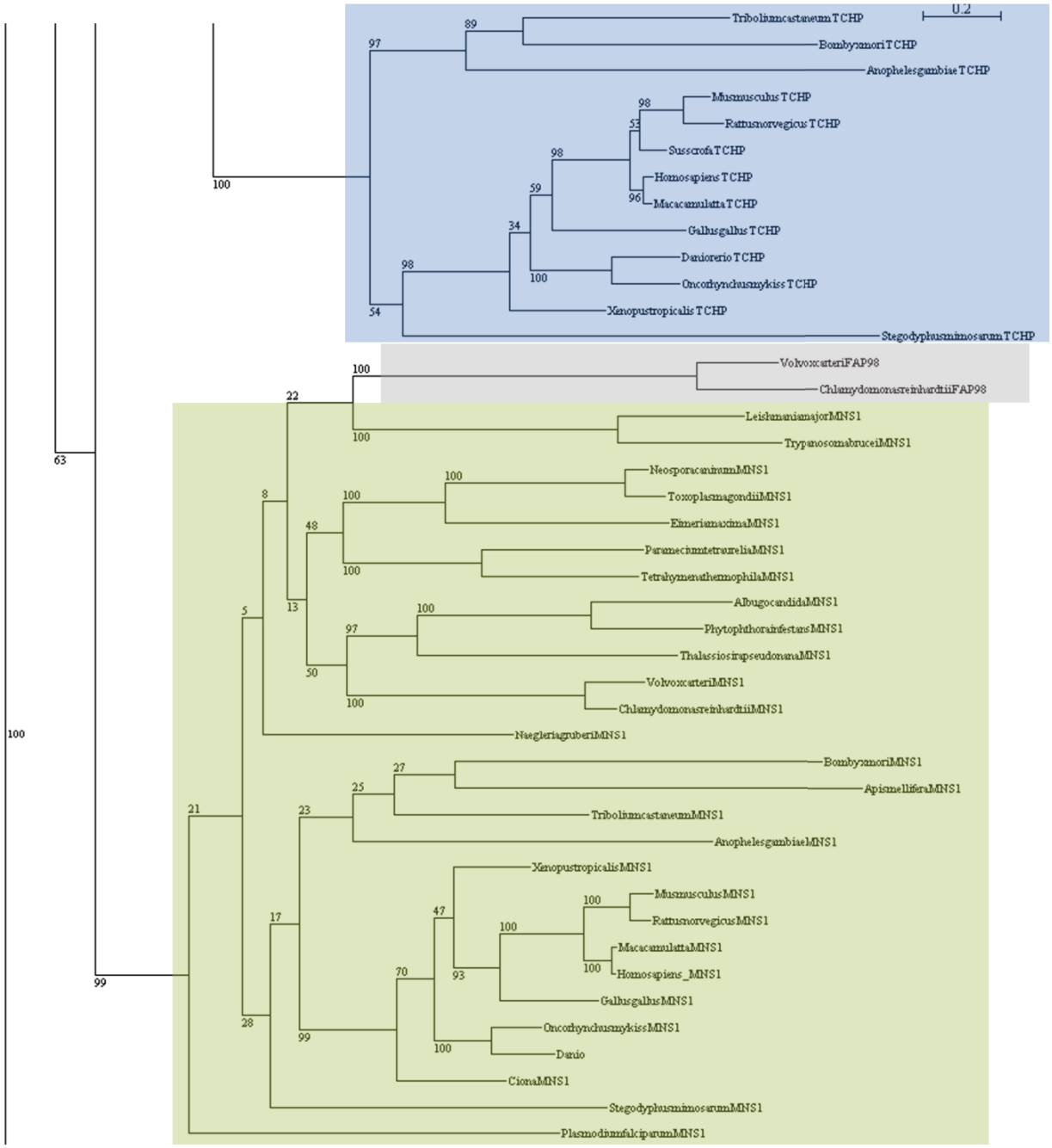
4.3.5 There are seven conserved TPH domain-containing proteins among eukaryotes

I used the species distribution list on Pfam to identify additional TPH domain-containing proteins i.e. proteins that included a TPH domain but were not orthologous to CCDC19, CCDC11 or MNS1. The list of TPH domains on Pfam contained 548 sequences across 126 species. I downloaded these 548 sequences in FASTA format from Pfam and used each one separately as the seed sequence in a BLASTp search to ascertain the nomenclature of the protein.

This approach yielded 7 discrete protein subfamilies that each contained a TPH domain. In addition to the three TPH domain-containing proteins already mentioned (CCDC19, CCDC11, MNS1) the other four subfamilies were CCDC173, FAP98, FAP241 and Trichoplein. As well as sharing a common TPH domain, these proteins are all coiled-coil proteins.

Once sequences from additional protein subfamilies had been identified they were used as bait in reciprocal BLASTp searches to detect orthologs and assess conservation within my reference set of organisms (n=39, see Table 9). Orthologs for the additional subfamilies of TPH domain-containing proteins were found using the same RBH method (see section 2.2.2 Identifying protein orthologs). The results are summarised in Table 17. These 7 subfamilies are distinct and group separately on a phylogenetic tree (Figure 45). In conclusion there are 7 main subfamilies of TPH domain-containing proteins in eukaryotes.


0.2
VolvoxcarteriFAP241
ChlamydomonasreinhardtiiFAP241
MacacamulattaCCDC173
HomosapiensCCDC173
RattusnorvegicusCCDC173
MusmusculusCCDC173
SusscrofaCCDEC173
GallusgallusCCDC173
XenopustropicalisCCDC173
OncorhynchusmykissCCDC173
DaniorerioCCDC173
CionaintestinalisCCDC173
AlbugocandidaCCDC173
PhytophthorainfestansCCDC173
ChlamydomonasreinhardtiiCCDC173
VolvoxcarteriCCDC173
ParameciumtetraureliaCCDC173
TetrahymenathermophilaCCDC173
LeishmaniamajorCCDC11
TrypanosomabruceiCCDC11
AlbugocandidaCCDC11
PhytophthorainfestansCCDC11
NeosporacaninumCCDC11
ToxoplasmagondiiCCDC11
EimeriamaximaCCDC11
VolvoxcarteriCCDC11
ChlamydomonasreinhardtiiCCDC11
ParameciumtetraureliaCCDC11
TetrahymenathermophilaCCDC11
ApismelliferaCCDC11
TriboliumcastaneumCCDC11
AnophelesgambiaeCCDC11
BombyxmoriCCDC11
OncorhynchusmykissCCDC11
DaniorerioCCDC11
CionaintestinalisCCDC11
XenopustropicalisCCDC11
MacacamulattaCCDC11
HomosapiensCCDC11
SusscrofaCCDC11
MusmusculusCCDC11
RattusnorvegicusCCDC11


0.2
TriboliumcastaneumTCHP
BombyxmoriTCHP
AnophelesgambiaeTCHP
MusmusculusTCHP
RattusnorvegicusTCHP
SusscrofaTCHP
HomosapiensTCHP
MacacamulattaTCHP
GalhusgalhusTCHP
DaniorerioTCHP
OncorhynchusmykissTCHP
XenopustropicalisTCHP
StegodyphusmimosarumTCHP
VolvoxcarteriFAP98
ChlamydomonasreinhardtiiFAP98
LeishmaniamajorMNS1
TrypanosomabruceiMNS1
NeosporacaninumMNS1
ToxoplasmagondiiMNS1
EimeriamaximaMNS1
ParameciumtetraureliaMNS1
TetrahymenathermophilaMNS1
AlbugocandidaMNS1
PhytophthorainfestansMNS1
ThalassiosirapseudonanaMNS1
VolvoxcarteriMNS1
ChlamydomonasreinhardtiiMNS1
NaegleriagruberiMNS1
BombyxmoriMNS1
ApismelliferaMNS1
TriboliumcastaneumMNS1
AnophelesgambiaeMNS1
XenopustropicalisMNS1
MusmusculusMNS1
RattusnorvegicusMNS1
MacacamulattaMNS1
Homosapiens_MNS1
GalhusgalhusMNS1
OncorhynchusmykissMNS1
Danio
CionaMNS1
StegodyphusmimosarumMNS1
PlasmodiumfalciparumMNS1

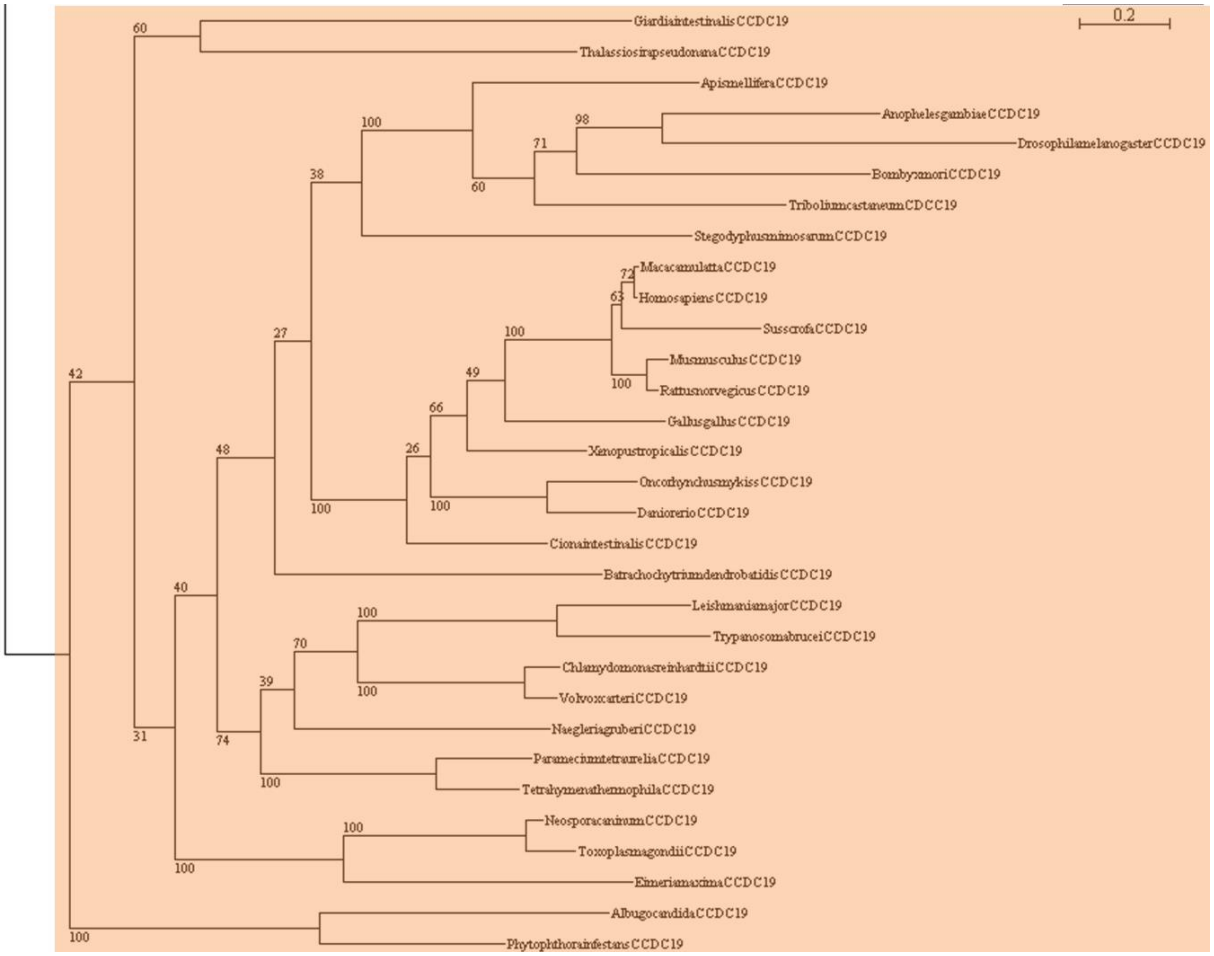


Figure 45: Phylogenetic tree for TPH domain-containing proteins

A maximum likelihood phylogenetic tree to show the relationship between the 7 subfamilies of TPH domain-containing proteins. Protein sub families are highlighted in yellow (FAP241), purple (CCDC173), red (CCDC11), blue (trichoplein), grey (FAP98), green (MNS1) and orange (CCDC19). 100 bootstraps were performed. Scale bar = 0.2 amino acid substitutions.

Table 17: Conservation of TPH domain-containing proteins across eukaryotes

The reference set of 39 eukaryotic organisms were investigated for the presence of orthologs. An empty circle (white) indicates no ortholog was detected by RBH. A full circle (black) indicates the presence of an ortholog that was detectable through RBH. A grey circle indicates a '-like' protein that is not a direct ortholog detectable through RBH.

Organism	CCDC11	CCDC19	CCDC173	FAP98	FAP241	MNS1	TCHP
Acanthamoeba castellani	○	○	○	○	○	○	○
Albugo candida	●	●	●	○	○	●	○
Anopheles gambiae	●	●	○	○	○	●	●
Apis mellifera	●	●	○	○	○	●	○
Arabidopsis thaliana	○	○	○	○	○	○	○
Batrachochytrium dendrobatidis	○	●	○	○	○	○	○
Bombyx mori	●	●	○	○	○	●	●
Caenorhabditis elegans	○	○	○	○	○	○	○
Chlamydomonas reinhardtii	●	●	●	●	●	●	○
Ciona intestinalis	●	●	●	○	○	●	○
Cryptosporidium parvum	○	○	○	○	○	○	○
Danio rerio	●	●	●	○	○	●	●
Dictyostelium discoideum	○	○	○	○	○	○	○
Drosophila melanogaster	○	●●	○	○	○	●●	○
Eimeria maxima	●	●	○	○	○	●	○
Gallus gallus	○	●	●	○	○	●	●
Giardia lamblia	○	●	○	○	○	○	○
Homo sapiens	●	●	●	○	○	●	●
Leishmania major	●	●	○	○	○	●	○
Macaca mulatta	●	●	●	○	○	●	●
Mus musculus	●	●	●	○	○	●	●
Naegleria gruberi	○	●	○	○	○	●	○
Neospora caninum	●	●	○	○	○	●	○
Oncorhynchus mykiss	●	●	●	○	○	●	●
Paramecium tetraurelia	●	●	●	○	○	●	○
Phytophthora infestans	●	●	●	○	○	●	○
Plasmodium falciparum	○	○	○	○	○	●	○
Rattus norvegicus	●	●	●	○	○	●	●
Saccharomyces cerevisiae	○	○	○	○	○	○	○
Schizosaccharomyces pombe	○	○	○	○	○	○	○
Stegodyphus mimosarum	○	●	○	○	○	●	●
Sus scrofa	●	●	●	○	○	●	●
Tetrahymena thermophila	●	●	●	○	○	●	○
Thalassiosira pseudonana	○	●	○	○	○	●	○
Toxoplasma gondii	●	●	○	○	○	●	○
Tribolium castaneum	●	●	○	○	○	●	●
Trypanosoma brucei	●	●	○	○	○	●	○
Volvox carteri	●	●	●	●	●	●	○
Xenopus tropicalis	●	●	●	○	○	●	●

4.3.6 TPH domain-containing proteins are absent from organisms that lack motile cilia

TPH domain-containing proteins were found to not be conserved in organisms that lack motile cilia/flagella (Table 17). Of the 39 reference organisms used in this study, 7 do not form motile cilia/flagella at any point in their lifecycle. These non-flagellated organisms were; the amoeba *A. castellani*, the land plant *A. thaliana*, the nematode *C. elegans*, the apicomplexan *C. parvum*, the slime mould *D. discoideum*, the yeast *S. cerevisiae* (budding yeast) and *S. pombe* (fission yeast). These organisms acted as negative controls for TPH domain-containing proteins as they were not expected to be conserved.

In conclusion, all seven subfamilies of TPH domain-containing proteins were consistently absent from organisms that do not build a motile flagellum.

4.3.7 Conservation of CCDC173

The TPH domain-containing protein, CCDC173, was identified through analysis of the TPH domains listed in Pfam. RBH analysis revealed that CCDC173 was not highly conserved, being only present in 16 of the 39 reference organisms. An alignment of CCDC173 ortholog sequences in Figure 46 shows a high level of identity amongst sequences from the phylum chordata. As well as being conserved in chordates, CCDC173 orthologs were also detected in algae, ciliates and oomycetes but not conserved in kinetoplastids, apicomplexa or arthropods. The aligned sequences were assembled into a phylogenetic tree (Figure 47).

In conclusion, CCDC173 was universally conserved in chordates, universally absent from arthropods and exhibited a strange conservation pattern in single celled eukaryotes.

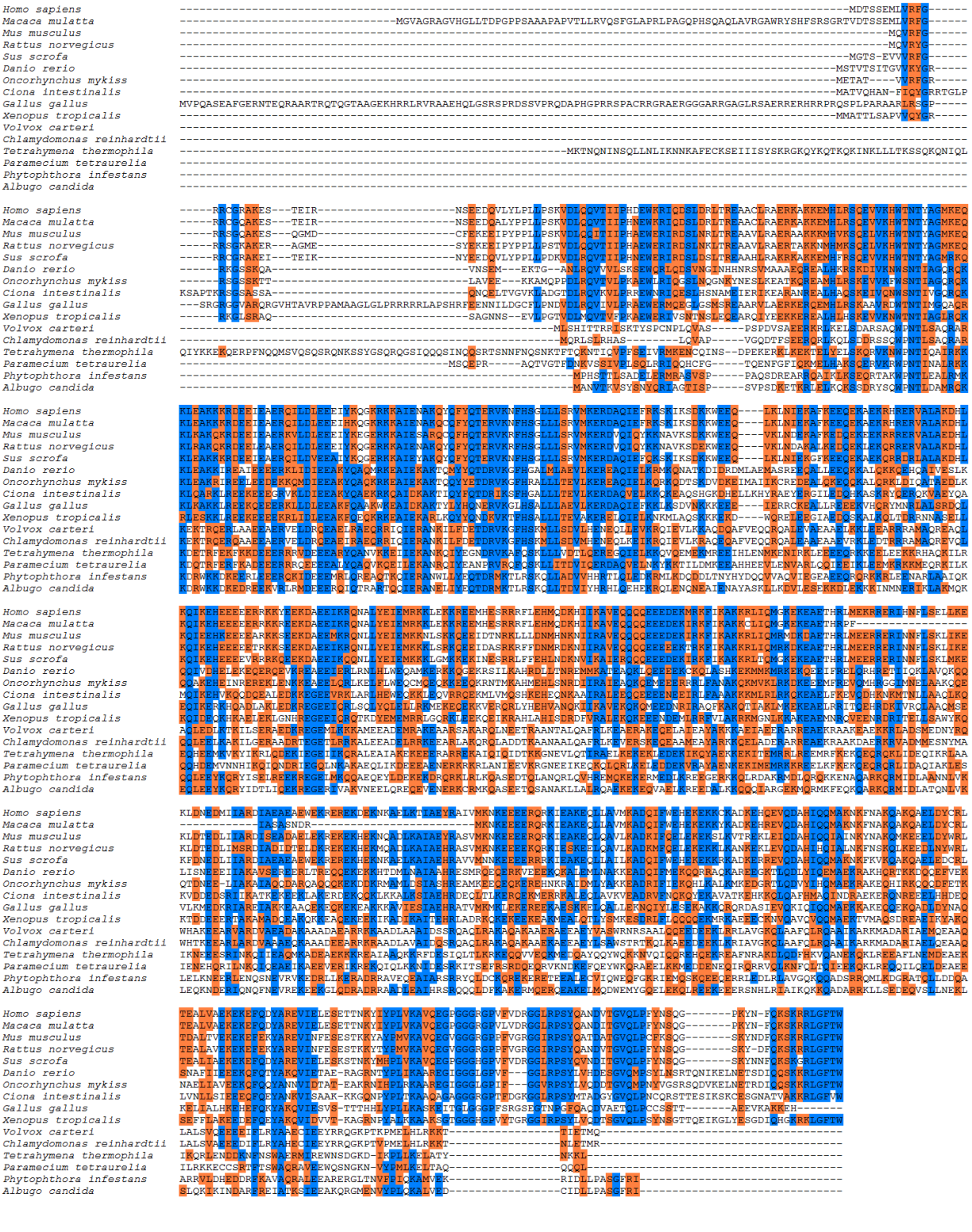


Figure 46: Alignment of CCDC173 orthologs

CCDC173 sequences were taken from across eukaryotes to assess the level of amino acid conservation between orthologs. Identical amino acids are coloured blue and similar amino acids are coloured orange.

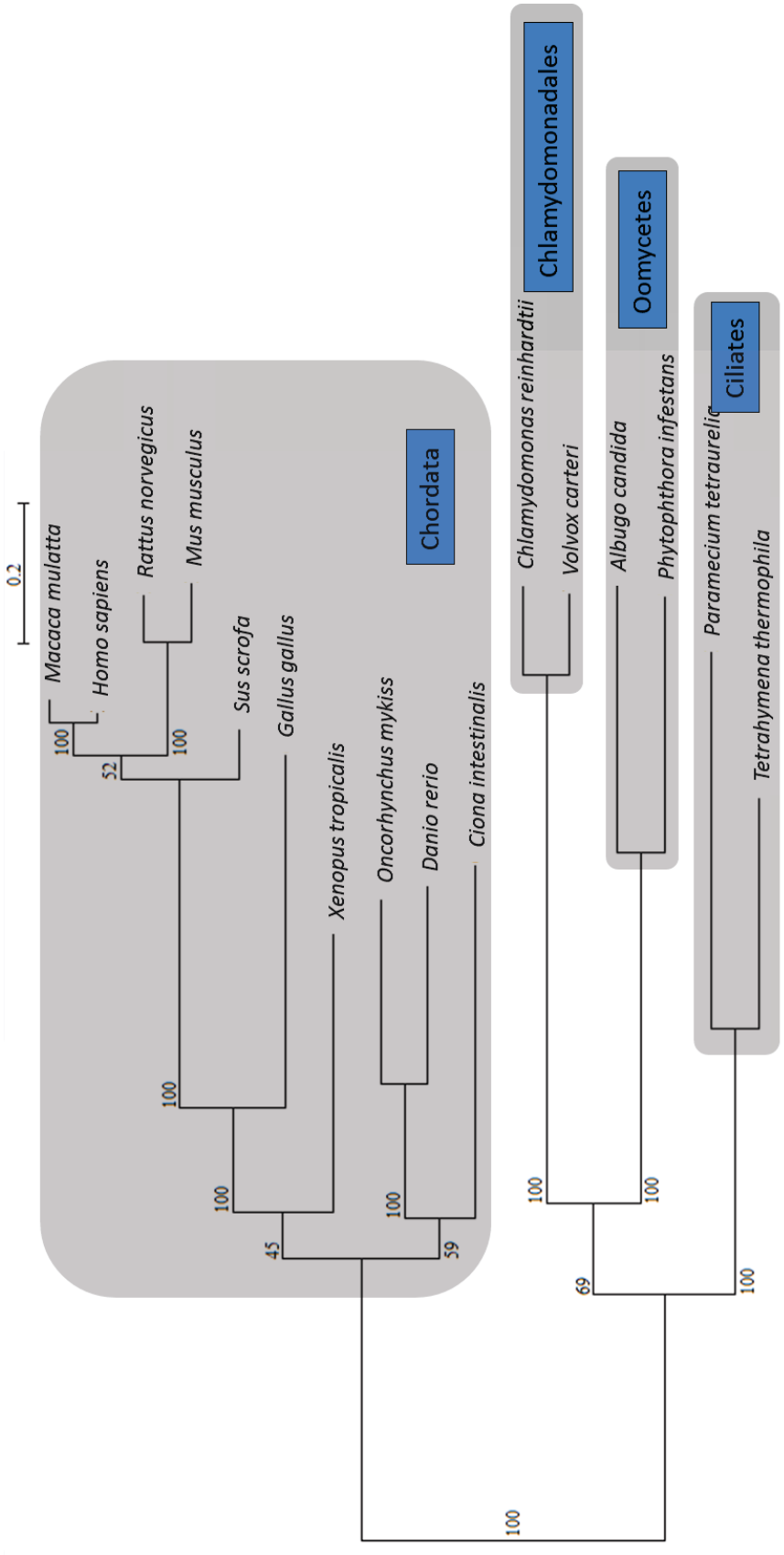


Figure 47: Phylogenetic tree of CCDC173 orthologs

Maximum likelihood phylogenetic tree of aligned CCDC173 sequences (n=16) from across eukaryotes Aligned sequences were bootstrapped with 100 replicates, bootstrap values are displayed at the clade nodes. The *s*cale bar indicates an evolutionary distance of 0.2 amino acid substitutions per position.

4.3.8 Conservation of FAP98 and FAP241

Two of the TPH domain-containing subfamilies are called flagellar associated proteins (FAP), which are so named after being detected in the *Chlamydomonas reinhardtii* flagellar proteome (Pazour *et al.*, 2005), although neither has been directly localised.

Orthologs of FAP98 and FAP241 were only detected in *Chlamydomonas reinhardtii* and *Volvox carteri,* which are both algae from the class Chlorophyceae. No other hits were detected during RBH analysis. An alignment of the sequences (Figure 48) demonstrates that the proteins are highly similar. In fact CrFAP98 and VcFAP98 share 54.1% amino acid identity (294/543 AA) between protein sequences and 68.1% similarity. CrFAP241 and VcFAP241 share 65.5% amino acid identity (97/148) between protein sequences and 83.1% similarity.

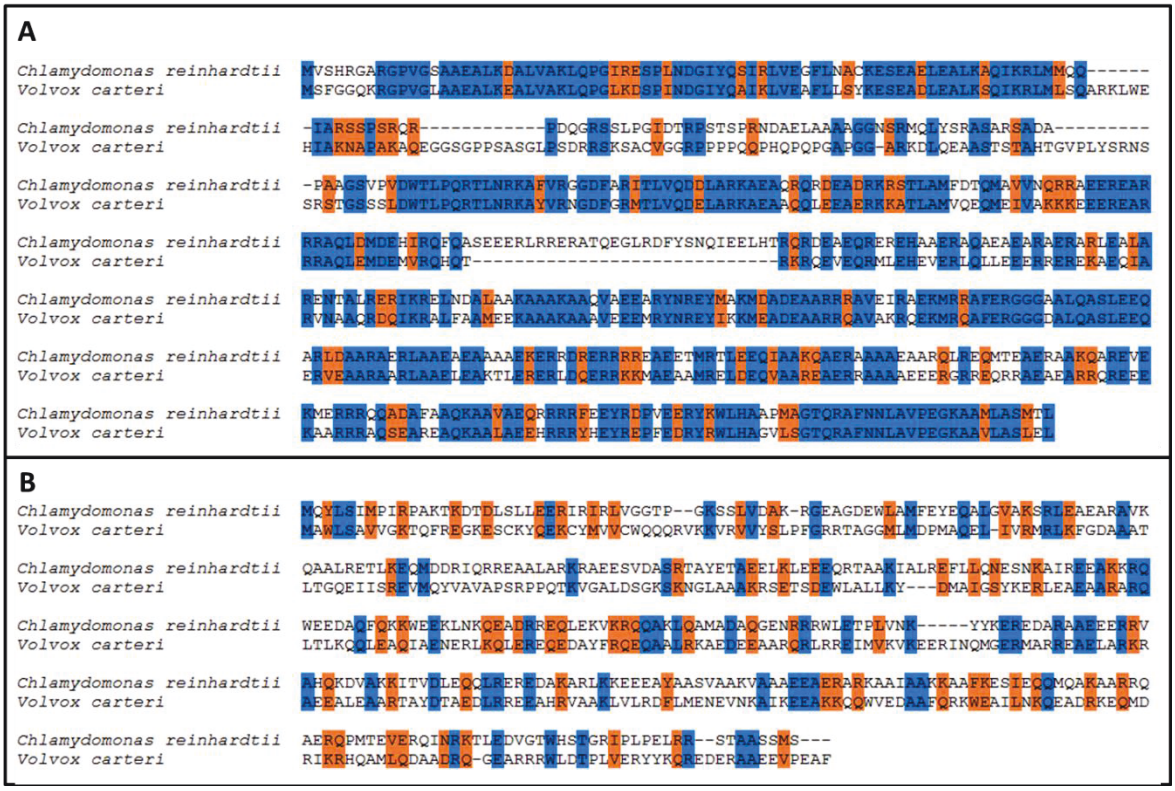


Figure 48: Alignments of FAP98 and FAP241

Alignment of the *C. reinhardtii* (top row) and *V. carteri* (bottom row) sequences for (A) FAP98 orthologs and (B) FAP241 orthologs. Identical amino acids are highlighted in blue and similar amino acids are highlighted in orange.

4.3.9 Conservation of trichoplein

As mentioned in section 4.4.8, trichoplein is one of the 7 subfamilies of TPH domain-containing proteins. Trichoplein orthologs were only detected in 13 of the 39 reference organisms by RBH analysis with an obvious restriction to metazoans. Alignment of orthologous trichoplein sequences (Figure 49) demonstrates the high similarity of orthologs between chordata species and that orthologs from arthropoda are more divergent in sequence. A trichoplein ortholog was not detected in *C. intestinalis* using RBH nor a relaxed BLASTp approach.

The trichoplein orthologs from *B. mori*, *S. mimosarum* and *A. gambiae* did not contain a TPH domain according to CDD or Pfam (Appendix 7) but I have classed them as trichoplein orthologs through RBH analysis. The *S. mimosarum* sequence available through the NCBI protein database was annotated as 'trichoplein' and 'partial'. A trichoplein ortholog was not detected in the insects *D. melanogaster* or *A. mellifera* through RBH analysis.

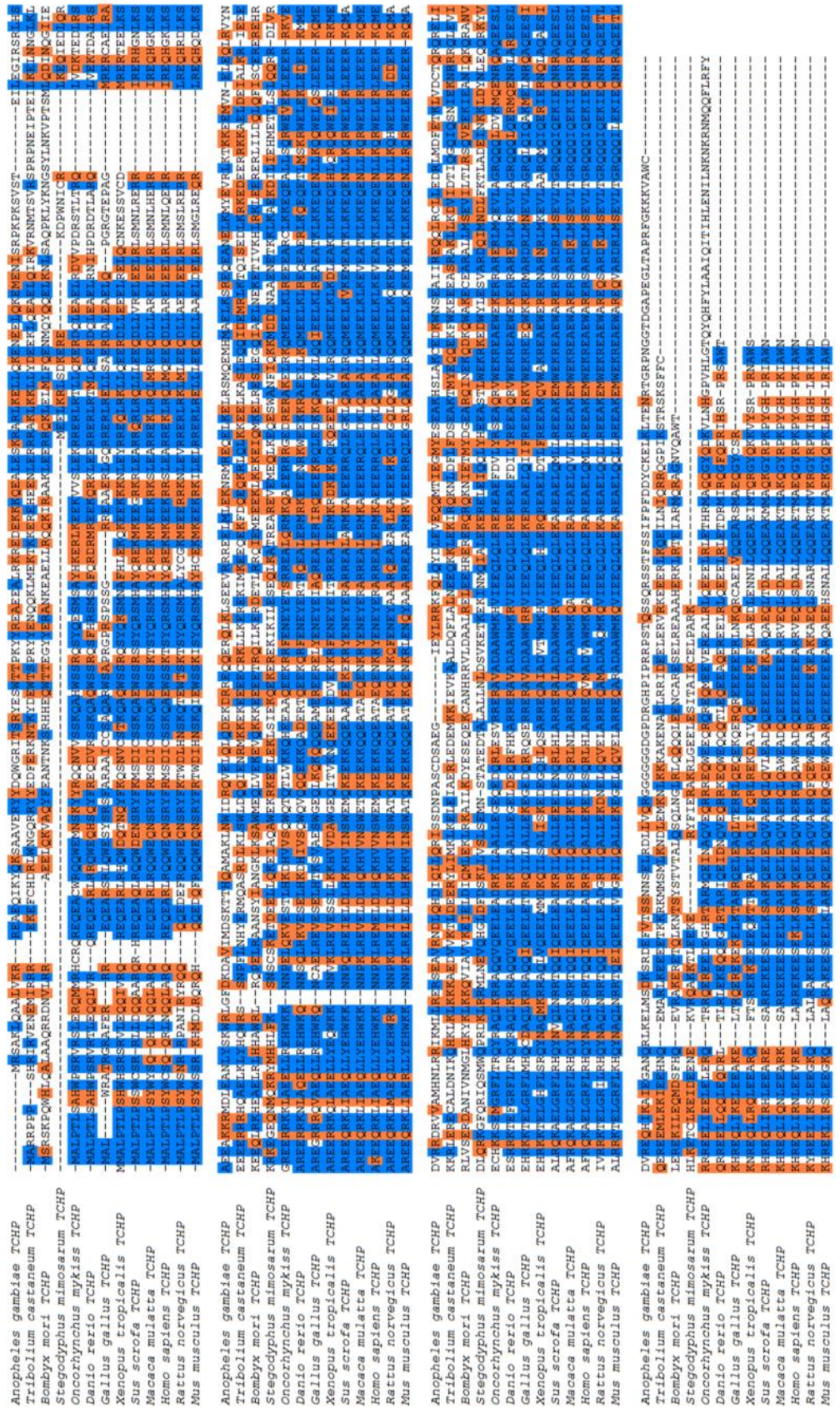


Figure 49: Alignment of trichoplein orthologs

Orthologous trichoplein sequences (n=13) were taken from across eukaryotes to assess the level of amino acid conservation between species. Identical amino acids are coloured blue and similar amino acids are coloured orange.

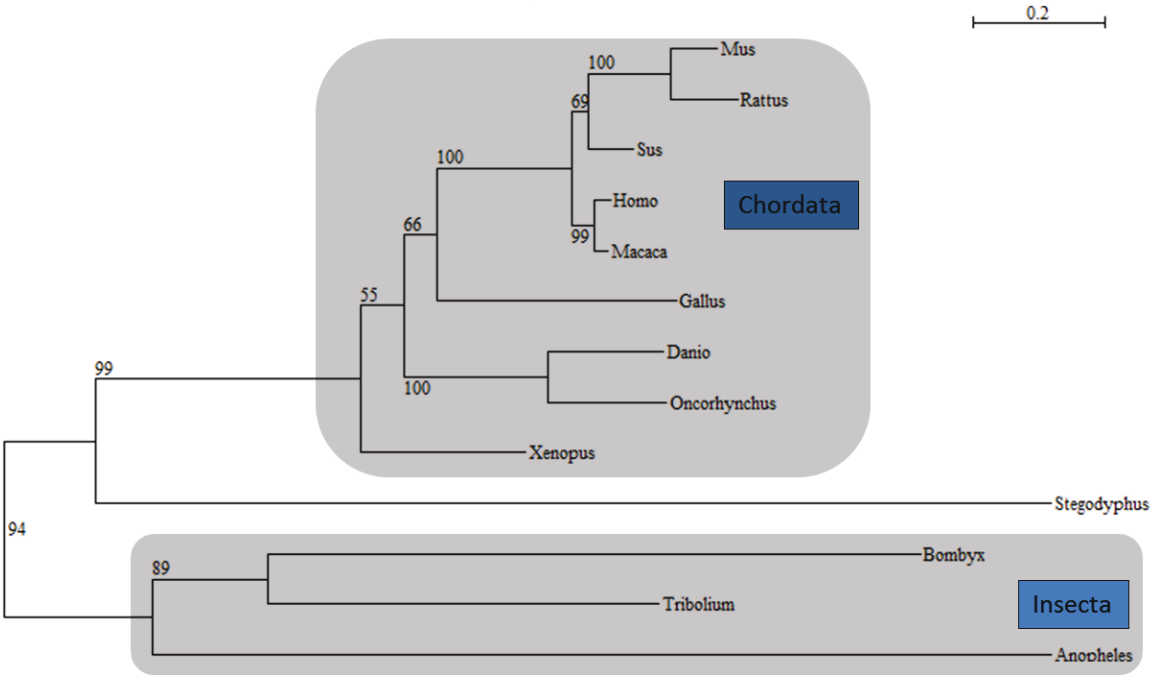


Figure 50: Phylogenetic tree of trichoplein orthologs

Aligned trichoplein sequences were bootstrapped with 100 replicates, bootstrap values are displayed at the clade nodes. The *s*cale bar indicates an evolutionary distance of 0.2 amino acid substitutions.

4.3.10 Elucidating the architecture of the TPH domain

To investigate the composition of the shared TPH domain, all proteins from the 39 reference organisms across the seven subfamilies were analysed *in silico* and TPH domain positions were annotated (Appendix 7). The average length of a TPH domain was 323 AA (n=113), which occupied 65.8% of the whole protein on average.

From the 39 organisms a total of 118 protein sequences had been identified through RBH analysis inclusive of the seven subfamilies. Protein lengths were retrieved from NCBI and the position of TPH domains were annotated by entering the amino acid sequence into two separate search engines; (i) the conserved domain database (CDD) (Marchler-Bauer *et al.*, 2005; Marchler-Bauer *et al.*, 2011) and (ii) Pfam (Punta *et al.*, 2012). Of the 118 sequences, five of the proteins did not contain a detectable TPH domain according to either CDD or Pfam. I have included these 5 sequences in my analysis as they were detected as orthologs through a RBH approach. Fifteen of the protein sequences had a TPH domain that was only detected by one of the two methods used but not both. In 14 cases CDD detected a TPH domain and Pfam did not and in only 1 case Pfam detected a TPH domain when CDD did not. In 98 sequences a TPH domain was detected as present by both CDD and Pfam but only 22 of those 98 were annotated at identical amino acid positions within the protein.

The TPH domain was excised from each protein (n=113) in silico and used to generate a hidden markov model (HMM) (Appendix 6). Pfam generates an automatic HMM for documented domains. However, the species contained within the Pfam species distribution tree for the TPH domain is heavily biased; The Pfam HMM logo is generated from 564 TPH domains recorded within Pfam. However 68% of those sequences are from metazoa (383/564), which is not truly representative of eukaryotes. I used a more diverse set of reference organisms to generate a HMM for the TPH domain.

4.3.11 Albatross does not belong to the TPH domain family

The protein Albatross has been described as containing a TPH domain in the literature (Sugimoto *et al.*, 2008) although I did not detect Albatross through my bioinformatics based search for TPH domain-containing proteins and Albatross orthologs are not listed in Pfam. For this reason, Albatross proteins are not included in the protein family phylogenetic tree (Figure 50).

Sugimoto and colleagues (2008) used both the mouse (NP_766159) and human (BAG71501) Albatross proteins in their study so these sequences were submitted to both Pfam and CDD. Neither search algorithm detected a TPH domain in the Albatross protein sequence from either species (Figure 51 A and B), HsCCDC19 was also included as a positive control where the known TPH domain is denoted in green (Figure 51; C).

Therefore I propose that Albatross does not contain a TPH domain.

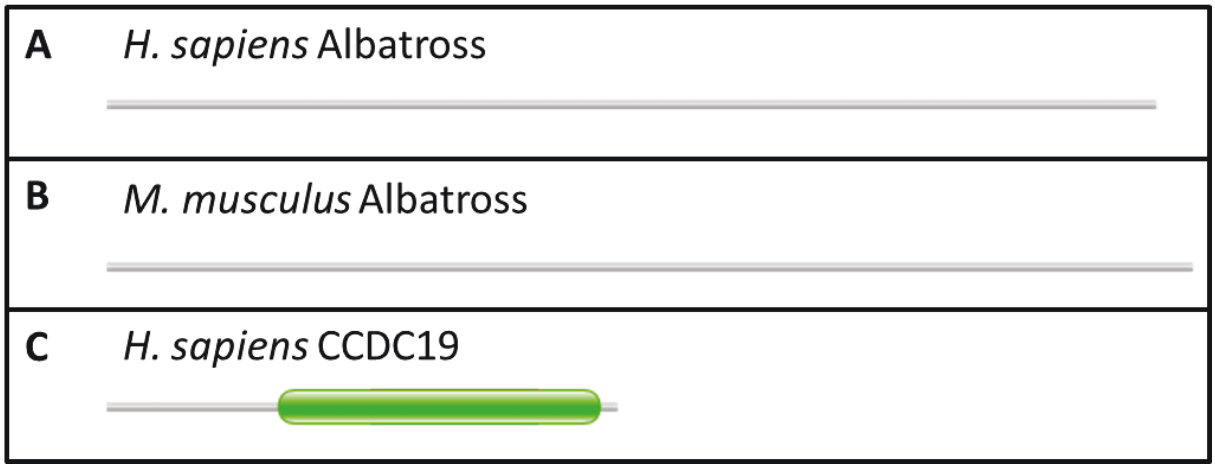


Figure 51: Pfam domain detection

The amino acid sequence of (A) *H.sapiens* Albatross BAG71501, (B) *M. musculus* Albatross NP_766159 and (C) (A) *H.sapiens* CCDC19 NP_036469 were entered into Pfam for domain detection. Grey bar indicates the protein and any detected domains are demonstrated in green.

4.3.12 TPH domain-containing proteins in ciliates

Paralogs created from whole genome duplications in *Paramecium* spp. are called 'ohnologs' (Aury *et al.*, 2006). Due to three certain whole genome duplication events within *Paramecium tetraurelia* it is possible to have up to 8 ohnologs for each original protein coding gene.

P. tetraurelia had two ohnologs for both CCDC173 and CCDC19, which I refer to as –A and –B. *P. tetraurelia* had four MNS1 ohnologs, which I refer to as –A, –B, –C and –D and only a single copy of CCDC11. The alignment between the four PtMNS1 ohnologs (Figure 52; A) showed the 4 sequences were highly similar. This was also evident in the percentage identity and similarity scores of the amino acid sequences between the four PtMNS1 ohnologs (Figure 52; B). From these values it is likely that the most recent whole genome replication event created MNS1-A and –B from a common source and MNS1-C and –D likewise as MNS1-A and B are most identical to each other (96.4%) and –C and –D also have a high identity value (94.5%). This hypothetical pattern is also supported by the arrangement of the phylogenetic tree branches (Figure 52; C).

The multiple ohnologs were unique to *P. tetraurelia* as the closely related ciliate, *T. thermophila* only had a single copy of the conserved TPH domain-containing proteins (Figure 52; C).

In conclusion, *P. tetraurelia* has ohnologs of CCDC19, CCDC173 and MNS1. These are most likely due to whole genome duplication events as these paralogs were not present in *T. thermophila.*

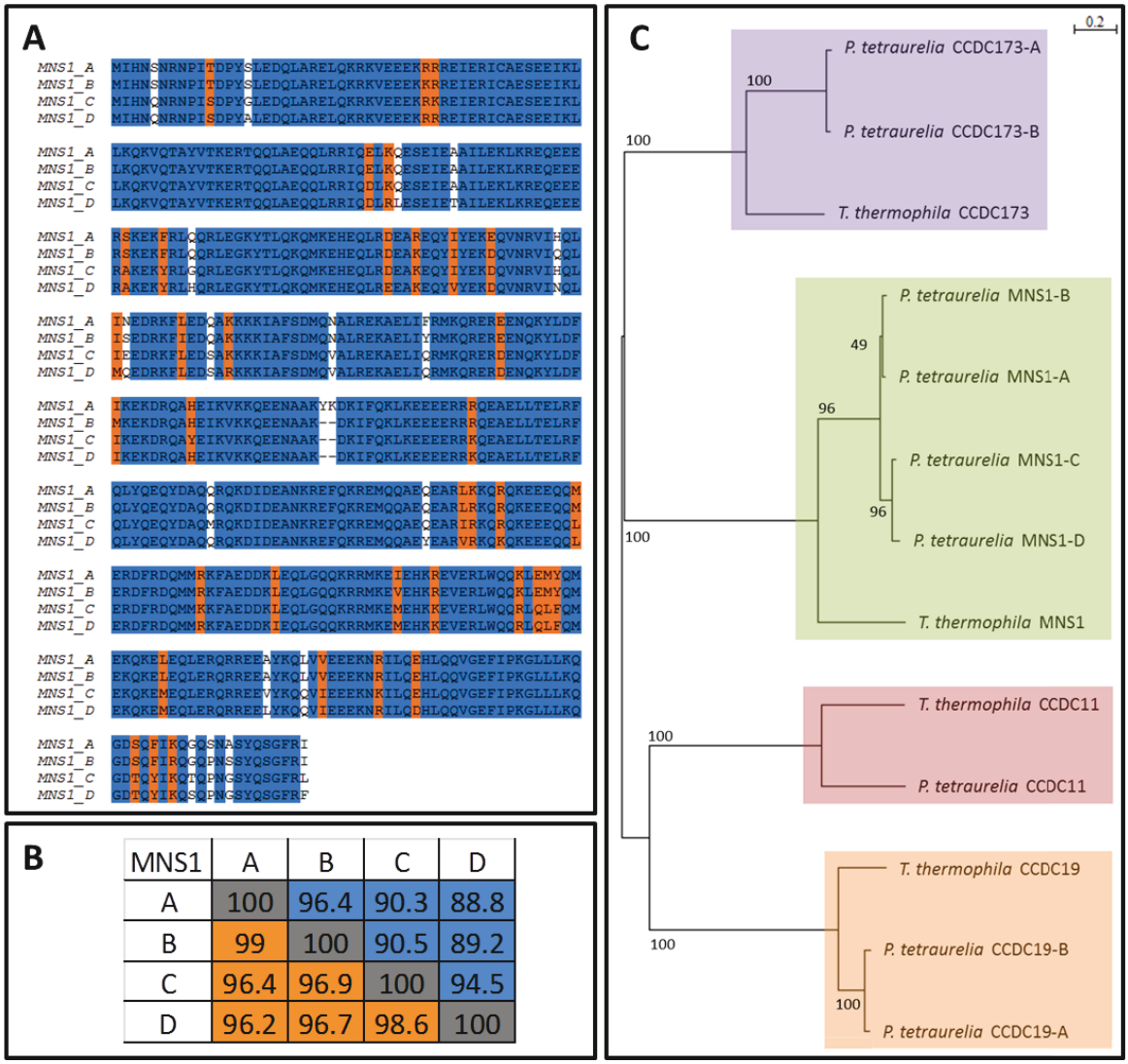


Figure 52: Ohnologs within *Paramecium tetraurelia*

A) Alignment of the four MNS1 ohnologs of *P. tetraurelia.* Identical amino acids are coloured blue and similar amino acids are coloured orange. B) Identity (blue) and similarity (orange) percentages between the four *P. tetraurelia* MNS1 ohnologs. C) A maximum likelihood tree showing the relationship between TPH domain-containing proteins of the ciliates *P. tetraurelia* and *T. thermophila* including ohnologs where applicable. Aligned sequences were bootstrapped with 100 replicates, bootstrap values are displayed at the clade nodes. Protein subfamilies are highlighted purple (CCDC173), red (CCDC11), green (MNS1) and orange (CCDC19). The s*cale bar indicates* an evolutionary distance of *0.2* amino acid *substitutions*.

4.3.13 Conservation of TPH domain-containing proteins in arthropods

Within arthropods I found orthologs of CCDC19, MNS1, CCDC11 and trichoplein using a reciprocal best blast hit (RBH) approach. CCDC173, FAP98 and FAP241 were not conserved in arthropods (Figure 53; A). Figure 53 (B) shows the relationship between arthropod TPH domain-containing proteins. The bootstrap values at the head of each cluster node for the four families' inferred high confidence (97, 92, 93, and 100) that the four clusters are distinct families (CCDC19, MNS1, CCDC11, trichoplein respectively).

Conservation of TPH domain-containing proteins among arthropods was not unanimous. Using RBH analysis I did not detect a trichoplein ortholog in *A. mellifera* or *D. melanogaster*. CCDC11 orthologs were not detected in *D. melanogaster* or *S. mimosarum* (Figure 53).

According to Pfam, *Drosophila melanogaster* contains four proteins that have a TPH domain. One of these proteins is a true ortholog of CCDC19 and this can be demonstrated using a RBH approach. No direct orthologs for CCDC11, MNS1 or trichoplein were detected in *D. melanogaster* using RBH. The additional three *D. melanogaster* proteins were not orthologous by the RBH method but when aligned and included in a phylogenetic tree with the known arthropods sequences, the three miscellaneous TPH domain-containing proteins grouped with either CCDC19 or MNS1 (Figure 53). None grouped with CCDC11 or trichoplein. Therefore I have called these divergent orthologs proteins CCDC19-like or MNS1-like, respectively.

There is one CCDC19-like protein in *D. melanogaster* and two MNS-like proteins, it is not clear how these '-like' proteins arose as they do not appear to be close paralogs. DmCCDC19 and DmCCDC19-like shared 25.7% identity at the amino acid level whilst DmMNS1-like A and DmMNS1-like B shared only 21.7% identity. I have deliberately not called either of the 'MNS1-like' proteins a direct ortholog of MNS1 because neither protein is detected as a true ortholog of MNS1 using RBH analysis.

A

Organism	CCDC11	CCDC19	CCDC173	FAP98	FAP241	MNS1	TCHP
Anopheles gambiae	●	●	○	○	○	●	●
Apis mellifera	●	●	○	○	○	●	○
Bombyx mori	●	●	○	○	○	●	●
Drosophila melanogaster	○	●●	○	○	○	●●	○
Stegodyphus mimosarum	○	●	○	○	○	●	●
Tribolium castaneum	●	●	○	○	○	●	●

B

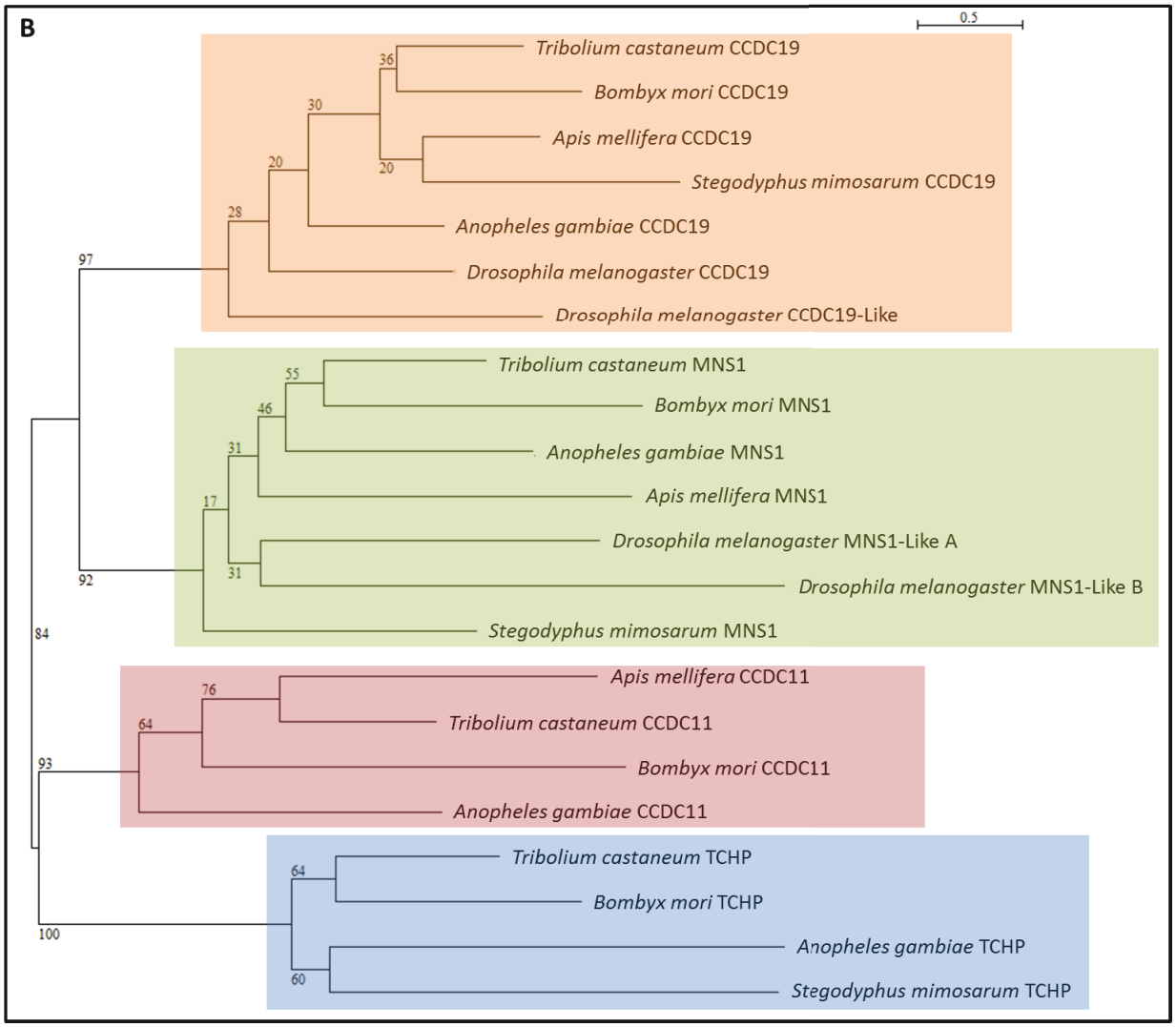


Figure 53: TPH domain-containing proteins within arthropods

(A) TPH domain-containing orthologs in arthropods. An empty circle (white) indicates no ortholog was detected by RBH. A full circle (black) indicates the presence of an ortholog that was detectable through RBH. A grey circle indicates a '-like' protein that is not a direct ortholog. (B) Phylogenetic tree showing the relationship between the TPH domain-containing protein sequences from different insects. Aligned sequences were bootstrapped with 100 replicates; bootstrap values are displayed at the nodes. The clade highlighted in Orange for CCDC19 orthologs, green for MNS1 orthologs, red denotes CCDC11 orthologs, and trichoplein cluster is in blue. The scale bar indicates an evolutionary distance of 0.5 amino acid substitutions.

It was not clear if this was an unusual conservation pattern of additional MNS1- and CCDC19-like proteins were specific to *Drosophila melanogaster*, *Drosophila* spp. or insects in general. To investigate this I looked at 12 *Drosophilidae* subspecies with fully sequenced genomes (Drosophila 12 Genomes *et al.*, 2007). The orthologous sequences were aligned and processed for a maximum likelihood phylogenetic tree (Figure 54), which showed the relationships between the protein subfamilies.

From the phylogenetic analysis (Figure 54) it was evident that *Drosophila* species have two distinct families of MNS1-like proteins (A and B) neither of which grouped with true MNS1 orthologs from other organisms (light green). An exception was MNS1-like B from *D. willistoni* (black arrowhead), which was positioned in the MNS1-like A group branched with *D. willistoni* MNS1-like A. Another anomaly in this phylogenetic analysis (Figure 54) was the position of *A. mellifera* MNS1 (arrow), which did not group with other arthropod MNS1 orthologs despite doing so in a previous phylogenetic tree (Figure 53, B).

In addition to the phylogenetic analysis of *Drosophila* spp. CCDC19 and CCDC19-like proteins (Figure 54), sequence identity percentages were also calculated (Appendix 4). Percentage sequence identity was also calculated for the *Drosophila* spp. MNS1-like proteins (Appendix 5). *Drosophila* does not possess a bona fide MNS1 ortholog as determined by RBH analysis and supported by the phylogenetic relationship of sequences (Figure 54). Neither MNS1-like clade (A or B) group with the 'true' MNS1 orthologs of other arthropods (light green).

In conclusion, CCDC11, CCDC19, MNS1 and trichoplein proteins are conserved within the arthropod lineage, but not uniformly. In addition to a bona fide CCDC19 ortholog, *D. melanogaster* has divergent orthologs of CCDC19 and MNS1.

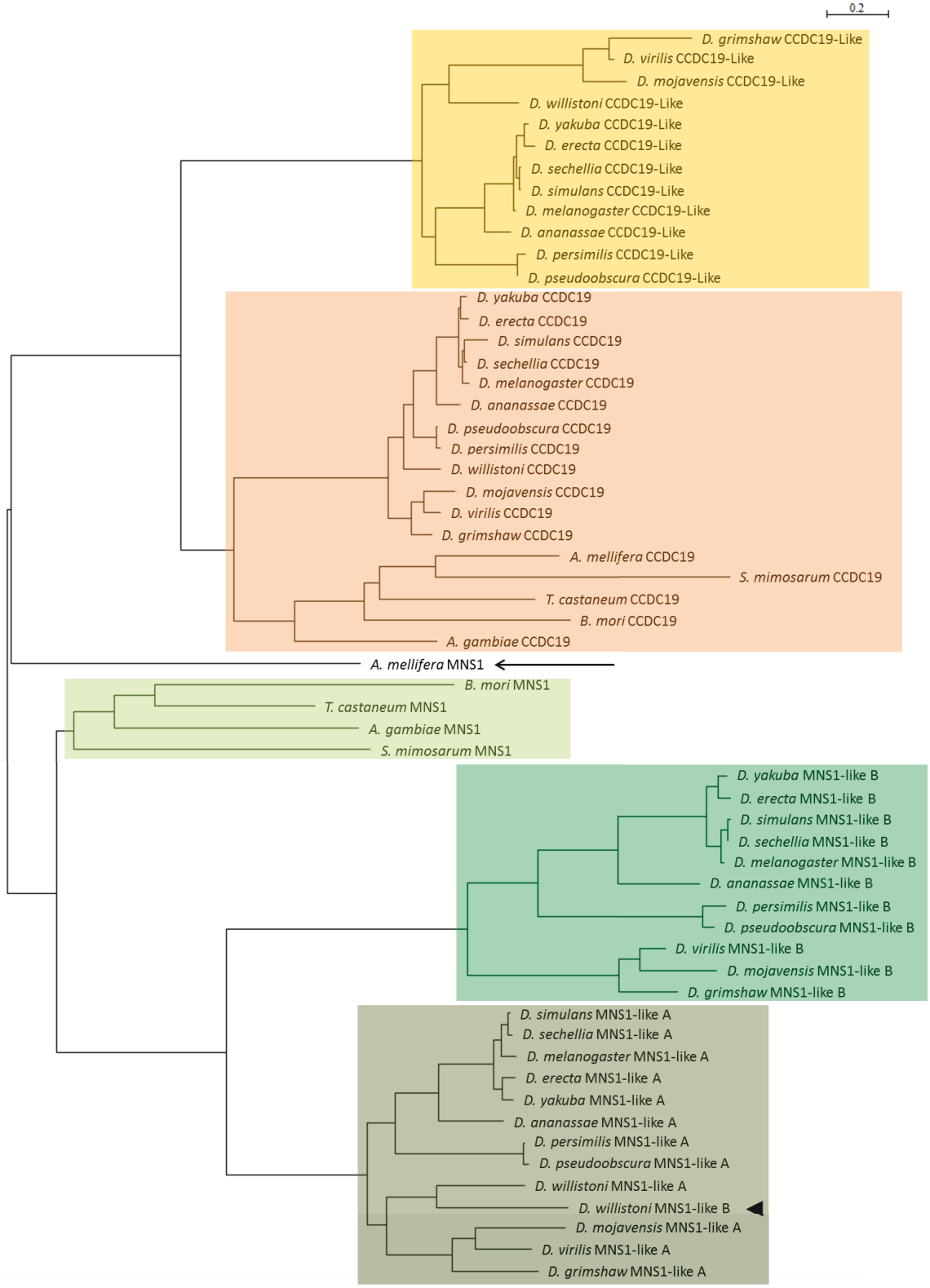


Figure 54: TPH domain-containing proteins within *Drosophila* spp.

Phylogenetic tree showing the relationship between the TPH domain-containing protein sequences from in 12 D*rosophila* species. MNS1 and CCDC19 orthologs from other arthropods are also included. Aligned sequences were bootstrapped with 100 replicates. The clade highlighted in orange for CCDC19 orthologs, light orange for CCDC19-like orthologs, light green for MNS1 orthologs, olive green for MNS1-like A orthologs and turquoise for MNS1-like B. The scale bar indicates an evolutionary distance of 0.2 amino acid substitutions

4.3.14 Unusual conservation of TPH domain-containing proteins in *plasmodium*

TPH domain-containing proteins were searched for within several apicomplexan species that are known to include a flagellated stage in their lifecycle (*P. falciparum, E. maxima, T. gondii* and *N. caninum*). *C. parvum* was also included as a negative control as this apicomplexan does not have a flagellated stage in its lifecycle and is known to lack other proteins associated with flagella (Briggs *et al.*, 2004a; Straschil *et al.*, 2010). *P. falciparum* does assemble a flagellum, albeit inside the cell rather than protruding from the main cell body. The flagellum of *P. falciparum* has the same 9+2 canonical ultrastructure of most eukaryotic flagella but does not require IFT to be assembled.

P. falciparum was the only flagellated organism not to have a CCDC19 ortholog, as determined by RBH analysis (Table 17). Orthologs of MNS1 were detectable by RBH in all four apicomplexa with flagellated lifecycle stages. No ortholog for CCDC11 was detected in *P. falciparum,* despite being conserved in other apicomplexa (Figure 55).

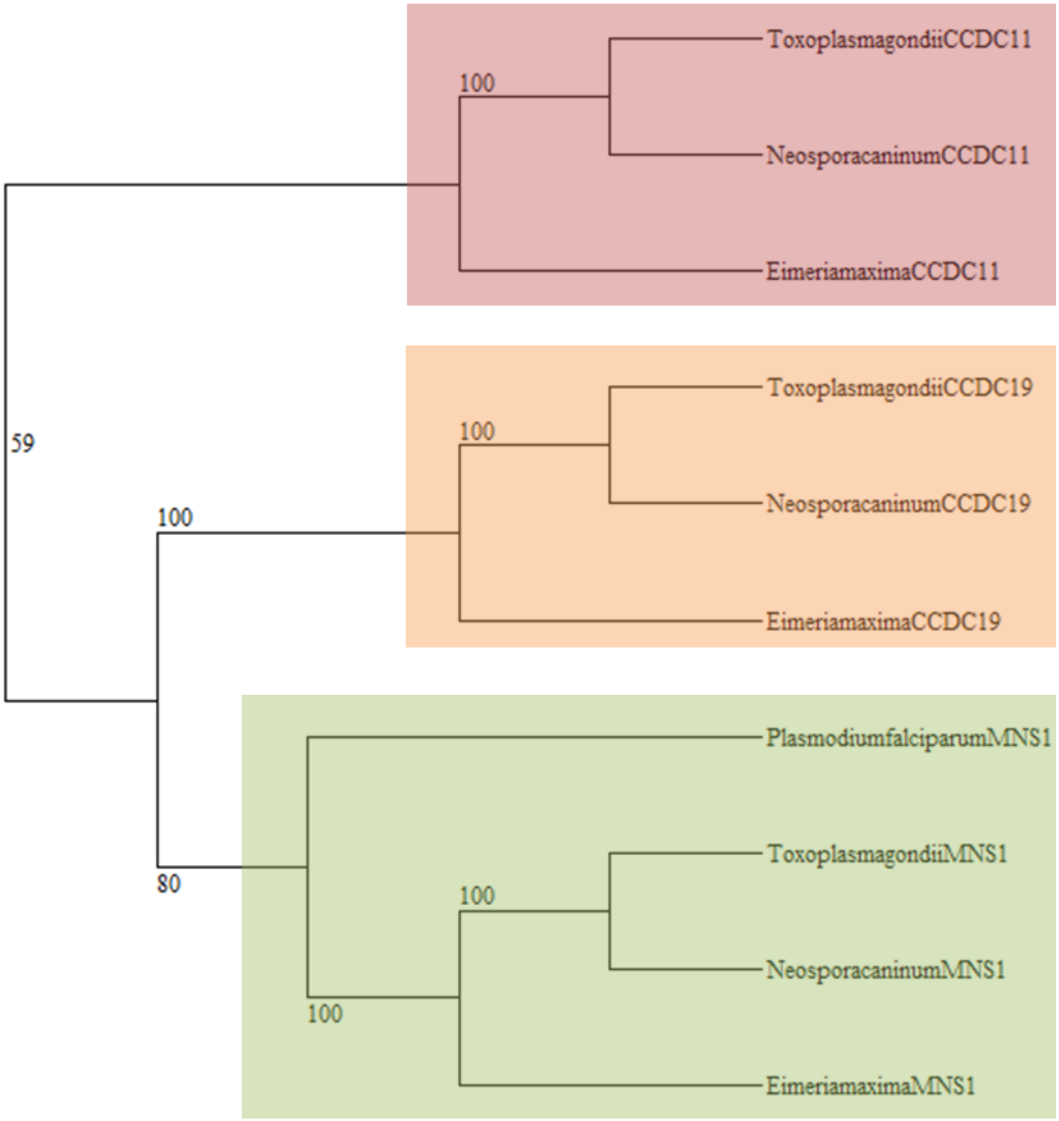


Figure 55: Cladogram of TPH domain-containing proteins within apicomplexa

A cladogram showing the relationship between the TPH domain-containing protein sequences conserved in Apicomplexan. 100 bootstraps were performed; bootstrap values are displayed at the nodes. The clade highlighted in red denotes CCDC11 orthologs, orange for CCDC19 orthologs and green for MNS1 orthologs.

According to the species list on Pfam, there are two TPH domain-containing proteins in *P. falciparum*, an ortholog of MNS1 and another protein that is not orthologous to any of the 7 TPH domain-containing proteins described previously and shall be referred to as a 'cryptic' TPH domain-containing protein.

No orthologs of the cryptic TPH domain-containing protein were detected in apicomplexa by RBH except in *Plasmodium* subspecies (Figure 56). The cryptic TPH domain-containing protein was unanimously conserved across all *Plasmodium* subspecies looked at and the orthologs shared a high sequence identity (Figure 56; C).

To find the relationship between the conventional TPH domain-containing proteins in Apicomplexa and the *Plasmodium* specific cryptic group, a phylogenetic analysis including both sets of proteins was performed (Figure 57). CCDC11 orthologs were excluded from the analysis.

A phylogenetic tree showed that apicomplexan MNS1 orthologs grouped together (green), as did the CCDC19 orthologs (orange) (Figure 57) whilst the cryptic *Plasmodium* specific proteins grouped together (grey). Evidence for the cryptic *Plasmodium* proteins forming a distinct group of TPH domain-containing protein was demonstrated by the length of the separate clade branch in a maximum likelihood phylogenetic tree. A wider BLASTp search did not detect orthologous proteins in other organisms, not even within Apicomplexa (taxid:5794), the *Plasmodium* genes are annotated as being *Plasmodium* specific in PlasmoDB (Aurrecoechea *et al.*, 2009) and no non-plasmodium orthologs are listed in OrthoMCL database (Li *et al.*, 2003).

In conclusion, *P. falciparum* has two TPH domain-containing proteins; an ortholog of MNS1 and an additional 'cryptic' TPH domain-containing protein.

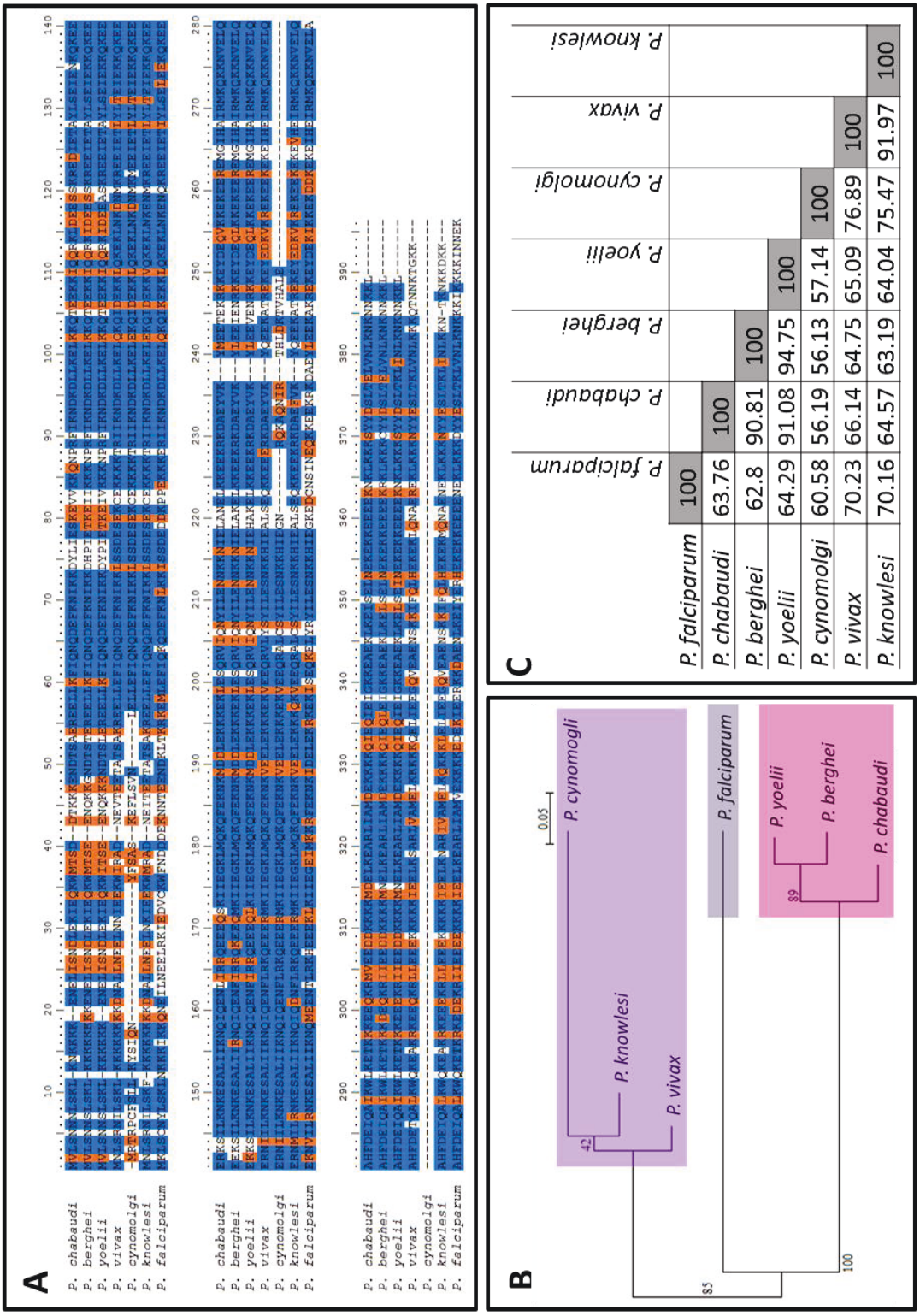


Figure 56: TPH domain-containing proteins in *Plasmodium* spp.

A) An alignment of the *Plasmodium* sequences for the unknown TPH domain-containing protein. Amino acids highlighted in blue are identical, similar amino acids are highlighted in orange. B) A phylogenetic tree showing the relationship between the sequences from different species. Bootstrap values are displayed at the nodes. The clade highlighted in pink denotes rodent infective species, grey for ape infective species and purple indicates monkey infective species. Based on (Hall, 2012) C) a table displaying the percentage similarity between the *Plasmodium* sequences.

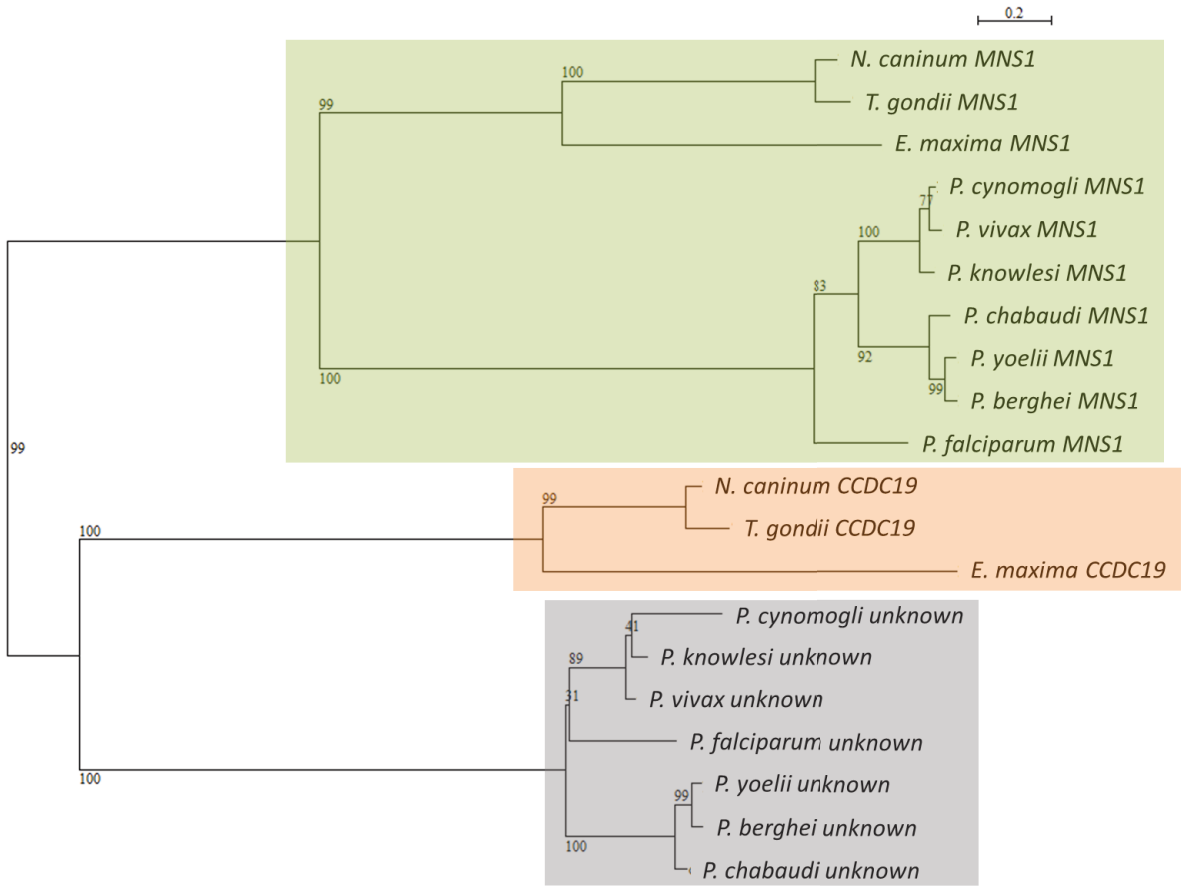


Figure 57: An additional TPH domain-containing protein in *Plasmodium* spp.

A maximum likelihood phylogenetic tree showing the relationship between the TPH domain-containing protein sequences conserved in Apicomplexan 100 bootstraps were performed; bootstrap values are displayed at the nodes. The clade highlighted in green are MNS1 orthologs, orange are CCDC19 orthologs and grey are the 'cryptic' plasmodium specific orthologs.

4.3.15 TPH domain-containing proteins are not universally present in flagella proteomes

As previously described, this project began by trying to identify uncharacterised flagellum proteins, one was TbCCDC19. Since that initial protein, several other proteins were identified that are related to CCDC19 through a shared TPH domain but it is not known if these additional family members had a link to flagellum form or function. To answer this question I analysed forty one relevant flagella/cilia proteomes (see Table 10) to assess the frequency of TPH domain-containing proteins being detected. Results are summarised in Table 18. When more than one copy of an ortholog, paralog or ohnolog is conserved I have indicated that in Table 18 by splitting the row (*Paramecium tetraurelia*) or indicating which protein was detected (*Drosophila melanogaster*).

To identify relevant proteomes, I conducted a review of the literature to find proteomic analyses of cilia, flagella, basal bodies, centrioles and centrosomes as this would provide information about the presence and expression of the TPH domain-containing proteins. This included several sperm proteomes (Amaral *et al.*, 2013; Dorus *et al.*, 2006; Nakachi *et al.*, 2011; Nynca *et al.*, 2014a; Skerget *et al.*, 2013; Wasbrough *et al.*, 2010). Studies were also included that were focussed on the mitotic spindle (Sauer *et al.*, 2005) and microtubule interactors (Gache *et al.*, 2010) as negative controls.

Table 18 shows that TPH domain-containing proteins are not listed in basal body/ centriole proteomes (Balestra *et al.*, 2013; Keller *et al.*, 2005) or the mitotic spindle proteome (Sauer *et al.*, 2005) and not listed in a microtubule interactome (Gache *et al.*, 2010). This indicates they are more likely associated with the flagellum however they are not present in proteomes of immotile cilia/flagella (Ishikawa *et al.*, 2012b; Liu *et al.*, 2007; Mayer *et al.*, 2009; Narita *et al.*, 2012). TPH domain-containing proteins were frequently detected in proteomes of motile cilia (Broadhead *et al.*, 2006; Ostrowski *et al.*, 2002; Pazour *et al.*, 2005) including sperm (Amaral *et al.*, 2013; Skerget *et al.*, 2013) although none were detected in the sperm proteome of rainbow trout (Nynca *et al.*, 2014a). TPH domain-containing proteins were not detected in cilia/flagella membrane proteomes (Oberholzer *et al.*, 2011; Yano *et al.*, 2013), which indicates they are cytoskeletal proteins.

Trichoplein was the only TPH domain-containing protein to be identified as a component of the centrosome (Andersen *et al.*, 2003; Jakobsen *et al.*, 2011), which is inclusive of the PCM and centrioles in *H. sapiens*.

In conclusion, TPH domain-containing proteins are commonly detected in the proteomes of motile flagella and are universally absent from proteomes of immotile flagella. Trichoplein was the only TPH domain-containing protein that was detected in the human centrosome proteome.

Table 18: Occurrence of TPH domain-containing proteins in proteomes.

Relevant proteomes were analysed for the presence of TPH domain-containing proteins. Detection of the protein is indicated by a green box and if the protein was not detected a red box. If the protein was not conserved (NC) in the organism that is indicated by a grey box.

Organism	Focus of Study	CCDC11	CCDC19	CCDC173	FAP98	FAP241	MNS1	TCHP	Reference
Chlamydomonas reinhardtii	Flagella	Yes	Yes	Yes	Yes	Yes	Yes	NC	(Pazour *et al.* , 2005)
	Flagella and basal body	No	No	No	No	No	Yes	NC	(Li *et al.* , 2004)
	Flagella	Yes	Yes	No	No	No	Yes	NC	(Stolc *et al.* , 2005)
	Basal body	No	No	No	No	No	No	NC	(Keller *et al.* , 2005)
Ciona intestinalis	Sperm	Yes	Yes	Yes	NC	NC	Yes	NC	(Nakachi *et al.* , 2011)
Drosophila melanogaster	Centrosome	NC	No	NC	NC	NC	No	NC	(Muller *et al.* , 2010)
	Cilia	NC	Yes	NC	NC	NC	No	NC	(Avidor-Reiss *et al.* , 2004)
	Sperm	NC	CG5062	NC	NC	NC	CG3610	NC	(Dorus *et al.* , 2006)
	Sperm	NC	CG5062	NC	NC	NC	CG3610	NC	(Wasbrough *et al.* , 2010)
	Centrosome	NC	No	NC	NC	NC	No	NC	(Dobbelaere *et al.* , 2008)
Giardia lamblia	Basal body	NC	No	NC	NC	NC	NC	NC	(Lauwaet *et al.* , 2011)
Homo sapiens	Mitotic spindle	No	No	No	NC	NC	No	No	(Sauer *et al.* , 2005)
	Centrosome	No	No	No	NC	NC	No	Yes	(Jakobsen *et al.* , 2011)
	Centrosome	No	No	No	NC	NC	No	Yes	(Andersen *et al.* , 2003)
	Sperm tail	Yes	Yes	No	NC	NC	Yes	No	(Amaral *et al.* , 2013)
	Motile cilia	No	Yes	No	NC	NC	Yes	No	(Ostrowski *et al.* , 2002)
	Sperm head and tail	Yes	Yes	Yes	NC	NC	Yes	No	(Baker *et al.* , 2013)
	Differentiation	No	Yes	No	NC	NC	Yes	No	(Ross *et al.* , 2007)
	Centriole	No	No	No	NC	NC	No	No	(Balestra *et al.* , 2013)
Macaca mulatta	Sperm	Yes	Yes	Yes	NC	NC	Yes	No	(Skerget *et al.* , 2013)
Mus musculus	Immotile cilia	No	No	No	NC	NC	No	No	(Ishikawa *et al.* , 2012)
	Immotile cilia	No	No	No	NC	NC	No	No	(Liu *et al.* , 2007)
	Motile cilia	Yes	Yes	No	NC	NC	Yes	Yes	(Hoh *et al.* , 2012)
	Ciliogensis	No	No	No	NC	NC	No	No	(McClintock *et al.* , 2008)
Naegleria gruberi	Flagellum	NC	Yes	NC	NC	NC	Yes	NC	(Fritz-Laylin and Cande, 2010)
Oncorhynchus mykiss	Sperm	No	No	No	NC	NC	No	No	(Nynca *et al.* , 2014)
Paramecium tetraurelia	Ciliary membrane	No	A) No B) No	A) No B) No	NC	NC	A) No B) No C) No D) No	NC	(Yano *et al.* , 2013)
	Cilia	Yes	A) Yes B) Yes	A) Yes B) Yes	NC	NC	A) Yes B) Yes C) No D) Yes	NC	(Arnaiz *et al.* , 2009)
	Reciliation	Yes	A) No B) No	A) No B) No	NC	NC	A) No B) No C) No D) Yes	NC	(Arnaiz *et al.* , 2010)
Phytophthora infestans	Flagellum	No	No	Yes	NC	NC	No	NC	(Judelson *et al.* , 2012)
Rattus norvegicus	Immotile cilia	No	No	No	NC	NC	No	No	(Mayer *et al.* , 2009)
Sus scrofa	Immotile cilia	No	No	No	NC	NC	No	No	(Narita *et al.* , 2012)
Tetrahymena thermophila	Motile cilia	No	Yes	No	NC	NC	No	NC	(Smith *et al.* , 2005)
	Pellicule	Yes	No	Yes	NC	NC	Yes	NC	(Gould *et al.* , 2011)
	Basal body	No	No	Yes	NC	NC	Yes	NC	(Kilburn *et al.* , 2007)
Trypanosoma brucei	Flagellum	Yes	Yes	NC	NC	NC	Yes	NC	(Broadhead *et al.* , 2006)
	Flagellum	Yes	Yes	NC	NC	NC	Yes	NC	(Hart *et al.* , 2009)
	Flagellum	No	No	NC	NC	NC	No	NC	(Baron *et al.* , 2007)
	Flagellum surface and matrix	No	No	NC	NC	NC	No	NC	(Oberholzer *et al.* , 2011)
Xenopus tropicalis	Epithelium	Yes	Yes	No	NC	NC	No	No	(Hayes *et al.* , 2007)
	Microtubule Interactome	No	No	No	NC	NC	No	No	(Gache *et al.* , 2010)

4.3.16 Predicted protein tertiary structure

Whilst it is important to demonstrate that proteins are orthologous by RBH it does not always mean that the tertiary structure of the protein is similar and therefore the function of the orthologous proteins could have diverged. The work I have carried out *in silico* and presented in this chapter is the foundation for work performed *in vitro* and presented in chapter five of this thesis. With this in mind I used the I-TASSER server to analyse the *H. sapiens* and *T. brucei* amino acid sequences of TPH domain-containing proteins (Figure 58). Similar protein tertiary structures would imply a more likely conserved function and therefore validate the experimental model, *T. brucei*. Human CCDC40 and trichohyalin were also submitted to the prediction software as negative controls. These proteins were chosen as negative controls because CCDC40 is also a highly conserved coiled-coil protein that is involved in flagellum function (Becker-Heck *et al.*, 2011) and trichohyalin shares sequence similarity with the TPH domain.

Predicted protein models are scored according to the predicted confidence in the model (c-score), proteins without a template a c-score usually range between -5 and +2. Values below -5 indicate a low confidence in the predicted model (the confidence values for models predicted in this study are as follows HsCCDC11 c-score of -0.9, TbCCDC11 c-score of -0.88, HsCCDC19 c-score of -2.74, TbCCDC19 c-score of -0.97, HsMNS1 c-score -0.91 and TbMNS1 c-score of -0.99. Whilst HsCCDC40 and HsTrichohyalin scored -0.4 and -0.96, respectively).

The average length of a TPH domain was 323 amino acids, which occupied 65.8% of the whole protein on average. It is therefore not surprising that TPH domain-containing proteins in humans and trypanosomes share a similar predicted tertiary structure (Figure 58).

In conclusion, the *H. sapiens* and *T. brucei* TPH domain-containing proteins share similar predicted tertiary structures, which implies similar function.

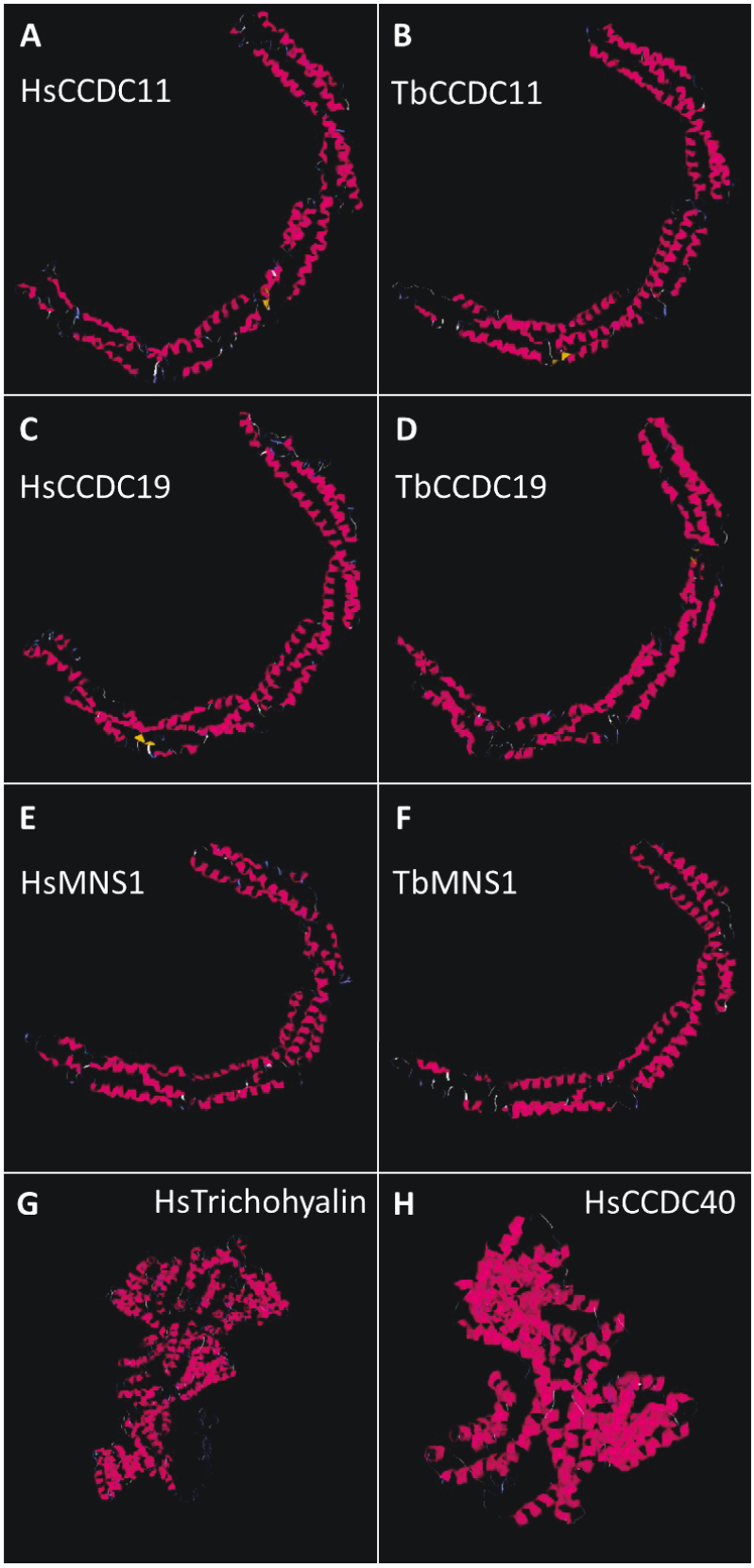


Figure 58: Tertiary structure for TPH domain-containing proteins

Tertiary protein sequences for HsCCDC11 (A), TbCCDC11 (B), HsCCDC19 (C), TbCCDC19 (D), HsMNS1 (E), TbMNS1 (F), HsTrichohyalin (G) and HsCCDC40 (H) were predicted by the I-TASSER sever. The coiled-coil regions are coloured pink.

4.4 Discussion

4.4.1 Conservation of CCDC19

In my initial bioinformatics screen, the protein encoded by Tb927.8.4580, TbCCDC19 was detected because of its presence in the *T. brucei* flagella proteome (Broadhead *et al.*, 2006) and the conservation of an ortholog in humans, HsCCDC19 is also known as nasopharyngeal epithelium specific gene 1 (NESG1). Bioinformatics analysis on CCDC19 orthologs revealed two interesting things. Firstly, the proteins contained a trichohyalin-plectin homology (TPH) domain (see page 148) and secondly, HsCCDC19/NESG1 has been linked to nasopharyngeal cancer (Li *et al.*, 1999; Liu *et al.*, 2011a; Liu *et al.*, 2011b). This is discussed further in chapter 5.

According to my work using the CDD and Pfam databases, the HsCCDC19 protein contained a TPH domain. However, there is no mention of this domain feature in the previous publications on CCDC19/NESG1 (Doudney *et al.*, 2001; Li *et al.*, 1999; Liu *et al.*, 2012; Liu *et al.*, 2011a; Liu *et al.*, 2011b). In the studies of human nasopharyngeal carcinomas, NESG1/HsCCDC19 is not localised at a cellular level so it is not known if HsCCDC19 localises to the cilium or not. Although highly conserved, no work has been performed on CCDC19 orthologs outside of the nasopharyngeal cancer field so despite being detected in several flagellar proteomes CCDC19 has not been localised to the cilium/flagellum in any organisms.

CCDC19 is highly conserved in eukaryotes and in this study orthologs were detected in 31 of the 39 organisms used (Table 17). Seven of the organisms did not form motile flagella and therefore CCDC19 orthologs were expected not to be present. The only organism that does assembled a motile flagellum that did not have a CCDC19 ortholog was *P. falciparum.* This is discussed further in section '4.4.10 TPH domain-containing proteins in apicomplexa'.

Despite the sub-cellular localisation of NESG1/CCDC19 not being published there is a large amount of evidence for CCDC19 being a component of motile flagella within the published proteomes. CCDC19 was not identified in proteomes of immotile cilia (Ishikawa *et al.*, 2012b; Liu *et al.*, 2007; Mayer *et al.*, 2009; Narita *et al.*, 2012) but several other

coiled-coil proteins were. This indicates the methodology used was not discriminative against cytoskeletal, coiled-coil proteins.

Evidence for CCDC19 being a component of motile flagella;

- CCDC19 was detected in several motile cilia/flagella proteomes in *H. sapiens* (Ostrowski *et al.*, 2002) (Amaral *et al.*, 2013; Baker *et al.*, 2013; Ross *et al.*, 2007).
- The *C. reinhardtii* CCDC19 ortholog (FAP45/BUG28) is present in the basal body proteome (Keller *et al.*, 2005) and flagellar proteome (Pazour *et al.*, 2005).
- Transcription of FAP45 is highly dysregulated during flagellar regeneration in *C. reinhardtii* (Stolc *et al.*, 2005).
- CCDC19 was found in the sperm proteome of *C. intestinalis* (Nakachi *et al.*, 2011)
- The *D. melanogaster* CCDC19 ortholog (CG11449) is present in *D. melanogaster* cillia proteome (Avidor-Reiss *et al.*, 2004)
- The sperm proteome of the macaque monkey contained CCDC19 (Skerget *et al.*, 2013)
- CCDC19 is unregulated during ciliogenesis of differentiating epithelial cells in *M. musculus* (Hoh *et al.*, 2012)
- CCDC19 is unregulated during differentiation of *N. gruberi* from the non-flagellated amoebae stage to the flagellated stage
- In *P. tetraurelia* CCDC19 was detected in the motile cilia proteome (Arnaiz *et al.*, 2009) and not in the membrane fraction (Yano *et al.*, 2013).
- CCDC19 was identified in the cilia proteome of the related ciliate, *T. thermophila* (Smith *et al.*, 2005) but not in the basal body (Kilburn *et al.*, 2007).
- CCDC19 is up-regulated during ciliogenesis of epithelial cells in frogs (Hayes *et al.*, 2007).
- TbCCDC19 (Tb927.8.4580) was listed in two flagellar proteomes from *T. brucei* (Broadhead *et al.*, 2006; Hart *et al.*, 2009).

Orthologs of CCDC19 have been detected in several species of organisms called oomycetes (Table 17) that have flagellated zoospores – reinforcing the fact that CCDC19 has a role in cilia/flagella. CCDC19 is not present in the yeast *S. cerevisiae* and S*. pombe* but is present in a group of true fungi, the chytrids (for review see (McLaughlin *et al.*, 2009), which have a flagellated zoospore during their lifecycle (Barr, 1981; Letcher *et al.*,

2006; Letcher *et al.*, 2008). An ortholog of CCDC19 was detected in the chytrid, *B. dendrobatidis,* which has a flagellated stage in its lifecycle (Longcore *et al.*, 1999). This conservation among flagellated fungi further cements the theory that CCDC19 is intrinsically linked to the structure and/or function of flagella and this claim links in with the proposal that the last eukaryotic common ancestor (LECA) had a flagellum (Wickstead and Gull, 2011).

4.4.2 Conservation of CCDC11

CCDC11 was not in the original list of candidates (Table 14) but was subsequently included for analysis due to the TPH domain within the protein sequence (Figure 40). CCDC11 was highly conserved with 24 out of the 39 organisms analysed having a CCDC11 ortholog. Seven of the organisms were negative controls so it was expected that CCDC11 would not be conserved. That left 32 flagellated organisms with 24 orthologs detected.

No ortholog for CCDC11 was detected using either RBH or a more general BLASTp search in the chicken, *G. gallus*. This was surprising because CCDC11 orthologs had been found in other chordata. Therefore the search parameters were widened to look at the proteomes of other bird species to see if orthologs of CCDC11 were present. An ortholog of CCDC11 was detected in the zebrafinch (*Taeniopygia guttata*). It is unknown if CCDC11 is truly not conserved in the chicken or if its absence was an artefact of faulty genome assembly or proteome translation.

Other flagellated organisms that lack CCDC11 orthologs include the flagellated chytrid *B. dendrobatidis,* the fruit fly, *D. melanogaster,* the spider *S. mimosarum,* the diatom *T. pseudonana* and the protozoa *G. lamblia, N. gruberi* and *P. falciparum. D. melanogaster* and *P. falciparum* are both known to have unusual examples of cilia biology (Briggs *et al.*, 2004b; Gottardo *et al.*, 2013; Riparbelli *et al.*, 2013), which could account for their lack of CCDC11 orthologs. *G. lamblia* and *N. gruberi* are both protozoa which are found near the base of the eukaryotic tree and therefore CCDC11 may have occurred in organisms which diverged later in evolution or CCDC11 may have been lost in these organisms separately. To accurately build a picture of CCDC11 occurance and loss, genome sequencing would need to be focussed on more unicellular eukaryotes. The diatom *T. pseudonana* is

unusual because sequencing of the genome revealed no central pair components (Armbrust *et al.*, 2004). An electron microscopic study of *T. lacustris* reveals that the structure of the cells motile flagellum is 9+0 (instead of 9+2), it does not have a central pair, only inner and outer dynein arms (Idei *et al.*, 2013). It is assumed that the flagellum of *T. pseudonana* is also 9+0 but this has not been experimentally proven.

There is a clinical report of a patient with a mutation within CCDC11 that caused *situs inversus* but examination of tracheal cilia and sperm found normal axonemes (Perles *et al.*, 2012). CCDC11 has recently been localised at a sub-cellular level to cilia and basal bodies (Narasimhan *et al.*, 2014) (This will be discussed in more detail in section 5.5 Discussion).

4.4.3 Conservation of MNS1

MNS1 is a coiled-coil protein that was first characterised in mouse sperm cells (Furukawa *et al.*, 1994). That is why its name includes the term 'meiosis'. It is now known that this protein is not meiosis specific. MNS1 has been localised to sperm flagella in mice (Lehti *et al.*, 2013; Vadnais *et al.*, 2014; Zhou *et al.*, 2012) but has not localised to the cilia/flagellum of other organisms or indeed to other cilia/flagella within the mouse.

In my analysis of the 39 reference organisms, 29 contained an ortholog of MNS1. Discounting the 7 negative controls, the remaining 3 organisms that are flagellated but without an MNS1 ortholog are; B*. dendrobatidis, D. melanogaster,* and *G. lamblia.* The lack of conservation of MNS1 within *D. melanogaster* is discussed in section '4.4.9 TPH domain-containing proteins in Arthropods'. The only TPH domain-containing protein conserved in B*. dendrobatidis* and *G. lamblia* is CCDC19. Without better ultrastructural studies on the flagella of these organisms it is hard to speculate why CCDC19 is conserved and other TPH domain-containing proteins are not.

The phylogenetic analysis of MNS1 orthologs presented some unexpected groupings. The sequence for *Apis mellifera* MNS1 did not group with the other insecta sequences (Figure 44), instead it shared a common branch with the MNS1 sequence of the apicomplexan, *Plasmodium falciparum.* In the phylogenetic analysis of all 7 TPH domain-containing

protein families (Figure 45), *A. mellifera* MNS1 does group with other insect sequences but the *P. falciparum* ortholog is situated on an individual branch of the MNS1 clade. In a phylogenetic analysis of just arthropod TPH domain-containing protein families (Figure 53) *A. mellifera* MNS1 groups in the MNS1 clade as expected but when additional *Drosophila* species sequences are included in the phylogenetic analysis, the position of *A. mellifera* MNS1 is displaced (Figure 54). This is not an unheared of phenomenon in bioinformatics. I have shown that the *A. mellifera* sequence (XP_006563783) is an MNS1 ortholog using reciprocal best blast hit approach and both Pfam and CDD detect a TPH domain within the protein sequence. Therefore I am confident that the *A. mellifera* sequence is a MNS1 ortholog, however, it is influenced and displaced by other sequences in phylogenetic trees for reasons other than incorrect orthology.

The *P. falciparum* MNS1 ortholog (XP_001349839), as mentioned above, is positioned strangely in the analysis of eukaryotic MNS1 orthologs despite other apicomplexan sequences being included in the analysis (Figure 44). In an isolated analysis of apicomplexan sequences, the *P. falciparum* MNS1 sequence does group with other MNS1 orthologs (Figure 55) and this is strengthened by the inclusion of other *Plasmodium* species MNS1 sequences (Figure 56). Pfam and CDD both detect a TPH domain within the protein sequence (XP_001349839; Appendix 7) and my analysis of MNS1 orthologs using reciprocal best blast hit approach detected this sequence within *P. falciparum* therefore I am confident that there is a true MNS1 ortholog within *P. falciparum.*

4.4.4 Conservation of CCDC173

CCDC173 was only conserved in 16 of the 32 flagellated organisms used for bioinformatics analysis (Table 17). As expected, CCDC173 was not present in the 7 non-flagellated organisms. See section ‘4.3.6 TPH domain-containing proteins are absent from organisms that lack motile cilia’.

Orthologs of CCDC173 were present in chordates but absent from arthropods (Figure 47). As well as being conserved in multi cellular eukaryotes CCDC173 orthologs were conserved in single celled organisms. CCDC173 was detected in oomycetes, ciliates and chlamydomonadales but not conserved in kinetoplastids or apicomplexa. Although this is

an unusual pattern of conservation, it is not unprecedented. The conservation of ζ tubulin is also unexpected due to conservation in kinetoplastids, oomycetes and ciliates but not in fungi or apicomplexa (Findeisen *et al.*, 2014).

There is no direct published work on CCDC173, only indirect work within the proteomes analysed here (Table 17). CCDC173 is present in 9 of the 41 proteomes that were examined (Table 18). Eleven of the proteomes were in organisms that did not possess an ortholog of CCDC173, meaning that CCDC173 had the potential to occur in only 30 of the proteomes. The proteomes where CCDC173 orthologs were not detected (but were conserved) are from centrosome, centriole and immotile cilia focussed studies. The 9 proteomes that detected CCDC173 are all from motile cilia/flagella, which indicates that CCDC173 has a role specific to motile axonemes.

CCDC173 was universally conserved in chordates, universally absent from arthropods and exhibited a strange conservation pattern in single celled eukaryotes. This could indicate that CCDC173 was lost from these eukaryotes in separate events, which has been previously documented for the loss of cilia/cilia proteins (Hodges *et al.*, 2011).

4.4.5 TPH domain-containing proteins in algae

FAP98 and FAP241 were only detected in *Chlamydomonas reinhardtii* and *Volvox carteri* (Figure 48), which are both algae from the class chlorophyceae. No other hits were detected during RBH analysis, not even highly divergent ones. It appears that conservation of FAP98 and FAP241 are restricted to chlorophyceae as no hits were detected in organisms outside of chlorophyceae. The lack of orthologs in other chlorophyceae could be due to a lack of genome sequencing within this group of organisms. In evolutionary terms there is still a gap between the unicellular algae, *C. reinhardtii* and the multicellular algae, *V. carteri*.

It is interesting that both FAP98 and FAP241 protein families are only detected in *C. reinhardtii* and *V. carteri.* To investigate if these protein families are truly two distinct subfamilies of the TPH domain-containing protein group, the sequences of FAP98 and FAP241 were compared to each other within the same organism. CrFAP98 and CrFAP241

share only 22.2% amino acid identity and VcFAP98 and VcFAP241 are only 33.6% identical. Another piece of evidence that FAP98 and FAP241 are not divergent orthologs of other TPH domain-containing proteins is the fact that *C. reinhardtii* and *V. carteri* have a total of 6 TPH domain-containing proteins, they are only lacking an ortholog of trichoplein, which is not conserved outside of metazoans. Therefore I am satisfied that FAP98 and FAP241 are truly distinct protein families.

4.4.6 Conservation of trichoplein

Trichoplein is the original TPH domain-containing protein where the TPH domain was first identified (Figure 59). The TPH domain was named due to sequence similarity with both plectin and trichohyalin (Nishizawa *et al.*, 2005).

Trichoplein has been listed as a component of the centrosome (Andersen *et al.*, 2003) but more recently an in-depth proteomic analysis of the human centrosome proposed that trichoplein is part of the mitotic spindle apparatus (Jakobsen *et al.*, 2011) with a GFP fused version of trichoplein localising to microtubules during mitosis (Jakobsen *et al.*, 2011). Trichoplein has no ortholog in *T. brucei,* although this is not surprising as the basal body of *T. brucei* does not participate in nuclear mitosis like the mammalian centriole does.

Trichoplein is also called mitostatin (TpMs) and has been described as a tumour suppressor gene in a range of cancer types (Cerqua *et al.*, 2010; Fassan *et al.*, 2011; Kim *et al.*, 2010; Vecchione *et al.*, 2009). Disruption of mitostatin allegedly leads to mitochondrial fragmentation although this phenotype has not been reported in recent studies (Kasahara *et al.*, 2014). Mitostatin has been localised to mitochondria (Cerqua *et al.*, 2010). However, the mitochondrial localisation of mitostatin (Cerqua *et al.*, 2010) conflicts with the centriole/centrosome localisation of trichoplein (Jakobsen *et al.*, 2011) and it is not clear how the two proteins (trichoplein and mitostatin) have originated, possibly due to alternative splicing.

Trichoplein was not listed in any of the flagella proteomes analysed (Table 18) but was detected in a transcriptional profiling experiment of differentiating mouse epithelial cells.

Trichoplein was up-regulated at the time point corresponding to basal body formation with levels decreasing in the time frame of cilia assembly (Hoh *et al.*, 2012). This indicates that trichoplein has a role associated with the centriole/basal body and not the axoneme of cilia/flagella. I found no trichoplein orthologs in single celled eukaryotes (Figure 50).

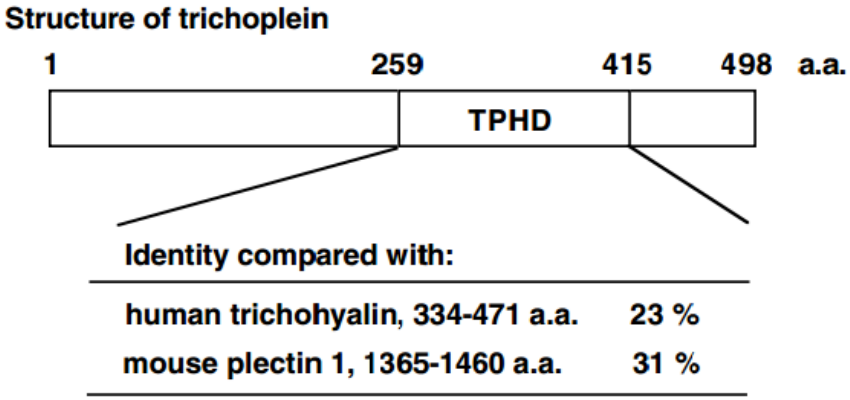


Figure 59: Structure of trichoplein

A diagram of the protein, Trichoplein including the trichohyalin-plectin homology (TPH) domain and its identity to regions of human trichohyalin (23%) and mouse plectin (31%). Diagram taken from (Nishizawa *et al.*, 2005)

Figure 60 summarises published interactions for trichoplein. Trichoplein interacts with ninein and outer dense fiber protein 2 (Odf2)/cenexin (Ibi *et al.*, 2011), which are known centrosomal/centriole components (Chen and Megraw, 2013; Ou *et al.*, 2002). Trichoplein has been found to interact strongly with keratins 8, 16 and 18 (Nishizawa *et al.*, 2005) and keratin 15 (Rual *et al.*, 2005). Other intermediate filaments that trichoplein interacts with include vimentin (Rual *et al.*, 2005) but Nishizawa and colleagues (2005) did not detect an interaction between trichoplein and vimentin. This suggests the interaction is weak or species/cell type specific. Trichoplein also interacts with Aurora kinase A to control primary cilia assembly (Inoko *et al.*, 2012). This pathway has been expanded and now it is known that trichoplein activates Aurora-A and therefore supresses axoneme extension. The feedback loop between trichoplein, Aurora-A and K+ channel tetramerisation domain-containing 17 (KCTD17) has been extensively investigated *in vitro* (Kasahara et al., 2014) (Figure 60).

4.4.7 Protein: protein interactions

The proteins containing a TPH domain are a highly conserved family of coiled-coil proteins. Proteins with coiled-coil domains are numerous and have varying roles, which is not surprising considering the generalised proposed function of coiled-coil domains is protein-protein binding (Lupas, 1996). It is established that proteins with coiled-coil domains have important structural roles and many have been found to localise at the centriole/basal body or at the cilia/flagella (Becker-Heck *et al.*, 2011; Fry *et al.*, 1998; Joo *et al.*, 2013; Yang *et al.*, 2002). Several coiled-coil domain-containing (CCDC) proteins are in the centrosome proteome (Jakobsen *et al.*, 2011).

Published literature on TPH domain-containing proteins and relevant interactome studies were queried to investigate any possible connections between TPH domain-containing proteins and known flagellum proteins. I collated this interaction data from eight separate publications and summarised the interactions individual publications are demonstrated by the use of different coloured lines within the interaction map (Figure 60). The majority of published interactions included in Figure 60 are just published simplistically as 'known interactions'; these are illustrated with a straight connecting line.

It is interesting that MNS1 and CCDC11 have a common interactor, kinesin family member 3A (KIF3A), identified by co-immunoprecipitation (Lehti *et al.*, 2013). CCDC19 was not detected. The authors speculate that KIF3A could be responsible for transporting MNS1 via IFT (Lehti *et al.*, 2013). It is known that MNS1 also interacts with itself *in vitro* to form dimers (Zhou *et al.*, 2012), which is indicated with an orange arrow (Figure 60). MNS1 has also been reported as interacting with mitofusin-2 (MFN2), which has been localised to the flagellum of boar sperm cells (Flores *et al.*, 2010).

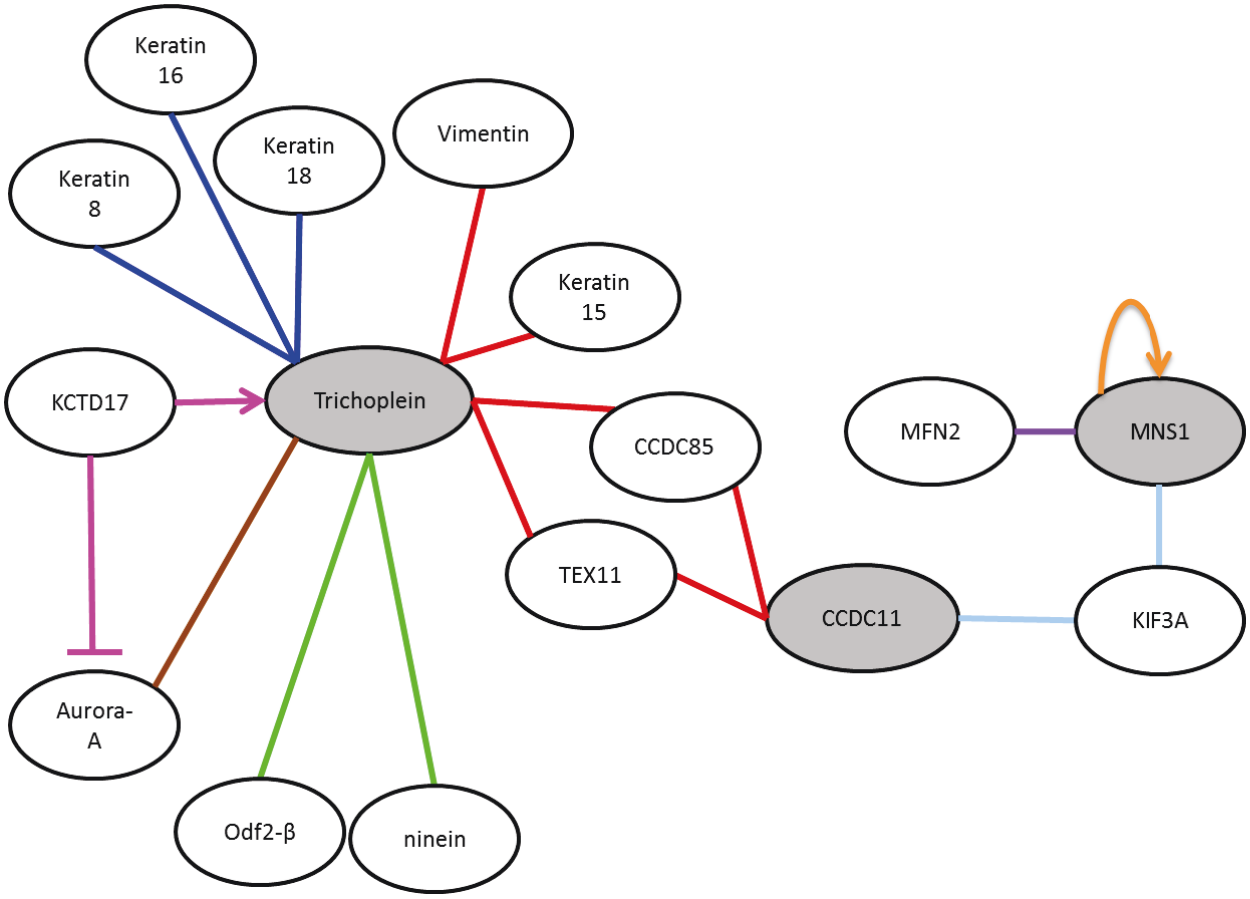


Figure 60 Published protein: protein interactions

Protein interaction data was collated from published literature. Interactions from separate publicationS were assigned a different coloured line – red (Rual *et al.*, 2005), purple (Vadnais *et al.*, 2014), orange (Zhou *et al.*, 2012), light blue (Lehti *et al.*, 2013), pink (Kasahara *et al.*, 2014), dark blue (Nishizawa *et al.*, 2005), green (Ibi *et al.*, 2011) and brown (Inoko *et al.*, 2012). TPH domain-containing proteins are highlighted in grey ovals with interacting proteins in white ovals. Abbreviations used in this figure; K^+ channel tetramerisation domain-containing 17 (KCTD17), outer dense fibre protein 2-β (Odf2-β), testis expressed 11 (TEX11), coiled-coil domain-containing protein 85 (CCDC85), mitofusin 2 (MFN2), kinesin family member 3A (KIF3A). Interactions between proteins are experimentally proven using yeast two-hybrid screens (Nishizawa *et al.*, 2005; Rual *et al.*, 2005; Zhou *et al.*, 2012) and/or co-immunoprecipitation (Ibi *et al.*, 2011; Inoko *et al.*, 2012; Kasahara *et al.*, 2014; Lehti *et al.*, 2013; Nishizawa *et al.*, 2005; Vadnais *et al.*, 2014). No publications with interaction data were found for CCDC19, FAP98, FAP241 or CCDC173 proteins.

4.4.8 TPH domain-containing proteins in ciliates

To investigate the conservation of TPH domain-containing protein I included two ciliates in the list of organisms (Table 9); *Paramecium tetraurelia* and *Tetrahymena thermophila.* These two ciliates were chosen as their genomes have been sequenced and they are both important experiment models used to study cilia assembly, structure and function.

It is known that *Paramecium tetraurelia* has undergone at least 3 whole genome duplications through evolution resulting in 39,642 protein coding genes (Aury *et al.*, 2006). This could have created up to 8 ohnologs for CCDC11, CCDC19, CCDC173 and MNS1 within *P. tetraurelia*. So when looking for TPH domain-containing proteins conserved within *P. tetraurelia* it was interesting to see what ohnologs had been retained or lost from the whole genome duplications. In fact there is evidence to suggest that when more than one ohnolog is retained, they are both transcribed (Plattner, 2010). This could be a mechanism to ensure that a large amount of a protein is available in a short time frame. In the case of *P. tetraurelia* that could represent the high demand for cilia proteins during ciliation.

Within *P. tetraurelia* a single copy of CCDC11 is present whilst CCDC19 and CCDC173 exist as two ohnologs. There are 4 ohnologs of MNS1. This is interesting because research shows there is a link between the level of gene expression and the likelihood of ohnolog retention (Gout *et al.*, 2010). By this premise, CCDC11 is the lowest expressed and MNS1 the highest expressed genes, which could indicate an important function for MNS1 within *P. tetraurelia.*

T. thermophila had a single copy of CCDC11, CCDC19, CCDC173 and MNS1, which confirms the duplication events are *Paramecium* specific. Neither *P. tetraurelia* nor *T. thermophila* contained an ortholog for FAP98, FAP241 or trichoplein.

4.4.9 TPH domain-containing proteins in Arthropods

As part of my 39 reference organisms I included 6 arthropods because of their importance as experimental models, both in cilia biology and in evolutionary biology. My choices of organisms were restricted to those whose genomes are sequenced. Reference arthopods

used in this study included one arachnid, *Stegodyphus mimosarum* and five insects *Drosophila melanogaster* (diptera), *Anopheles gambiae* (diptera), *Tribolium castaneum* (coleopteran), *Apis mellifera* (hymenoptera) and *Bombyx mori* (lepidoptera). No arthropods possessed orthologs of CCDC173, FAP98 or PAF241.

CCDC11 orthologs were not detected in the fruit fly, *D. melanogaster* or the spider, *S. mimosarum*, despite CCDC11 orthologs being detected in other arthropods (Figure 53). To date there is only one complete spider genome for *S. mimosarum* and a partially assembled draft genome for a second spider, *Acanthoscurria geniculate* (Sanggaard *et al.*, 2014). I did not detect a CCDC11 ortholog in *A. geniculate* either. Due to a lack of genomes sequenced from arachnids it is hard to prove if CCDC11 is truly not conserved in arachnids. In addition to more spider genome sequencing, it would be informative for ultrastructural studies to focus on cilia/flagella of spider cells as some species have an unusual 9+3 microtubule arrangement in the axoneme of sperm (Baccetti *et al.*, 1970).

All 6 arthropoods possessed an ortholog of CCDC19 but *D. melanogaster* had an additional CCDC19-like protein (Figure 53 and Figure 54). It is not clear why *D. melanogaster* has this additional divergent CCDC19 protein but it was detected in two sperm proteomes (Dorus *et al.*, 2006; Wasbrough *et al.*, 2010) and therefore appears to be connected to cilia/flagella. The direct ortholog of CCDC19 was identified in a bioinformatics analysis searching for genes involved in ciliogenesis (Avidor-Reiss *et al.*, 2004). The only motile flagella present in insects are the flagella of the sperm cells, which can be structurally divergent from sperm cells of chordata with microtubule arrangements of 9+0, 9+3, 9+1 and 9+9+1 being found within diptera alone (Jamieson *et al.*, 1999).

In addition to a CCDC19-like protein, *D. melanogaster* had two MNS1-like proteins. Using a reciprocal best blast hit approach; *D. melanogaster* had no detectable direct ortholog of MNS1. However, the species tree function within Pfam listed four *D. melanogaster* proteins that contained a TPH domain. When I entered these proteins as a seed sequence of a BLASTp search, the results contained MNS1 orthologs of other species. Therefore *D. melanogaster* has duplicate MNS1-like proteins that are a feature restricted to *Drosophila*

species (Figure 54). Direct MNS1 orthologs were found in the 5 other arthropods used in this study (Figure 53).

Using a reciprocal best blast hit approach I found a trichoplein ortholog in *A. gambiae, B. mori* and *S. mimosarum* but when the protein sequences are entered as a search query in Pfam or CDD, no TPH domain is detected. The protein sequence from *S. mimosarum* (KFM76999) is annotated on NCBI as 'partial', which could account for this lack of detectable TPH domain as it is 362 amino acids in length and the trichoplein sequence of *T. castaneum* is 504 amino acids. The *B. mori s*equence (XP_004928902) is annotated within NCBI as 'trichoplein' by an unknown annotator.

4.4.10 TPH domain-containing proteins in apicomplexa

Apicomplexa are an important group of parasitic protozoa that are responsible for several human and animal infectious diseases (Darde *et al.*, 2008; Rider and Zhu, 2008). The centriole of apicomplexa is unusual as it consists of 9 outer singlet microtubules and a single central microtubule (for reviews see (Francia and Striepen, 2014; Morrissette and Sibley, 2002).

No TPH domain-containing proteins were detected in *C. parvum,* which does not assemble a flagellum during its lifecycle and therefore this result fits my hypothesis of TPH domain-containing proteins being linked to motile cilia/flagella. Orthologs of CCDC173, FAP98, FAP241 and trichoplein were not detected in apicomplexa. *P. falciparum, E. maxima, T. gondii* and *N. caninum* all contained orthologs of MNS1 (Table 17).

Orthologs of CCDC11 were detected in *E. maxima, T. gondii* and *N. caninum* but not in *P. falciparum* (Table 17). The motile flagellum of *P. falciparum* is unusual, as it is assembled independently of IFT (Briggs *et al.*, 2004a), however, it is still assembled in the canonical 9+2 microtubule configuration (Sinden *et al.*, 1978) and many known flagella proteins are conserved (Carvalho-Santos *et al.*, 2010). In addition to lacking CCDC11, *P. falciparum* does not have a CCDC19 ortholog (Table 17). It is not clear why *P. falciparum* lacks an

ortholog of CCDC19 because CCDC19 is highly conserved across eukaryotes and is present in other apicomplexa species that also form flagella during the lifecycle.

However, *P. falciparum* possess a TPH domain-containing protein that is specific to plasmodium species (Figure 56) that I have termed the 'cryptic' TPH domain-containing protein. The cryptic TPH domain-containing protein could be linked to the unusual assembly of the *P. falciparum* flagellum compared to other eukaryotes. It is possible that the cryptic TPH domain-containing protein of *Plasmodium sp.* is a highly divergent ortholog of CCDC19.

4.4.11 Albatross

During my review of the literature I came across a protein called Fas-binding factor 1 (FBF-1), which binds to Fas, a tumour necrosis factor receptor (TNFR) (Yanagisawa *et al.*, 1997). FBF-1 is also known as Albatross and has been localised to the cytoplasm (Schmidt *et al.*, 2000) and adherens junctions between cells (Sugimoto *et al.*, 2008).

Albatross has been described as containing a TPH domain at amino acids 625-910 (285 in length) (Sugimoto *et al.*, 2008). Further examination revealed that Albatross is in fact, not a bona fide member of the TPH domain-containing protein family as no TPH domain could be detected by Pfam or CDD (Figure 51). Further evidence of a TPH domain being absent from Albatross is shown in Figure 61 and described below.

The original TPH domain was characterised in the protein trichoplein (Nishizawa *et al.*, 2005). The TPH domain within trichoplein has 23% sequence identity with amino acids 334-471 of human trichohyalin (Figure 61, A; red section) and 31% sequence identity with amino acids 1365-1460 of mouse plectin 1 (Figure 61, B; red section).

The proposed TPH domain within Albatross had been compared to different sections of trichohyalin and plectin (Figure 61; blue sections) to that of the original criteria for a TPH domain (Nishizawa *et al.*, 2005). The false TPH domain within albatross has 25% sequence identity with amino acids 123-369 of human trichohyalin (Figure 61, A; blue section) and

22% sequence identity with amino acids 2073-2401 of mouse plectin 1 (Figure 61, B; blue section).

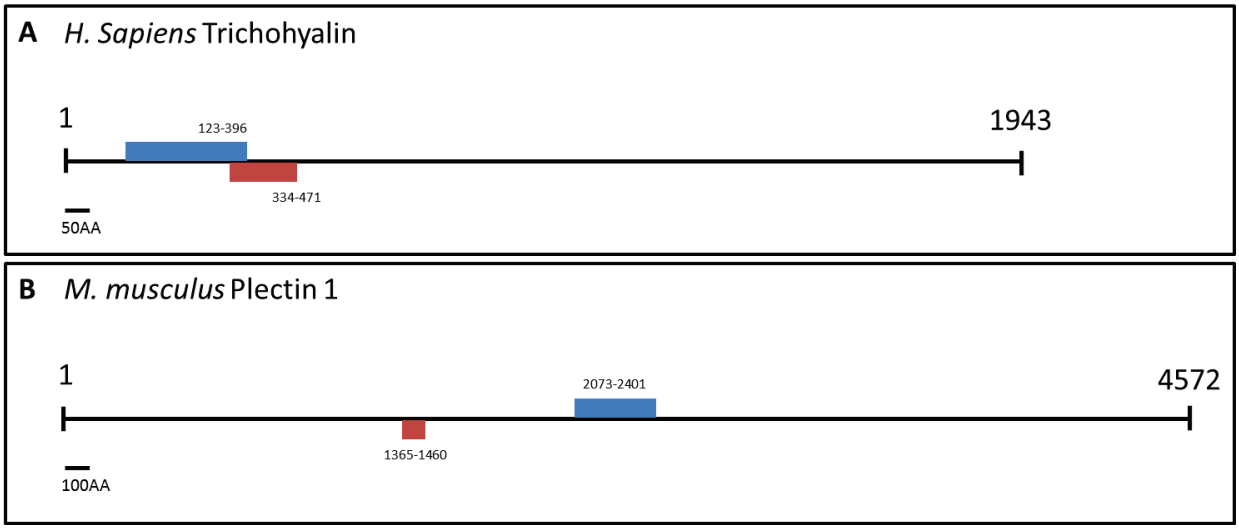


Figure 61: Protein architecture of trichohyalin and plectin 1

Schematics of (A) *H.sapiens* trichohyalin and (B) *M. musculus* Plectin 1 proteins. Areas indicated in red are the original criteria used to discover the TPH domain (Nishizawa *et al.*, 2005). Areas indicated in blue are used to imply that Albatross contains a TPH domain (Sugimoto *et al.*, 2008).

Therefore, using the original criteria published for identification of a TPH domain (Nishizawa *et al.*, 2005), Albatross does not have a bona fide TPH domain.

4.4.12 Proteome analysis

Despite the high number of published proteomes available my analysis was hindered by the absence of a 'complete set' of proteomes, meaning a proteome of immotile cilia, motile cilia and basal body/centriole from the same organism. The analysis included 41 proteomes that focused on sperm cells, immotile cilia, motile cilia, centriole and centrosomes from a wide range of organisms (Table 18). Looking at the absence or presence of proteins in immotile cilia would be more informative if there existed an equivalent motile cilia proteome, ascertained using the same methodology, to compare between in the same organism.

There are several published sperm proteomes, presumably due to the ease of sample extraction and lack of tissue dissection needed. These can be extremely informative

especially when presented in a compartmentalised way as it can help pinpoint if the protein of interest is present in the flagellum, sperm head or semen (Baker *et al.*, 2013; Nynca *et al.*, 2014a; Nynca *et al.*, 2014b). TPH domain-containing proteins were detected in all except one sperm proteome. This exception was the proteome of rainbow trout sperm (Nynca *et al.*, 2014a) despite orthologs of CCDC11, CCDC19, CCDC173, MNS1 and trichoplein being conserved in *O. mykiss*. This sperm proteome was quite small with only 206 significant proteins being included in the final list, which included the cytoskeletal proteins tektin and WDR16. The sperm proteomes of other chordata are considerably larger with 1429 proteins in *H. sapiens* sperm (Baker *et al.*, 2013) and 1247 proteins in M*. mulatta* sperm (Skerget *et al.*, 2013).

In the *P*. *tetraulia* cilia membrane proteome, of the 267 detected proteins, only 3 were conserved in *T. brucei* (Yano *et al.*, 2013). As expected, none of the TPH domain-containing proteins were detected in the membrane fraction. The *D. melanogaster* ortholog of CCDC19 (CG11449) was found in a screen for genes essential for ciliogenesis (Avidor-Reiss *et al.*, 2004) but is not present in the centrosome proteome (Muller *et al.*, 2010), which would have included the centriole therefore this suggests that the DmCCDC19 also has a axoneme specific localisation rather than centriole/basal body function. Although I would not expect CCDC19 to be detected in the giardia basal body proteome, it is interesting to note that CCDC39, CCDC146 and CCDC147 were present. This indicates that sample extraction was effective for preservation of cytoskeletal and coiled-coil proteins.

It is likely that TPH domain-containing proteins are axonemal proteins because of their conservation in the genomes of organisms with cilia/flagella and because of their detection in motile cilia/flagella proteomes. It is unlikely that TPH domain-containing proteins are basal body proteins because the proteins are not present in basal body proteomes (Kilburn *et al.*, 2007; Lauwaet *et al.*, 2011) of organisms that are known to have these proteins. The exception is trichoplein as this was detected in the human centrosome (Andersen *et al.*, 2003; Jakobsen *et al.*, 2011), which is likely due to being present at the centriole (Jakobsen *et al.*, 2011; Kasahara *et al.*, 2014).

The most interesting observation is that TPH domain-containing proteins were not found in proteomics studies of primary/sensory cilia (Ishikawa *et al.*, 2012b; Liu *et al.*, 2007; Mayer *et al.*, 2009; Narita *et al.*, 2012), only in motile cilia. This pattern supports the hypothesis that TPH domain-containing are linked to axonemal motility.

4.4.13 A possible role for TPH domain-containing proteins in *T. brucei*

As stated previously, a function of the TPH domain in trichoplein has been found to be to a keratin binding domain (Nishizawa *et al.*, 2005). This is particularly interesting in trypanosome biology because the genome does not encode a gene for keratin (Berriman *et al.*, 2005). In fact no genomic evidence of intermediate filaments was found in *T. brucei* or the related kinetoplastids, *T. cruzi* and *Leishmania major* (El-Sayed *et al.*, 2005; Ivens *et al.*, 2005). However, there are genes within the TriTryp database (Aslett *et al.*, 2010) that have been annotated as 'putative trichohyalin' (Tb927.11.11480) and 'plectin-like' (TcCLB.503897.110) (Appendix 10). Although TcCLB.503897.110 appears to not have a syntenic ortholog in *T. brucei*. It is not clear who annotated these accessions with plectin/trichohyalin and the bioinformatics evidence is not provided. These annotations taken together with the confirmed presence of TPH domain-containing proteins within *T. brucei* could suggest that of a group of divergent, intermediate-filament type proteins does exist within *T. brucei* and possibly other protozoa. There is now precedent for unusual conservation of proteins that are highly divergent (Akiyoshi and Gull, 2014).

Indeed, a lamin-like protein has been discovered in *T. brucei* (DuBois *et al.*, 2012) and another unconventional intermediate filament-type protein in *D. discoideum* (Batsios *et al.*, 2012) when it was previously thought that intermediate filament proteins were absent from protozoa. Incorrect annotation of the genome could explain this because genes that were deemed 'unlikely' orthologs such as predicted proteins that were missing expected domains were not annotated as orthologs. This may have been the case for divergent intermediate filament related proteins.

It is known that intermediate filaments within host cells are important for establishing parasitic infection. For instance, keratin is involved in the attachment of *Trichomonas vaginalis* to the vaginal epithelium (Vilela and Benchimol, 2011). It has been proposed

that intermediate filament binding proteins may be involved in host cell invasion by intracellular parasites such as *T. cruzi* (Butler and Tyler, 2012) but the mammalian infective stages of *T. brucei* remain extracellular in the bloodstream of the mammalian host so intermediate filament binding proteins must have another role in this lifecycle form or presumably they would have been shed from the genome. An important stage within the *T. brucei* lifecycle is for the epimastigote form of the parasite to attach to the epithelia cells within the insect vector salivary glands (Beattie and Gull, 1997; Tetley and Vickerman, 1985) where the attachment plaques have been described as 'hemidesmosome-like' (Bauer, 1984; Kohl and Bastin, 2005) but no components of these plaques have been identified on a molecular level. Similar adhesion plaques have also been visualised for *Leishmania* spp. (Wakid and Bates, 2004) so these attachment plaques are conserved within kinetoplastids and presumably the components of these plaques will be highly similar between kinetoplastid species.

It is not clear if kinetoplastid TPH domain-containing proteins could perform a role in mammalian host cell invasion. If this were the case then intracellular infective parasites, such as *T. cruzi* or *Leishmania* spp. would provide a better model to explore this possibility. When keratin is knocked down in host mammalian cells *T. cruzi* were still able to invade but the amastigote forms once inside the cell could not divide properly in the absence of keratin (Claser *et al.*, 2008). It is possible that the TPH domain-containing proteins within trypanosomes are the proteins that bind to the keratin in host cells or the epithelia of the insect vector in epimastigote form (Figure 9).

4.4.14 Conclusions

The work presented in this chapter shows that I have used extensive bioinformatics analysis to identify seven distinct sub-families of proteins that contain a trichohyalin plectin homology (TPH) domain, which constitutes an unpublished family of proteins. In addition to the seven highly conserved families there are two additional restricted groups of TPH domain-containing proteins; the cryptic TPH domain-containing protein of *Plasmodium* spp. and the MNS1-like (A and B) family in *D. melanogaster*.

I have also investigated the conservation of TPH domain-containing proteins across a range of eukaryotes and from this I can conclude that TPH domain-containing proteins are restricted to organisms that build a motile flagellum/cilium during their lifecycle. This hypothesis is strengthened by the examination of published proteomic studies (Table 18). Out of the TPH domain-containing proteins only two, MNS1 and trichoplein, are published as containing a TPH domain (Nishizawa *et al.*, 2005; Vadnais *et al.*, 2014) and no one else has proposed that these proteins exist as a family with a common domain.

As well as identifying a family of proteins widely conserved among eukaryotes I have identified 3 genes that are conserved within *T. brucei* that encode for proteins that contain a TPH domain; TbCCDC19, TbMNS1, TbCCDC11.

5. Characterisation of Trypanosomal TPH domain-containing proteins

5.1 Collaborations

Some of the work presented in this chapter was performed in collaboration with Robbie Crickley (RC), Oxford Brookes University. Procyclic form cells were grown, fixed, immunolabelled and imaged by SB. RC measured the perimeter and surface area of cells, data are presented in Figure 71, Figure 73 and Figure 75. Measurement of flagellum length and posterior end length was performed by SB. Numerical data were processed and figures made by SB.

5.2 Introduction

In chapter three, a family of proteins that each contained a trichohyalin-plectin homology (TPH) domain was identified. There were 7 sub families; CCDC11, CCDC19, CCDC173, FAP98, FAP241, MNS1 and trichoplein. I investigated the conservation of these 7 sub families across eukaryotes and found that these proteins were highly conserved in eukaryotes that built motile cilia/flagella. TPH domain-containing proteins were not conserved in eukaryotes that did not assemble cilia/flagella or those that only built immotile cilia/flagella.

I identified three TPH domain-containing proteins that were conserved within *Trypanosoma brucei*, TbCCDC19, TbCCDC11 and TbMNS1. These were all detected in the *T. brucei* flagella proteome (Broadhead *et al.*, 2006). Of the three TPH domain-containing protein families conserved in *T. brucei*, only MNS1 has previously been localised to a flagellum in mouse sperm (Zhou *et al.*, 2012).

5.3 Aims

The aim of the work presented in chapter is to investigate the localisation and function of all three *T. brucei* TPH domain-containing proteins. *T. brucei* is an ideal model organism to study flagellum proteins as the flagellum remains intact through the cell cycle and the new flagellum grows alongside the old flagellum and each flagellum is identifiable as the old or the new. This work is the first evidence that an ortholog of CCDC19 is localised to a motile flagellum and the first time an ortholog of MNS1 to be localised to the motile flagellum of a cell type that can undergo mitosis. MNS1 has previous only been localised to male gamete cells of mice (Zhou *et al.*, 2012). CCDC11 has been localised to the basal body of zebrafish (Narasimhan *et al.*, 2014).

The research aims of this work are:

- To evaluate the localisation profiles of candidate proteins, identified through bioinformatics, using endogenous expression methods
- To ascertain the function of these proteins through RNAi and/or ectopic expression of the protein within *T. brucei*

5.4 Results

5.4.1 Endogenous expression of all three TPH domain-containing proteins localise to the flagellum in *T. brucei.*

To study each of the three TPH domain-containing proteins in *T. brucei*, I generated procyclic cell lines that expressed a YFP fusion version of TbCCDC11 (Tb927.5.1230), TbCCDC19 (Tb927.8.4580) or TbMNS1 (Tb927.6.4520) from one endogenous locus. Each protein was tagged with YFP at the N terminus using the pPOT system (Dean *et al.*, 2015) or the pEnT5 system (Kelly *et al.*, 2007). See section 2.3.1 Primer design and plasmid construction.

To determine the localisation of the three TPH domain-containing proteins, cells were detergent extracted, fixed in -20°C methanol and immunolabelled with an anti-PFR2 antibody, L8C4 (Kohl *et al.*, 1999) and treated with DAPI to label the DNA containing organelles (panels not shown). See section '2.6.1 Light microscopy'. The cell cycle stages of *T. brucei* are annotated as xKxN where xK is the number of kinetoplasts in the cell and xN is the number of nuclei. The kinetoplast contains the mitochondrial genomic DNA and has a separate DNA synthesis phase to the nucleus.

Fluorescence microscopy showed that each of the YFP::TPH domain-containing fusion proteins localised to the flagellum in procyclic *T. brucei* in detergent extracted cytoskeletons (Figure 62, Figure 28, Figure 64) and whole cells (data not shown). The localisation of YFP::TbCCDC11, YFP::TbCCDC19 and YFP::TbMNS1 remained visible after detergent extraction. Therefore TPH domain-containing proteins are cytoskeletal proteins.

Recently, a new method for PCR tagging was developed for the model organism *T. brucei* (Dean *et al.*, 2015). This long primer PCR approach was used to modify one endogenous Tb927.5.1230 allele so that a fusion protein of TbCCDC11, tagged with YFP at the N terminus, would be constitutively synthesised. YFP::TbCCDC11 localised to the flagellum in G1 phase cells, known as 1K1N (Figure 62; A). This stage of the cell cycle is identified by the presence of a single nucleus (blue arrowhead) and a single flagellum, which is subtended by a single set of basal bodies (BB; Figure 62; D, black arrows) and a single

kinetoplast. As the cell cycle continues the kinetoplast DNA undergoes S-phase and the basal bodies duplicate. The next distinct stage in the cell cycle identifiable by light microscopy is known as 2K1N (Figure 62; B). At this stage the cell still has a single nucleus but the kinetoplast DNA has divided and the basal bodies have segregated, with each mature basal body assembling a flagellum. The cell now has 2 flagella, known as the old flagellum (OF) and new flagellum (NF). The new flagellum grows alongside the old flagellum and is always positioned more posterior to the old flagellum (Robinson *et al.*, 1995). The YFP::TbCCDC11 signal is present in the NF as it extends (Figure 62; B and C).

The final stage in the cell cycle is a 2K2N cell where the nuclear DNA has undergone S phase and mitosis to create two nuclei (blue arrowheads). In a 2K2N cell (Figure 62; C) the new flagellum has extended to almost its full length prior to cytokinesis.

The YFP signal was found at the proximal end of the flagellum (Figure 62; E, bracket) close to the basal body (Figure 62; D) but the PFR immunolabelling began at a position more distal to the endogenous YFP signal (Figure 62; F). Using the software ImageJ, a standard deviation projection was performed on the YFP and L8C4 inset panels (Figure 62; G), which produced a combined image view showing the common pixels (Figure 62; G, green) between the YFP::TbCCDC11 signal (Figure 62; G, yellow) and the PFR immunolabelling (Figure 62; G, red). This showed that there were common and unique pixels from the YFP and PFR signal and therefore YFP::TbCCDC11 does not co-localise with the protein, PFR2, the protein immunolabelled by L8C4.

L8C4 is an established antibody for the immunolabelling of PFR2 protein in the PFR lattice (Kohl *et al.*, 1999) throughout the cell cycle. The PFR structure is not present along the whole length of the flagellum, it is only present once the axoneme has exited the flagella pocket of the cell and extends to the distal tip of the flagellum (Bastin *et al.*, 1999a; Bastin *et al.*, 1999b). In detergent extracted cytoskeletons, the flagellar pocket is no longer present due to its membranous composition but the gap between the basal bodies and the L8C4 label is clear (Figure 62; F). The 9+2 microtubule structure of the axoneme is continuous from tubules A and B of the triplet basal bodies. The YFP::TbCCDC11 does not localise at the basal bodies (Figure 62; D and E, BB) but does begin in accordance with the start of the outer doublet microtubules, at a position more proximal to the L8C4

immunolabelling. The offset in the signals from YFP and L8C4 (Figure 62; E, F, G) strongly suggests that TbCCDC11 is present in the axoneme rather than the PFR.

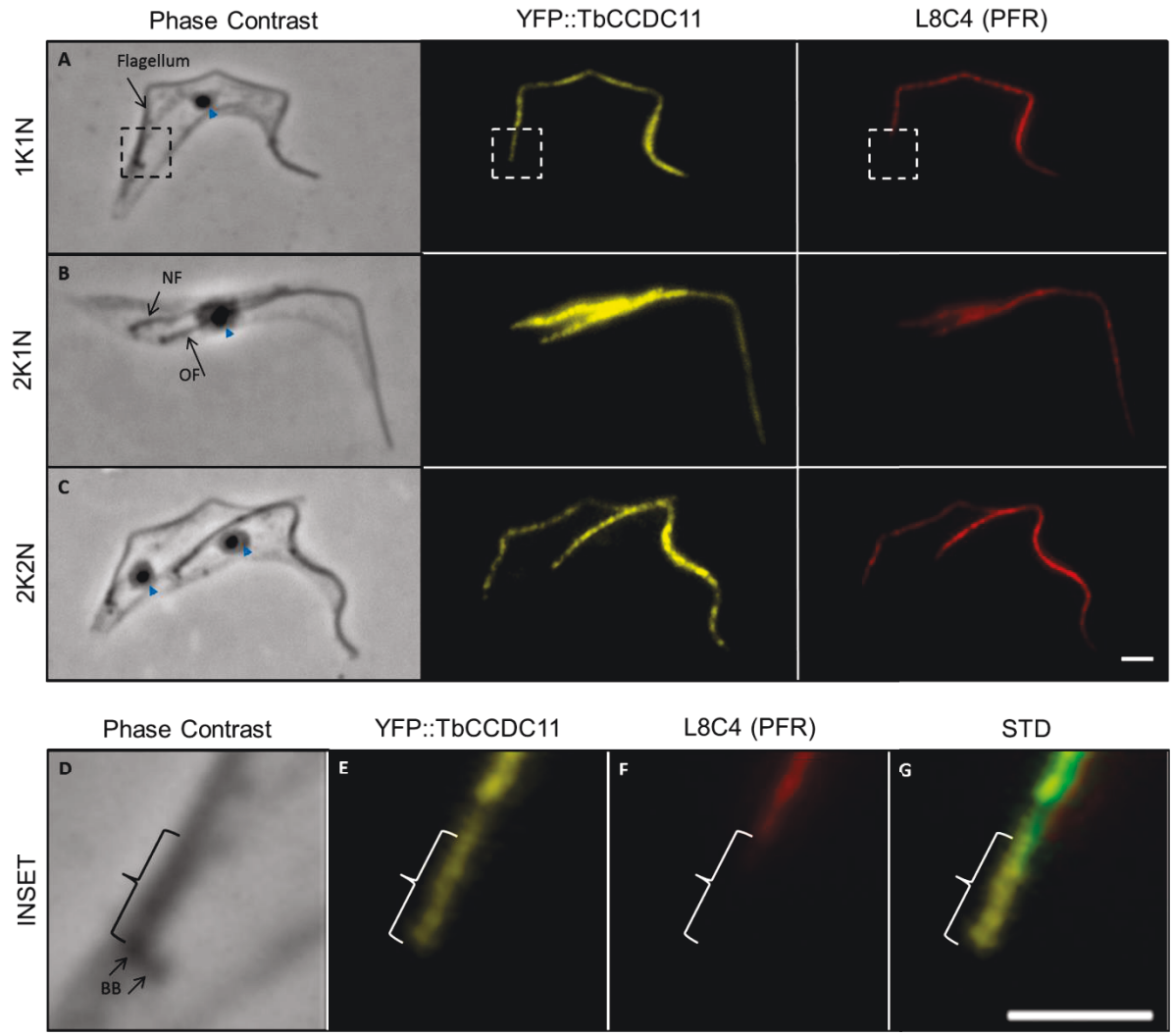


Figure 62: TbCCDC11 localises to the flagellum throughout the cell cycle

TbCCDC11 was tagged at the N terminus with YFP (YFP::TbCCDC11; yellow) and expressed from the endogenous locus in procyclic *T. brucei*. Cells were detergent extracted, fixed and immunolabelled with L8C4 to label the PFR. The YFP::TbCCDC11 signal can be seen localised to the flagellum in 1K1N (A), 2K1N (B) and 2K2N (C) cells. Insets show the proximal end of the flagellum from the 1K1N cell (A). (D) Phase contrast image of the basal bodies and the proximal end of the flagellum. (E) YFP::TbCCDC11 signal, which is localised to the flagellum but not the basal bodies. (F) Immunolabelling of the PFR with L8C4. This signal does not label the proximal end of the flagellum (bracket). (G) A z-projection of the standard deviation (STD) between panels E and F in ImageJ, shared pixels are represented in green. NF; new flagellum, OF; old flagellum, BB; basal bodies. Blue arrowheads; nucleus. Scale bars = 2µm.

The TbCCDC19 protein was tagged with YFP at the N terminus using the pEnT5 vector system (Kelly *et al.*, 2007) to replace one endogenous locus. This plasmid was stably transfected into the SmOx parental cell line (Poon *et al.*, 2012) and YFP:TbCCDC19 was expressed under the native promoter (promoter unknown). Similarly to YFP::TbCCDC11, YFP::TbCCDC19 also has a localisation offset from the PFR immunolabelling (Figure 28; E-G) and localises to the old and new flagella throughout the cell cycle (Figure 28; A-C). The inset images (Figure 28; D-G) from the 1K1N cell (Figure 28; A) demonstrate that the YFP::TbCCDC19 is absent from the basal bodies (BB; Figure 28; D and E, arrows). The YFP::TbCCDC19 (Figure 28; E) signal is evident at a position more proximal on the flagellum to the L8C4 label (Figure 28; F), which suggests YFP::TbCCDC19 had an axonemal localisation rather than PFR. The absence of labelling from the triplet microtubule basal body suggests TbCCDC19 may have a function connected to the outer doublet microtubules in the axoneme.

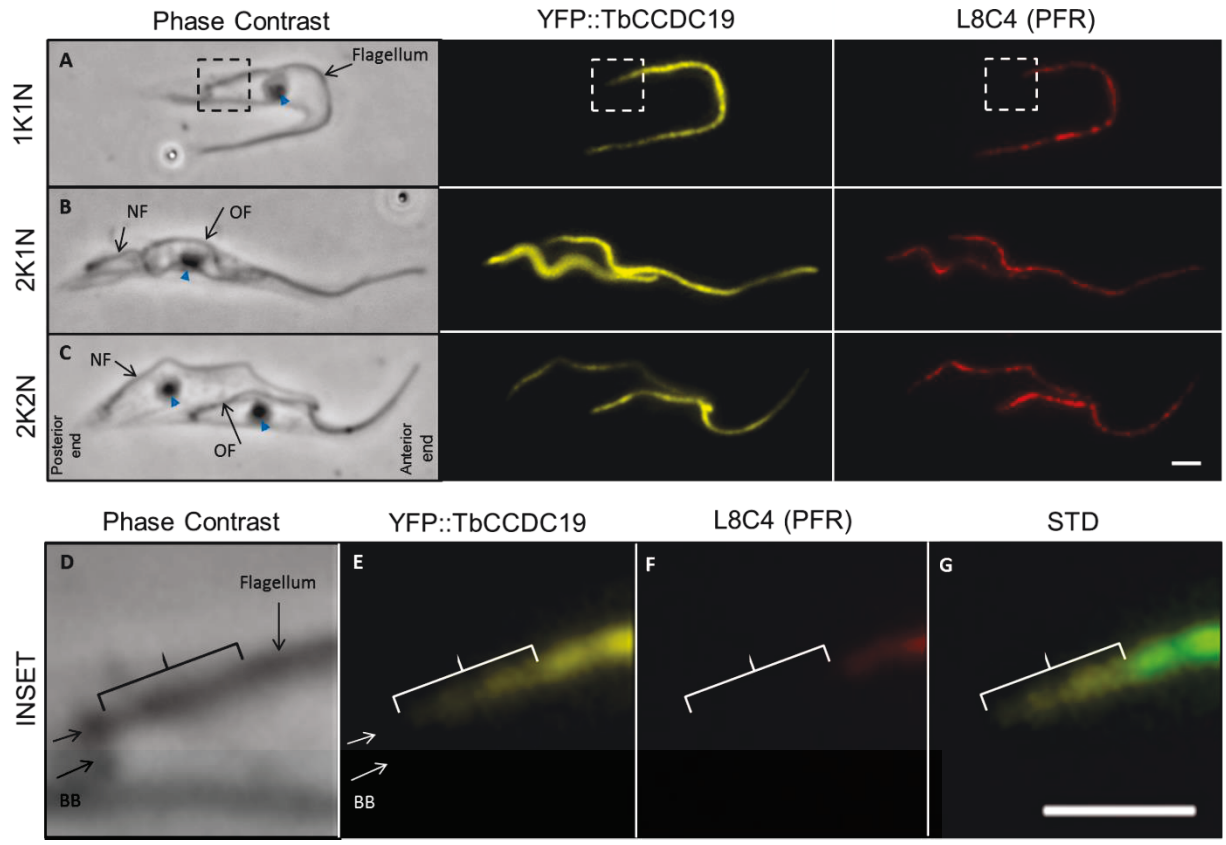


Figure 63: TbCCDC19 localises to the flagellum throughout the cell cycle

TbCCDC19 was tagged at the N terminus with YFP (YFP::TbCCDC19; yellow) and expressed from the endogenous locus in procyclic *T. brucei*. Cells were detergent extracted, fixed and immunolabelled with L8C4 to label the PFR. The YFP::TbCCDC19 signal can be seen localised to the flagellum in 1K1N (A), 2K1N (B) and 2K2N (C) cells. Insets (D-G) show the proximal end of the flagellum from the 1K1N cell (A). (D) Phase contrast image of the basal bodies (BB) and the proximal end of the flagellum (black bracket). (E) YFP::TbCCDC19 signal, which is localised to the flagellum but not the basal bodies (BB; white arrows). (F) Immunolabelling of the PFR with L8C4. This signal does not label the proximal end of the flagellum (white bracket). (G) A z-projection of the standard deviation (STD) between panels E and F in ImageJ, shared pixels are represented in green. NF; new flagellum, OF; old flagellum, BB; basal bodies. Blue arrowheads; nucleus. Scale bars = 2µm.

TbMNS1 was tagged with YFP at the N terminus using the pPOT system (Dean *et al.*, 2015) to target Tb927.6.4520 from one endogenous locus. Homologous recombination of the PCR product modified one allele so expression of YFP::TbMNS1 was constitutively active. YFP::TbMNS1 localised to the flagellum of *T. brucei*. The YFP::TbMNS1 signal was visible on the old flagellum and new flagellum of cells with 2 flagella (Figure 64; B, 2K1N and C, 2K2N). This implies that YFP::TbMNS1 is stably incorporated into the flagellum as it grows and does not have cell cycle dependant expression. All three trypanosomal TPH domain-containing proteins displayed a similar localisation. YFP::TbMNS1 also had a localisation offset from the L8C4 immunolabelling of PFR2. The standard deviation projection (Figure 64; G) produced from the YFP and L8C4 inset panels (Figure 64; E and F, respectively) showed that there were common pixels (Figure 64; G, green) but the PFR immunolabelling (Figure 64; F) did not fully overlap with the YFP::TbMNS1 signal (Figure 64; E).

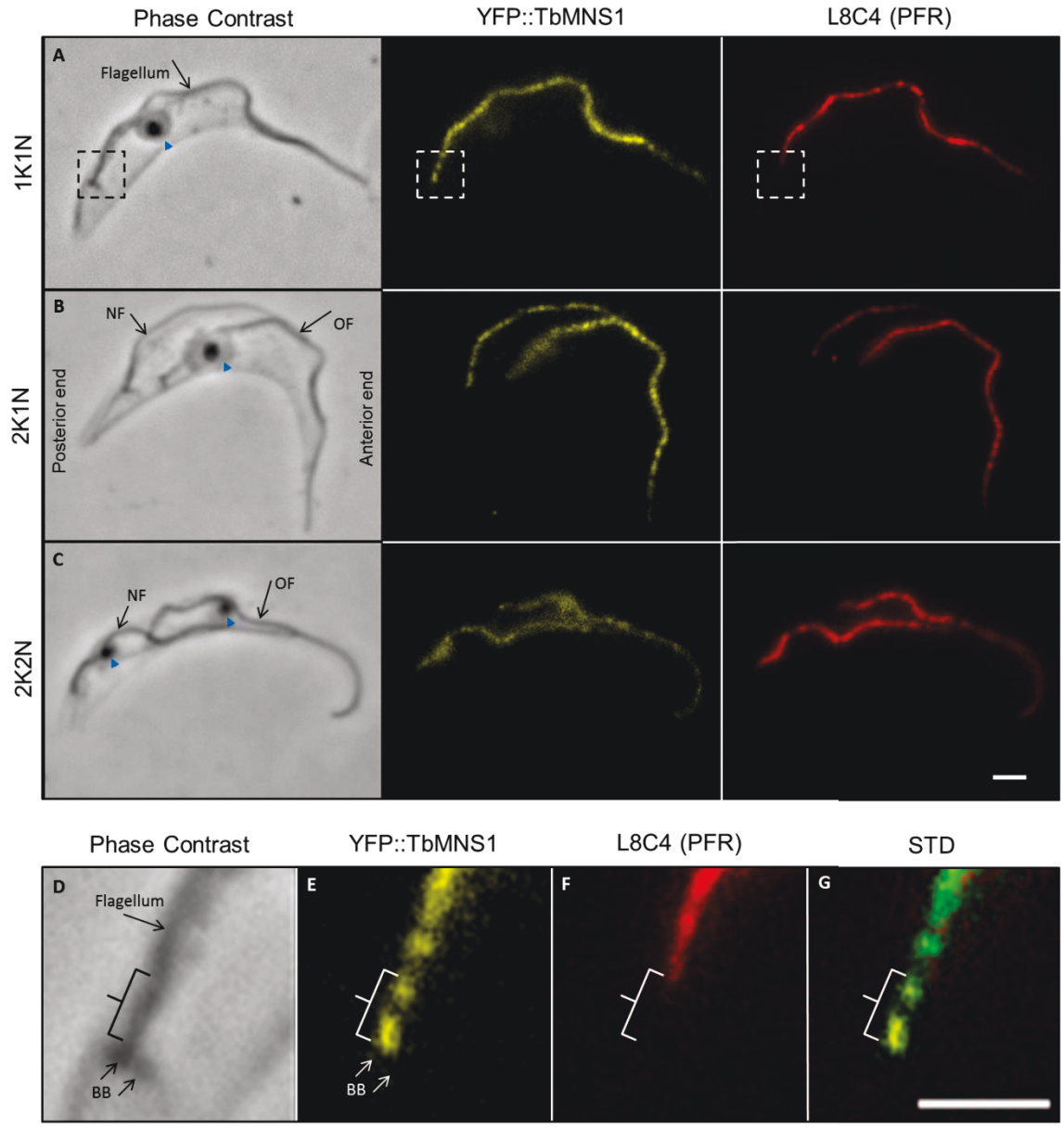


Figure 64: TbMNS1 localises to the flagellum throughout the cell cycle

TbMNS1 was tagged at the N terminus with YFP (YFP::TbMNS1; yellow) and expressed from the endogenous locus in procyclic *T. brucei*. Cells were detergent extracted, fixed and immunolabelled with L8C4 to label the PFR. The YFP::TbMNS1 signal can be seen localised to the flagellum in 1K1N (A), 2K1N (B) and 2K2N (C) cells. Insets (D-G) show the proximal end of the flagellum from the 1K1N cell (A). (D) Phase contrast image of the basal bodies (BB) and the proximal end of the flagellum (black bracket). (E) YFP::TbCCDC19 signal, which is localised to the flagellum but not the basal bodies (BB; white arrows). (F) Immunolabelling of the PFR with L8C4. This signal does not label the proximal end of the flagellum (white bracket). (G) A z-projection of the standard deviation (STD) between panels E and F in ImageJ, shared pixels are represented in green. NF; new flagellum, OF; old flagellum, BB; basal bodies. Blue arrowheads; nucleus. Scale bars = 2µm.

5.4.2 Knockdown by RNAi of TbCCDC11 affects the cell cycle in *T. brucei.*

To assess the function of TbCCDC11, the Tb927.5.1230 gene was targeted for knockdown by RNA interference (RNAi). A stable procyclic cell line was generated that constitutively expressed YFP::TbCCDC11 from one endogenous locus and was also transfected with an inducible p2T7-177 RNAi vector (Wickstead et al., 2002) against Tb927.5.1230.

Cell growth was assessed for cell populations post induction of RNAi and compared to the growth of the non-induced cell population (see section 2.4.10 Continuous growth curve). The culture density was measured every 24 hours for 72 hours (Figure 65; A). Figure 65 (A) shows the continuous growth for a population of non-induced (black) and induced (red) populations of the YFP::TbCCDC11/p2T7_TbCCDC11 cell line. This shows that RNAi of TbCCDC11 has an effect on cell growth because the RNAi induced cells do not grow as well as the non-induced population. However, depletion of TbCCDC11 is not lethal to the cells.

To assess any disruption to normal cell cycle progression the DNA containing organelles were counted in a population of cells from an asynchronous culture (induced or non-induced flask) (see section 2.4.11 Cell cycle analysis). To stain the DNA, the cells were treated with DAPI. This permitted visualisation of the kinetoplast and nucleus within whole cells. For each cell line, the DNA containing organelles were quantified by preparing a slide sample of cells per time point throughout the RNAi induction. Normal cell types fall into the 1K1N (blue), 2K1N (red) and 2K2N (green) categories. The 'others' category (purple) represents abnormal cell types including zoids (1K0N), 1K2N, 0K1N and multinucleate cells. The DNA containing organelles were quantified through the induction (Figure 65; B). In the non-induced population of TbCCDC11 2.22% (n=8) were abnormal cell types, referred to as 'others'. Through the induction, by 72 hours this had increased to 15.14% (n=56) of the population. There was a slight decrease in 1K1N cells from 78.39% at 0 hours to 71.08% at 72 hours post induction.

Knockdown of TbCCDC11 was assessed by quantifying the percentage of cells in the population that had a YFP signal, which remained localised to the flagellum. The percentage of cells in the non-induced population that had a YFP::TbCCDC11 signal along

the flagellum was 95.7% (n=580), which decreased to 15.5% (n=459) 72 hours post induction of RNAi against TbCCDC11.

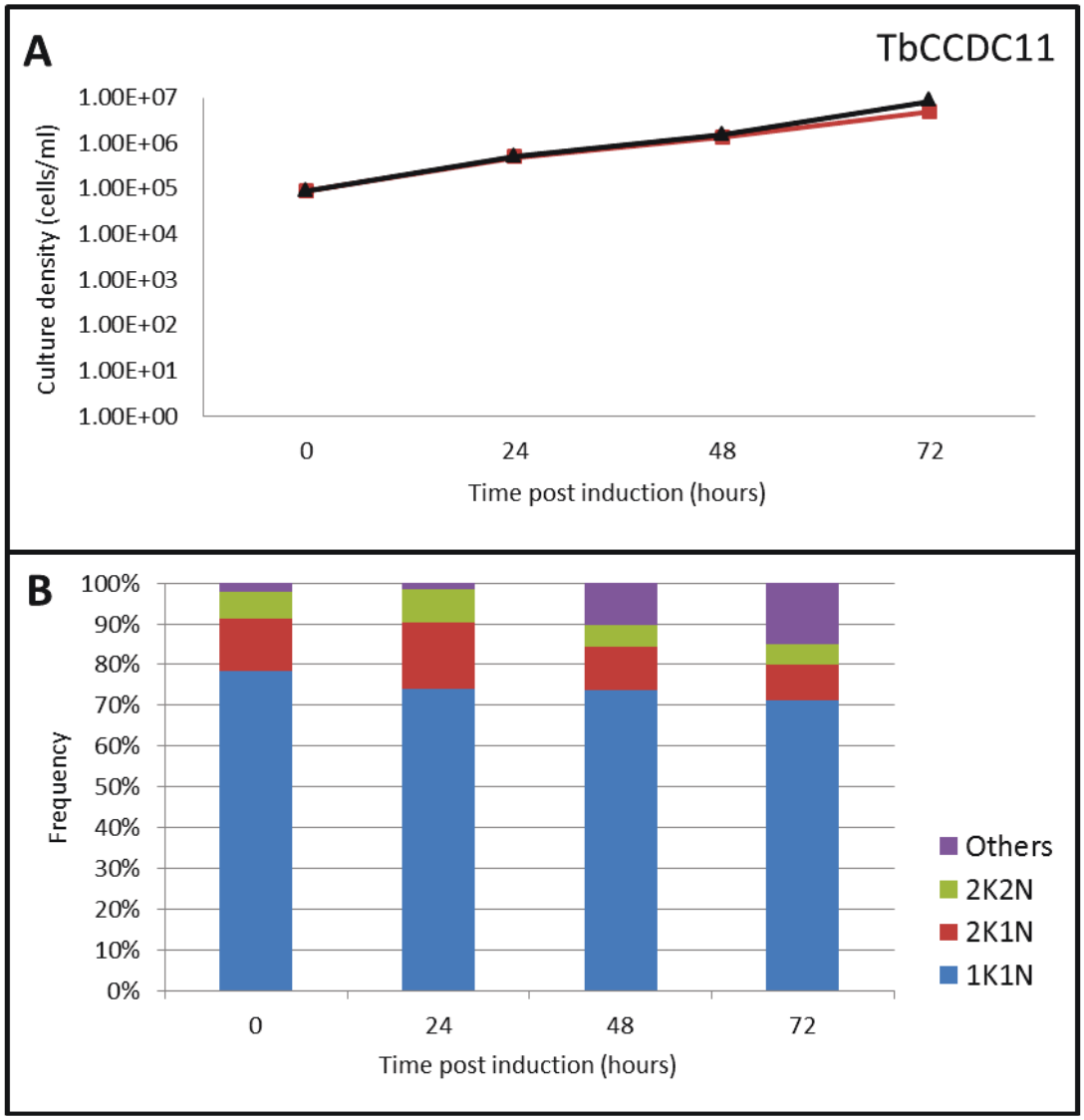


Figure 65: RNAi of TbCCDC11

(A) Continuous growth curve of non-induced (black) and induced (red) populations of TbCCDC11. (B) Quantification of DNA containing organelles in 0 , 24, 48 and 72 hours post induction (n= 361, 365, 413 and 370, respectively). Cell types were classified as 1K1N (blue), 2K1N (red), 2K2N (green) or abnormal 'others' (purple). Kinetoplast = K, nucleus = N.

5.4.3 Knockdown by RNAi of TbCCDC19 affects the cell cycle in *T. brucei.*

To assess the function of TbCCDC19, the Tb927.8.4580 gene was targeted for knockdown by RNA interference (RNAi) using an inducible vector, p2T7-177 (Wickstead et al., 2002). The RNAi vector against Tb927.8.4580 was transfected into the same cell line that constitutively expressed YFP::TbCCDC19 from one endogenous locus.

Cell growth was assessed by plotting a continuous growth curve (see section 2.4.10 Continuous growth curve). This showed that knockdown of TbCCDC19 has a detrimental effect on *T. brucei* cell growth by 72 hours post induction (Figure 66; A) but ablation of TbCCDC19 is not lethal to procyclic cells.

Cells were treated with DAPI to visualise the kinetoplast and nucleus (see section 2.4.11 Cell cycle analysis). Quantification of DNA containing organelles (Figure 66; B) shows a significant decrease in 1K1N cells (blue) from 90.25% of the population at 0 hours (non-induced) to only 64.13% by 72 hours post induction. This decrease in 1K1N cells is mirrored by an increase in other abnormal cell types (purple) from 0.5% in the non-induced population to 15.98% by 72 hours post induction.

Efficacy of the RNAi was assessed by quantifying the percentage of cells in the population that had a YFP signal, which remained localised to the flagellum. The percentage of cells in the non-induced population with a YFP::TbCCDC19 signal localised to the flagellum was 32.1% (n=1607), which decreased to 16.9% (n=1878) 72 hours post induction of RNAi against TbCCDC19.

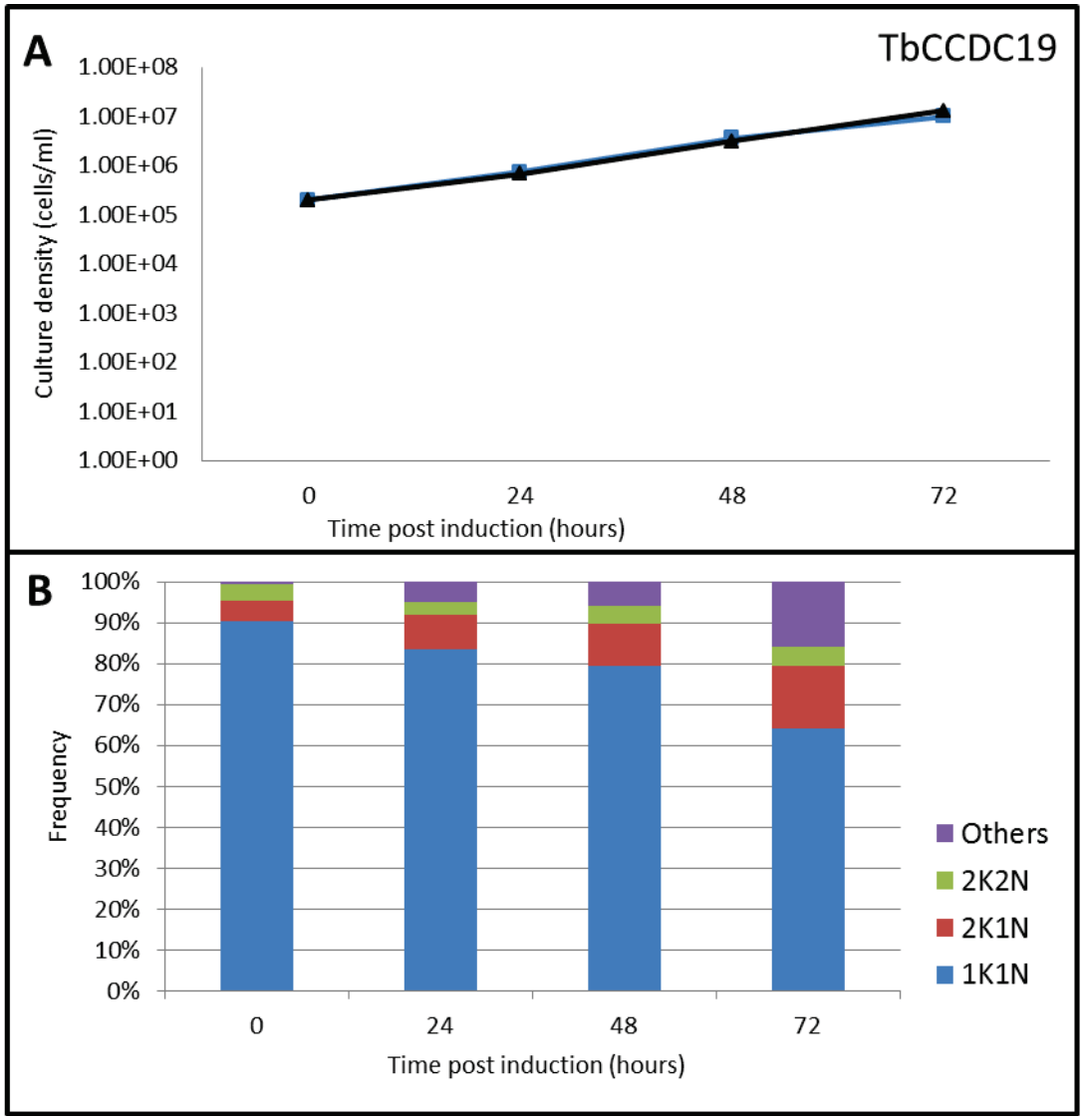


Figure 66: RNAi of TbCCDC19

(A) Continuous growth curve of non-induced (black) and induced (blue) populations of TbCCDC19. (B) Quantification of DNA containing organelles in 0 , 24, 48 and 72 hours post induction (n= 605, 632, 682 and 513, respectively). Cell types were classified as 1K1N (blue), 2K1N (red), 2K2N (green) or abnormal 'others' (purple). Kinetoplast = K, nucleus = N.

5.4.4 Knockdown by RNAi of TbMNS1 affects the cell cycle in *T. brucei.*

To assess the function of TbMNS1, an inducible p2T7-177 RNAi vector (Wickstead et al., 2002) against Tb927.5.1230 was created and transfected into the pPOT_YFP::TbMNS1 procyclic cell line.

A continuous curve was performed (see section 2.4.10 Continuous growth curve) to assess the effect of TbMNS1 ablation of procyclic form cell growth (Figure 67; A). There was no difference in the growth rate of *T. brucei* between the two populations; non-induced (black) and induced (green), by 72 hours post induction.

Quantification of DNA containing organelles (see section 2.4.11 Cell cycle analysis) (Figure 67; B) showed that between 0 and 24 hours of induction the percentage population of 2K1N and 2K2N cells increased but then by 48 and 72 hour post induction the percentage of 2K1N and 2K2N decreased and an increase in the percentage of 'other' cells was more prominent – from 0.98% to 9.09% (0 to 72 hours).

Knockdown of TbMNS1 was assessed by quantifying the percentage of cells in the population that had a YFP signal, which remained localised to the flagellum. The percentage of cells in the non-induced population that had a YFP::TbMNS1 signal along the flagellum was 91.6% (n=498), which decreased to 14.7% (n=300) 72 hours post induction of RNAi against TbMNS1.

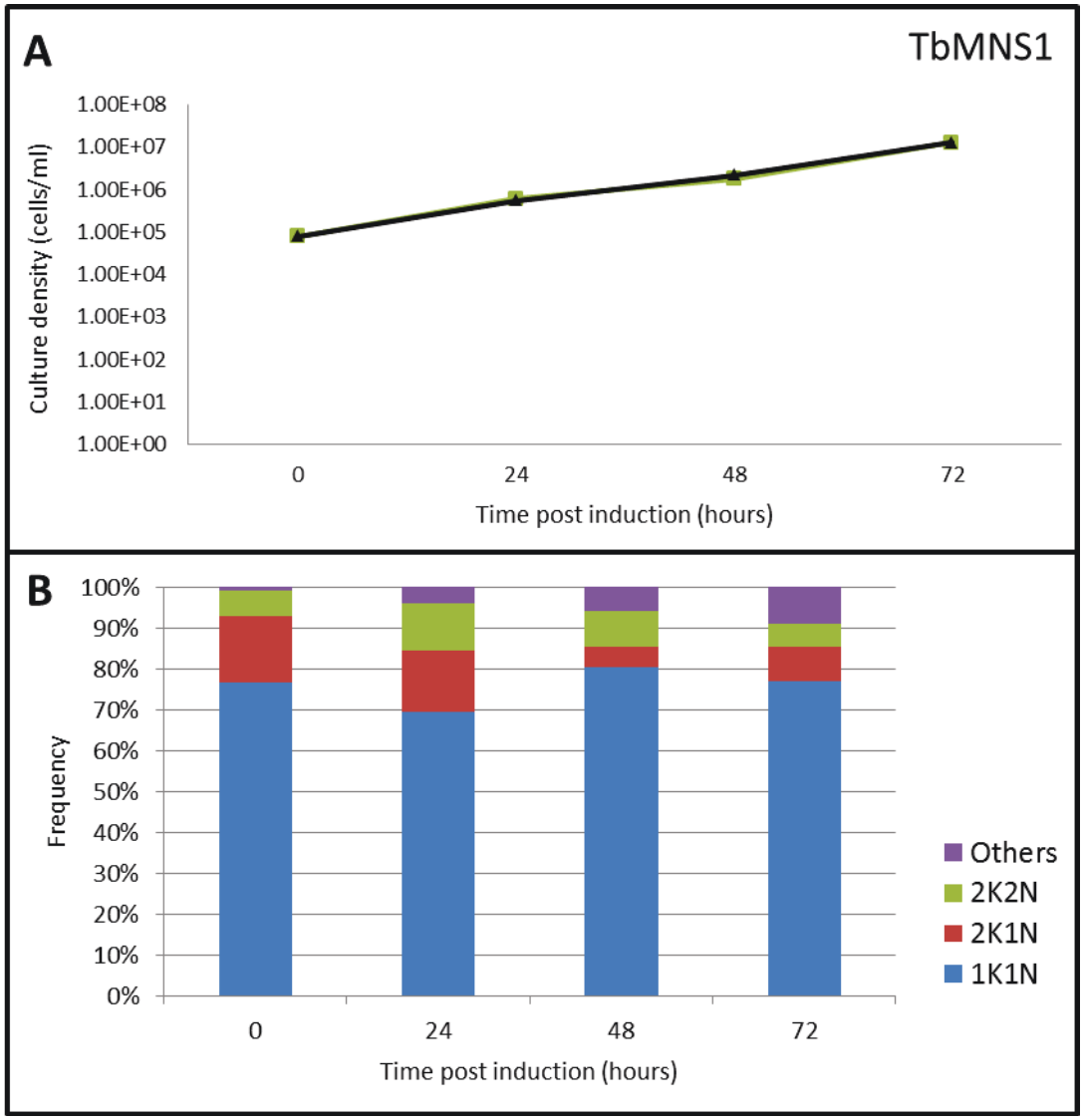


Figure 67: RNAi of TbMNS1

(A) Continuous growth curve of non-induced (black) and induced (green) populations of TbMNS1. (B) Quantification of DNA containing organelles in 0 , 24, 48 and 72 hours post induction (n= 305, 340, 332 and 341, respectively). Cell types were classified as 1K1N (blue), 2K1N (red), 2K2N (green) or abnormal 'others' (purple). Kinetoplast = K, nucleus = N.

5.4.5 TPH domain-containing proteins are not required for flagellum assembly

During the course of RNAi inductions against the three trypanosomal TPH domain-containing proteins, cells were examined by light microscopy. This showed that a flagellum was still assembled after target knockdown (Figure 68) and therefore none of the three conserved TPH domain-containing proteins are required for assembling the flagellum in *T. brucei.* However it was observed in suspension culture that the morphology of the induced cells was abnormal when compared to the non-induced cells. The functionality of the assembled flagellum is investigated in section '5.4.5 TPH domain-containing proteins are not required for flagellum assembly'.

To investigate the difference in cell morphology, cells from non-induced populations (Figure 68; A, C, E) and induced populations (72 hours post induction: Figure 68; B, D and F) were analysed. Cells were fixed and immunolabelled with a well characterised anti β-tubulin antibody, KMX-1 (Birkett *et al.*, 1985). KMX-1 has been raised against an antigen from amoebal tubulin and strongly recognises the β-tubulin isotype found in the subpellicular microtubules of *T. brucei.* The flagellum microtubules of *T. brucei* contains a different β-tubulin isotype to other populations of microtubules within the cell (Gallo and Anderton, 1983). The microtubules in the flagellum were only weakly labelled by KMX-1 in comparison to the subpellicular corset microtubules.

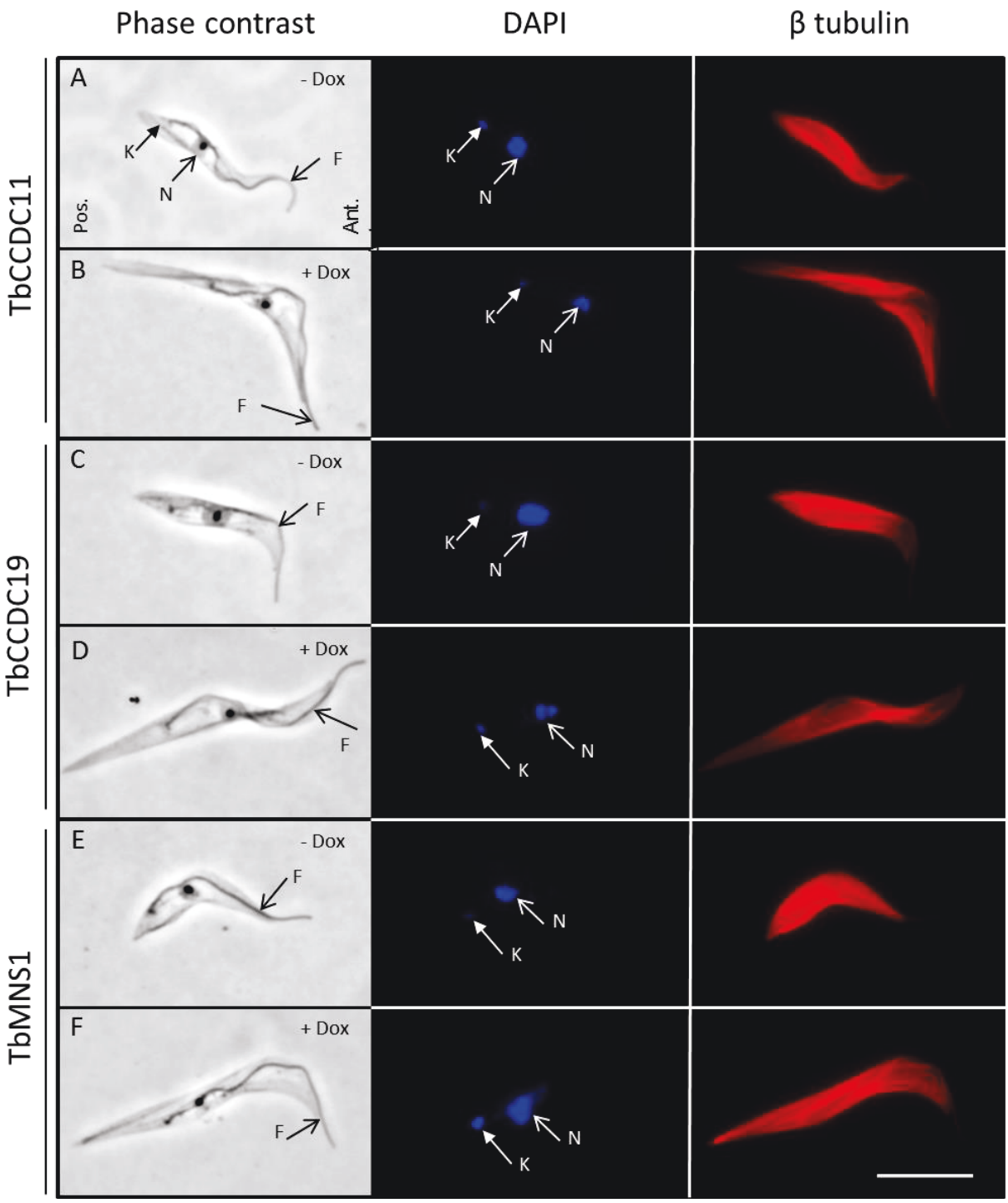


Figure 68: Flagellum assembly not affected by knockdown of TPH domain-containing proteins.

Example 1K1N1F cells from induced and non-induced populations of TbCCDC11 (A, B), TbCDC19 (C, D) and TbMNS1 (E, F). (A, C, E) non-induced cells (- Dox). (B, D, F) Cells from 72 hours post induced populations (+Dox), which still have a flagellum (F). β tubulin was labelled using the KMX-1 antibody (Birkett *et al.*, 1985). All cells are in the same orientation with the posterior end (Pos.), anterior end (Ant.) annotated in panel A. Scale bar = 10µm.

It is common in trypanosome cell biology to measure the paraflagellar rod (PFR) using immunolabelling as a proxy for axoneme/flagellum length (Rotureau *et al.*, 2014). In this work I developed a different approach to directly measure the axoneme length rather than PFR. The ImageJ plugin, NeuronJ (Meijering *et al.*, 2004) was used to semi-automate tracing the flagellum (Figure 24; B, pink line). This was performed on inverted phase contrast images to give a pseudo-fluorescence effect. The length of assembled flagella was measured in non-induced (0 hours) and induced cells (72 hours) for all three cell lines. GraphPad (www.graphpad.com) was used to perform statistical analysis on the data. Only 1K1N1F cells with a single kinetoplast (1K), a single nucleus (1N) and a single flagellum (1F) were used in this analysis to avoid cell cycle related changes to flagellum length and/or cell morphology. For methodology refer to sections '2.6.4 Measurement of flagellum length' and '2.6.5 Measurement of cell body dimensions'.

To examine the effect knockdown of TbCCDC11 had on the length of the flagellum, flagellum length was measured in non-induced 1K1N1F cells (n=90; Figure 69; dark red) and cells 72 hours post induction of RNAi (n=92; Figure 69; light red). The average length of the flagellum in the non-induced population was 20.8µm and the average length of the flagellum in the induced population was 21.5µm. Flagellum length measurements were statistically analysed using GraphPad, an unpaired *t* test was used because the same variable (flagellum length) was being measured in the two independent samples (induced and non-induced). The unpaired *t* test gave a P-value of 0.1287 and therefore found no statistically significant difference between flagellum length of the non-induced (n=90) and the induced cells (n=92) when TbCCDC11 was knocked down.

To examine the effect knockdown of TbCCDC19 had on the length of the flagellum, flagellum length was measured in non-induced 1K1N1F cells (n=61; Figure 69; dark blue) and cells 72 hours post induction of RNAi (n=76; Figure 69; light blue). The average length of a flagellum in the non-induced population was 18.9µm and the average length of a flagellum in the induced population was 19.8µm. Statistical analysis using an unpaired *t* test gave a P value of 0.0262 and therefore the difference in length was statistically significant (*) between the non-induced (n=61) and the induced cells (n=76) for knockdown of TbCCDC19.

Comparison between 1K1N1F cells from non-induced (n=83; Figure 69; dark green) and induced (n=88; Figure 69; light green) populations of TbMNS1 RNAi revealed a significant statistical (**) difference in flagellum length. Statistical analysis using an unpaired t test gave a P value of less than 0.0001. The average length of a flagellum in the non-induced population (n= 83) was 19.2µm and the average length of a flagellum in the induced population (n=88) was 21.1µm.

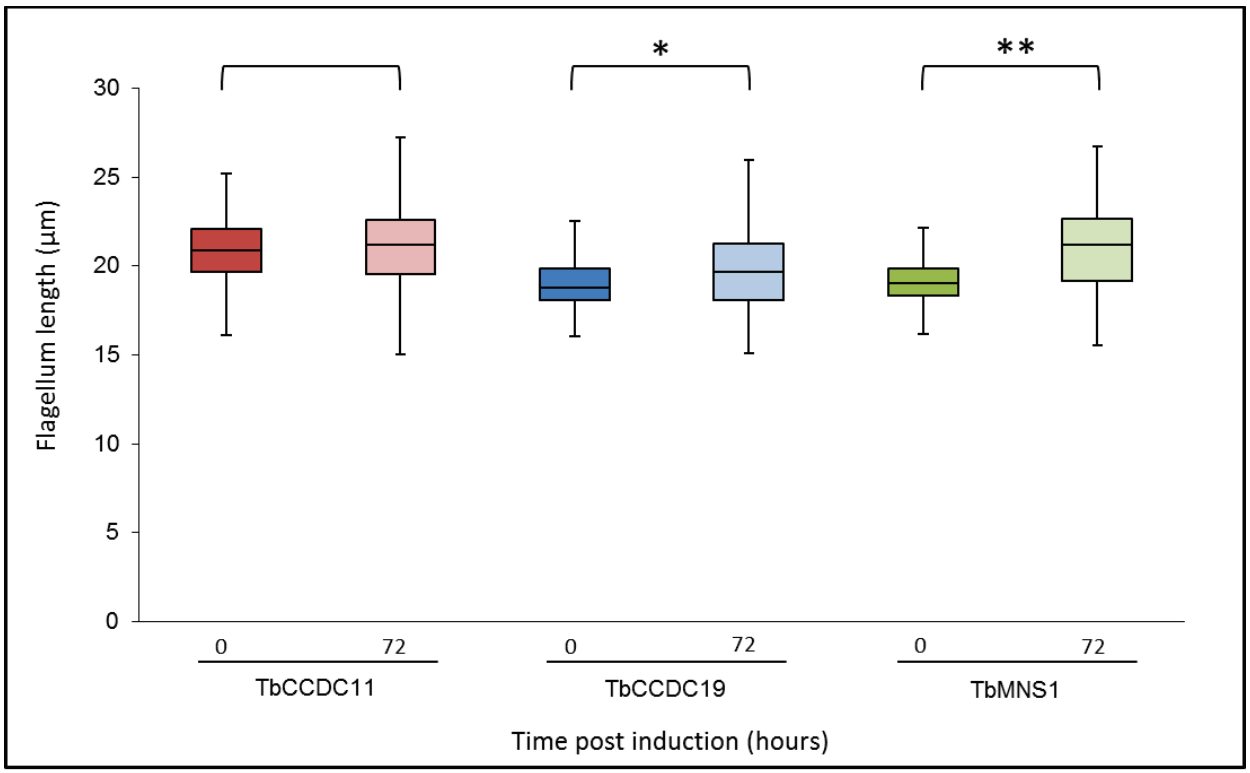


Figure 69: Flagellum length variation in cell populations

Box and whisker plots to demonstrate the variation of flagellum length of 1K1N1F cells. Populations analysed were; TbCCDC11 0 and 72 hours (n=90 and 92. P-value 0.1287), TbCCDC19 0 and 72 hours (n=61 and 76. P-value 0.0262*) and TbMNS1 0 and 72 hours (n=83 and 88. P-value <0.0001**). Statistical significance was calculated using an unpaired t test through GraphPad (* and ** = statistically significant).

5.4.6 Knockdown by RNAi of TbCCDC11 affects the regulation of cell morphology

The cytoskeleton of *T. brucei* has been well characterised (Robinson *et al.*, 1995; Sherwin and Gull, 1989). It is known that the microtubule cytoskeleton undergoes remodelling as part of a normal cell cycle, which is why this analysis was only performed on 1K1N1F cells to avoid cell cycle related changes. It has been established that the morphology of a trypanomastigote cell form is influenced by external factors (e.g. host environment) (Wheeler *et al.*, 2013a). This work analysed procyclic form cells that had been grown axenically in suspension culture.

From the cells analysed for flagellum length (Figure 69), 25 cells from each data set were further analysed. Additional measurements were made for the posterior end length, cell perimeter and 2D cell surface area. For methodology refer to sections '2.6.4 Measurement of flagellum length' and '2.6.5 Measurement of cell body dimensions'. These four morphometric data sets were compared directly and used to assess correlations in changes to cell morphology. The relationship strength between the four variables was assessed by calculating the Pearson correlation coefficient (R).

There was no significant difference in the average length of the flagellum for TbCCDC11 when non-induced cells (n=90) and induced cells (n=92) were statistically analysed (Figure 69). Although the average length of the flagellum increased from 20.8µm to 21.5µm, it was not a statistically significant increase.

Twenty five cells were chosen for further investigation from the non-induced and induced populations because the morphological change between the populations was so apparent by light microscopy (Figure 68). The cells used had to be suitable for analysis in ImageJ, in particular thresholding to produce the 'mask' for perimeter and surface area measurements (see Figure 24; C and D). Procyclic *T. brucei* cells are highly motile in culture and were settled onto the glass microscopy slides whilst live and motile. Cells that settled and attached onto the glass slide in a contorted form or overlapping were not suitable for 2D surface area or perimeter analysis and cells that had settled in a straight conformation were easier to measure the posterior end length.

Direct comparison of the 2D surface area between the non-induced (n=25) and induced (n=25) cells (Figure 70; A) provided the evidence that the average cell surface area had increased by 32.9% from 43.1µm^2 to 57.3µm^2. This was a statistically significant increase (P-value = < 0.0001**) according to an unpaired *t* test.

The average perimeter of cells increased by 31.5% (Figure 70; B) from an average of 41.3µm for non-induced cells (n=25) to an average of 54.3µm in the induced population (n=25). These data were statistically significant when evaluated using an unpaired *t* test (P-value = < 0.0001**).

Observations by light microscopy showed that knockdown of TbCCDC11 caused an elongation at the posterior end of the cell body (Figure 68; A and B). This lengthening was confirmed by measuring the posterior end. The length of the posterior end was defined as the distance between the basal bodies and the posterior end of the cell body (Figure 24; A). These data are presented on a box and whisker plot (Figure 70; C). It was found that there was a statistically significant increase (P-value = < 0.0001**) in the length of the posterior end of the cell body from an average of 3.36µm in the non-induced cells (n=25) to an average length of 7.01µm in the induced cells (n=25). This was an increase of 108.6%. An unpaired *t* test was applied to assess statistical significance of this single variable (posterior end length) between the two populations, non-induced and induced.

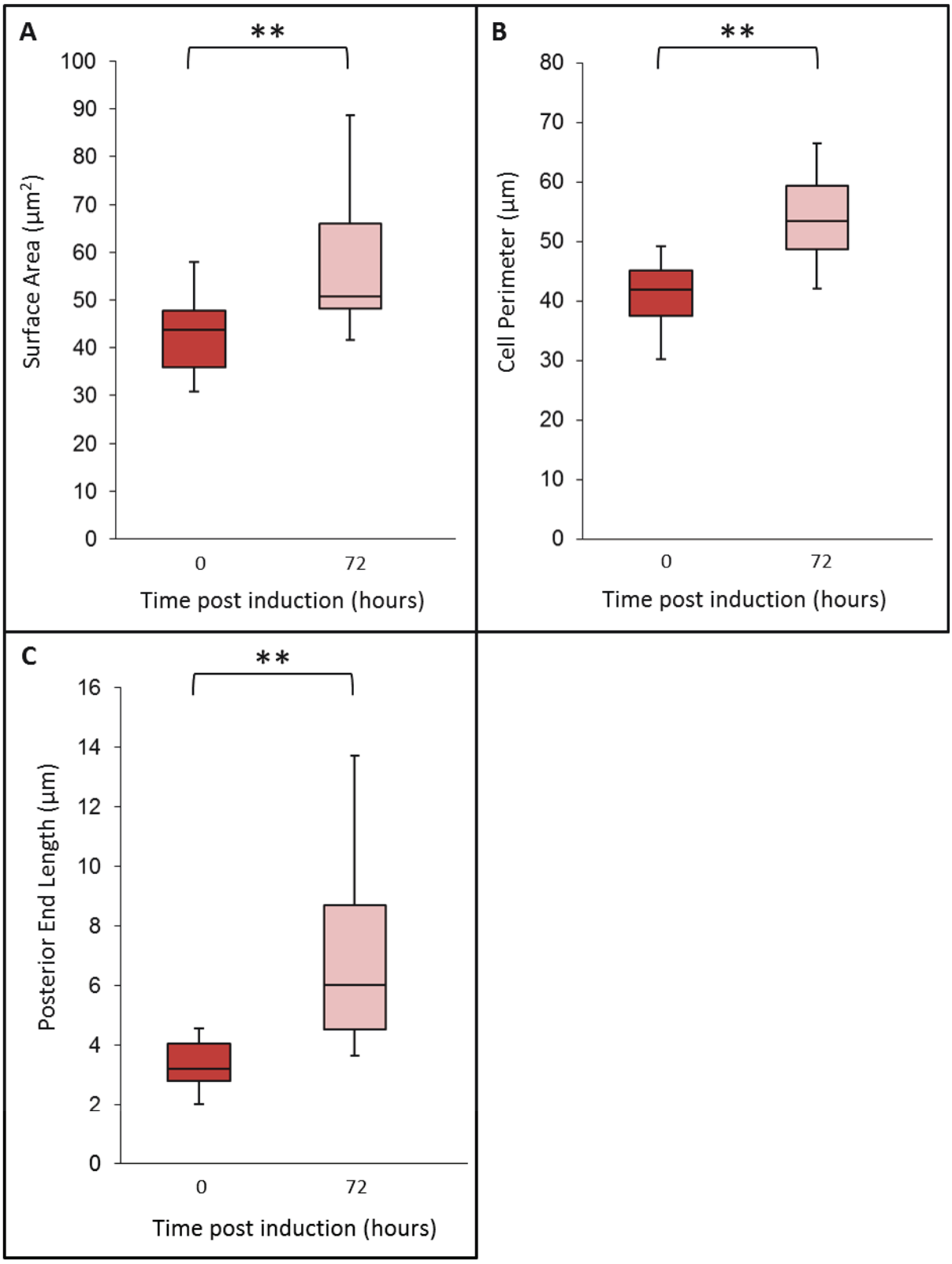


Figure 70: Cell measurements affected by TbCCDC11 knockdown

Box and whisker plots to demonstrate the variation in cell dimensions of 1K1N1F cells for TbCCDC11 knockdown. Populations analysed were TbCCDC11 0 hours (n=25) and 72 hours (n=25). (A) 2D surface area (P-value = <0.0001**). (B) Cell body perimeter (P-value = <0.0001**). (C) Posterior end length (P-value = <0.0001**). Statistical significance was calculated using an unpaired *t* test through GraphPad (* and ** = statistically significant).

When the length of the posterior end was compared to another variable, the length of the flagellum, data points for non-induced cells (n=25) grouped tightly together for TbCCDC11 (Figure 71; A; white). In the non-induced there was a weak positive correlation (R= 0.3538), which was not statistically significant (P-value = 0.08273) between the two variables, the length of the flagellum and the length of the posterior end. After 72 hours induction of the RNAi the induced cell data points (n=25) were widely dispersed on the plot (Figure 71; A; grey) and had a negative correlation (R= -0.237, P-value = 0.251947). This meant that knockdown of TbCCDC11 disrupted the regulation between the length of the flagellum and the length of the cell body at the posterior end.

Due to the increase in the surface area of the cells, post knockdown of TbCCDC11, it was expected that there would also be an increase in the cell perimeter. There was a statistically significant correlation between the 2D surface area and the perimeter of the cell body (R= 0.823, P-value = < 0.00001**) in the non-induced cell population (Figure 71; B; white, n=25). The strength of this relationship dropped in the induced population (Figure 71; B; grey, n=25) although it was still a positive correlation (R= 0.7034, P-value = 0.000088**).

Plotting flagellum length against cell perimeter (Figure 71; C) revealed that the non-induced population (n=25) had a strong positive correlation between these two measurements (Figure 71; C; white. R=0.793, P-value = < 0.00001**) but after 72 hours of RNAi targeted against TbCCDC11, the correlation was weakened. In the induced population (n=25), the data points (Figure 71; C; grey) where more dispersed on the plot (R=0.328, P-value = 0.109676) and the strong correlation between flagellum length and cell perimeter was disrupted by TbCCDC11 knockdown.

Correlation between flagellum length and 2D surface area of the cell (Figure 71; D) was positive in the non-induced cell population (Figure 71; D; white, n=25). This positive relationship between the two variables remained strong in the induced population (Figure 71; D; grey, n=25). The correlation coefficient of the non-induced cells (R=0.6601, P-value = 0.00033*) was similar to that of the induced cells (R= 0.6578, P-value = 0.000352*). This suggests that despite the RNAi, the cell has to keep the ratio between

flagellum length and surface area relatively constant. This is presumably to try and ensure effective propulsive motility, regardless of cell speed.

Figure 71 (E) shows the relationship between the length of the cell posterior end and the cell perimeter. The data points from the non-induced cells (n=25; white) clustered together whilst the data points from the induced population (n=25; grey) were more dispersed on the x axis. The non-induced population showed a weak positive correlation between the variable (R=0.491, P-value = 0.012755*). Seventy two hours post TbCCDC11 knockdown, it was apparent that there was a large increase in the posterior end length but the positive correlation with cell perimeter was not disrupted in the induced population(R = 0.692, P-value = 0.000128**).

The relationship between the posterior end length and the cell 2D surface area (Figure 71; F) was a strong positive correlation (R= 0.7095) in the non-induced cells (n=25; white), which was statistically significant (P-value = 0.000071**). This strong correlation was lost after knockdown of TbCCDC11 because for the data for the induced cell population (n=25; grey) had a positive correlation coefficient close to 0 (R = 0.1226, P-value = 0.559337).

The statistical analysis of TbCCDC11 morphometric analysis is summarised in Table 19, which shows the strength of the relationship between cell dimensions (R) and the P-value for that relationship.

Table 19: Summary of TbCCDC11 morphometric statistical analysis.
Collation of the Pearson correlation coefficients (R) and the P-values from non-induced and induced cell populations of the TbCCDC11 RNAi cell line analysed in Figure 71.

TbCCDC11		**Flagellum length**		**Posterior end length**		**Perimeter**	
Posterior end length	R =	0.3538	-0.2368				
	P-value =	0.08273	0.251947				
Perimeter	R =	0.7926	0.3278	0.4907	0.6918		
	P-value =	< 0.00001 (**)	0.109676	0.012755 (*)	0.000128 (*)		
Surface area	R =	0.6601	0.6578	0.7095	0.1226	0.8226	0.7034
	P-value =	0.00033 (*)	0.000352 (*)	0.000071 (**)	0.559337	< 0.00001 (**)	0.000088 (**)
		non-induced	induced	non-induced	induced	non-induced	induced

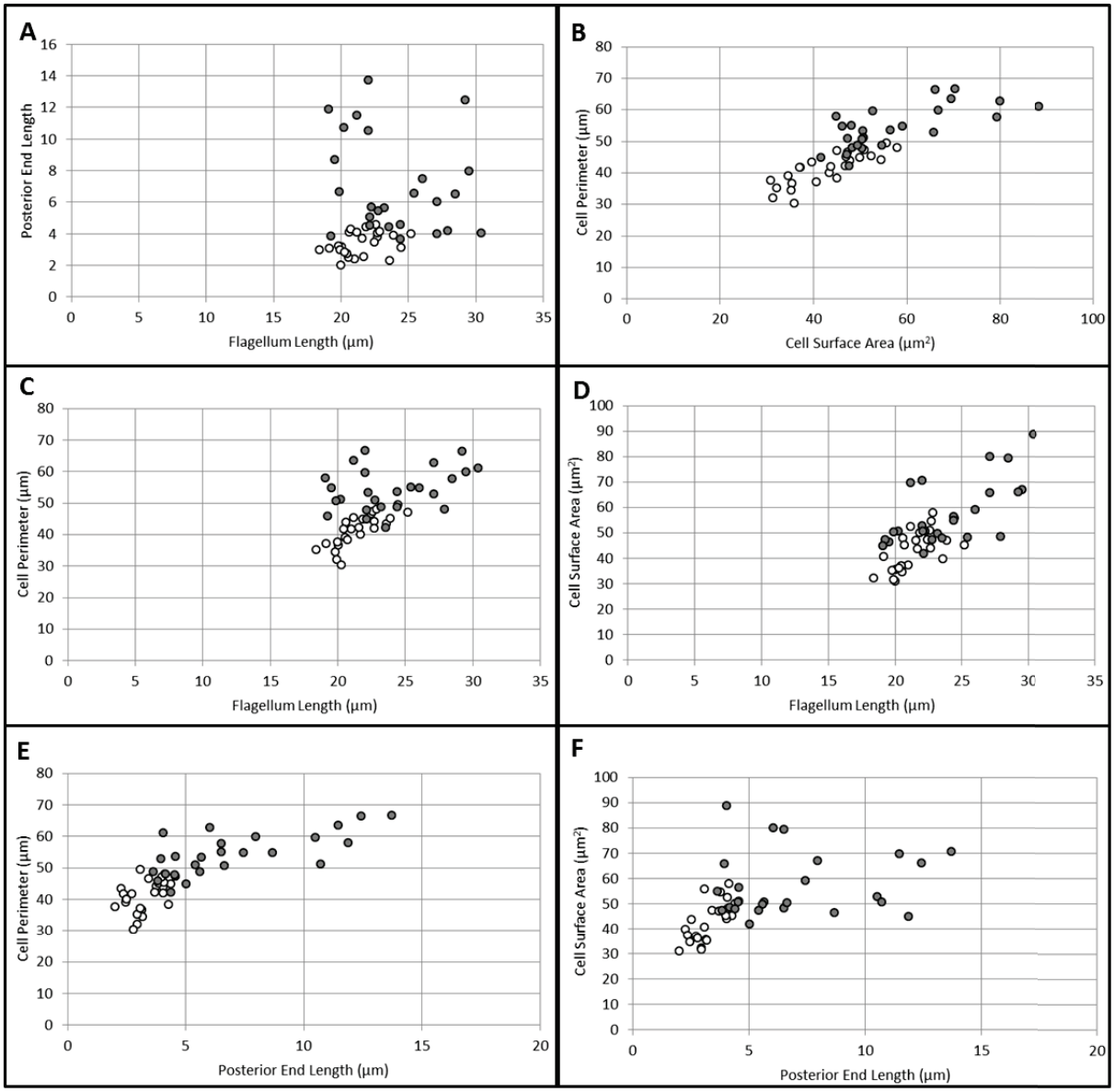


Figure 71: Cell morphometrics for TbCCDC11

Scatter graphs for morphometric analysis. 25 cells from each population of 0 hours (white data points) and 72 hours post induction of RNAi (grey data points) were used for analysis. Axes are annotated separately for each plot (A-F)

5.4.7 Knockdown by RNAi of TbCCDC19 affects the regulation of cell morphology

Once it had been confirmed that knockdown of TbCCDC19 did affect the length of the flagellum (Figure 69), other morphometric values were analysed in conjunction with flagellum length for 25 cells. It has been published that cells that build a shorter flagellum, form a smaller cell body (Kohl *et al.*, 2003). I wanted to investigate if the opposite was true; would cells that assembled a longer flagellum, produce a larger cell body?

Post-knockdown of TbCCD19, it was found that the surface area increased by 30.6%, from an average of 46.92µm^2 in non-induced cells (n=25; light blue) to 61.28µm^2 in induced cells (n=25; dark blue). The data were plotted in box and whisker format (Figure 72; A) and when evaluated using an unpaired *t* test there was a statistically significant difference (P-value = < 0.0001**) between the two populations.

As the surface area of cells increased after 72 hours of TbCCDC19 knockdown, so did the cell body perimeter. The average perimeter of a cell increased by 41.2% (Figure 11; B) from 37.12µm for non-induced cells (n=25; dark blue) to an average of 52.40µm in the induced population (n=25; light blue). This was a statistically significant increase in the cell perimeter (P-value = < 0.0001**).

The length of the posterior end of the cell body increased from an average of 3.17µm in the non-induced cells (n=25; dark blue) to an average length of 6.92µm in the induced cells (n=25; light blue). That is a 118.3% increase in the posterior end of a cell. This increase in length can be seen from the box and whisker plot (Figure 72; C). The statistical significance of this elongation was assessed using an unpaired *t* test (P-value = < 0.0001**).

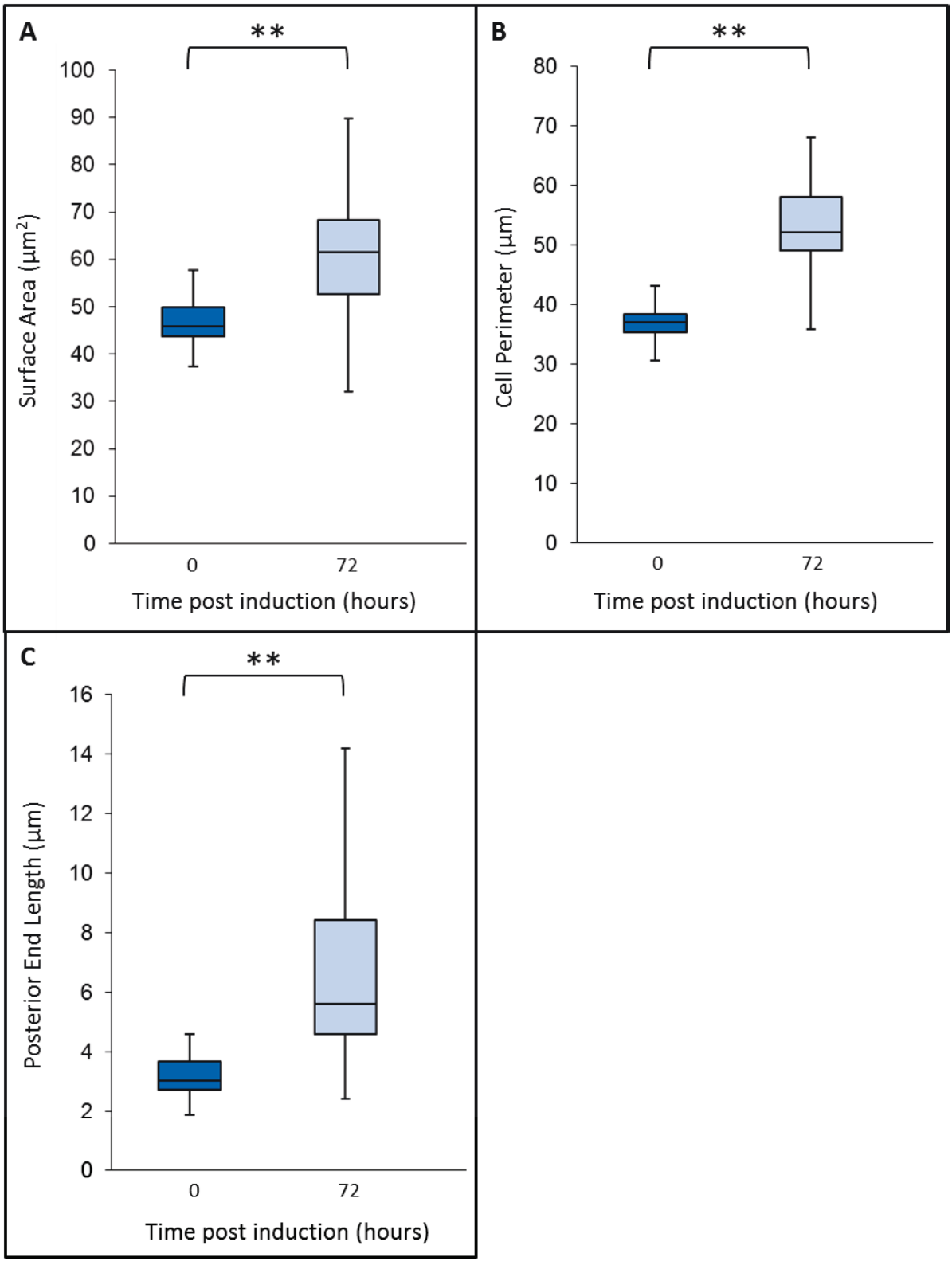


Figure 72: Cell measurements affected by TbCCDC19 knockdown

Box and whisker plots to demonstrate the variation in cell dimensions of 1K1N1F cells for TbCCDC19 knockdown. Populations analysed were TbCCDC19 0 hours (n=25) and 72 hours (n=25). (A) 2D surface area (P-value = <0.0001**). (B) Cell body perimeter (P-value = <0.0001**). (C) Posterior end length (P-value = <0.0001**). Statistical significance was calculated using an unpaired *t* test through GraphPad (* and ** = statistically significant).

Comparison between the length of the posterior end and the length of the flagellum showed data points for non-induced cells (n=25) grouped tightly together for TbCCDC19 (Figure 73; A; white). In the non-induced population there was a statistically significant, strong positive correlation (R= 0.5162, P-value = 0.008252*) between the length of the flagellum and the length of the posterior end. After 72 hours induction of the RNAi the induced cell data points (n=25) were widely dispersed on the plot (Figure 73; A; grey) and had a statistically significant negative correlation (R= -0.0824, P-value = < 0.00001**). This meant that knockdown of TbCCDC19 disrupted the regulation between the length of the flagellum and the length of the cell body at the posterior end.

Due to the increase in the surface area of the cells, post knockdown of TbCCDC19, it was expected that there would also be an increase in the cell perimeter. There was not a statistically significant correlation between the 2D surface area and the perimeter of the cell body (R= 0.4452, P-value = 25738) in the non-induced cell population (Figure 73; B; white, n=25). However there was a positive correlation between the two variables in the induced population (Figure 71; B; grey, n=25) (R= 0.7063, P-value = 0.00008**). This showed that a change of surface area affected the cell perimeter when TbCCDC19 was knocked down.

Plotting flagellum length against cell perimeter (Figure 73; C) revealed that the non-induced population (n=25) statistically had no correlation between these two measurements (Figure 73; C; white. R=1.467, P-value = 0.484085). After knockdown of TbCCDC19, the mathematical correlation was stronger (R=0.3558, P-value = 0.080887) but not statistically significant, for the induced population data points (n=25; Figure 73; C; grey).

There was no statistically significant correlation between flagellum length and 2D surface area of the cell (Figure 73; D) in the non-induced cell population (Figure 73; D; white, n=25). Although statistically there was a correlation between the two variables in the induced population (Figure 73; D; grey, n=25). The correlation coefficient of the non-induced cells (R=0.1222, P-value = 0.560629) was weaker to that of the induced cells (R= 0.5504, P-value = 0.00436*) but the data points on the graph show a clear shift towards the upper right for induced data points (Figure 73; D; grey, n=25).

The relationship between the length of the posterior end and the cell perimeter (Figure 73; E) showed that in the non-induced population (n=25; white) the data points clustered together whilst the data points from the induced population (n=25; grey) were more dispersed along the x axis. The non-induced population showed a weak positive correlation between the variables (R=0.1233, P-value = 0.557079), which was not statistically significant. Seventy two hours post TbCCDC19 knockdown, it was apparent that there was a large increase in the posterior end length and this was accompanied by a statistically significant positive correlation with an increase in cell perimeter (R = 0.6486, P-value = 0.000453*).

Both the non-induced cells (n=25; white) and the induced cells (n=25; grey) had a weak positive relationship between posterior end length and 2D surface area (Figure 73; F). Neither correlation coefficients were statistically significant, in the non-induced cells R= 0.225 (P-value = 0.279544) and this coefficient was similar in the induced cells, R = 0.2331 (P-value = 0.262133).

The statistical analysis of TbCCDC19 morphometric analysis is summarised in Table 20, which shows the strength of the relationship between cell dimensions (R) and the P-value for that correlation.

Table 20: Summary of TbCCDC19 morphometric statistical analysis.

Collation of the Pearson correlation coefficients (R) and the P-values from non-induced and induced cell populations of the TbCCDC19 RNAi cell line analysed in Figure 73.

TbCCDC19		**Flagellum length**		**Posterior end length**		**Perimeter**	
Posterior end length	R =	0.5162	-0.0824				
	P-value =	0.008252 (*)	< 0.00001 (**)				
Perimeter	R =	0.1467	0.3558	0.1233	0.6486		
	P-value =	0.484085	0.080887	0.557079	0.000453 (*)		
Surface area	R =	0.1222	0.5504	0.225	0.2331	0.4452	0.7063
	P-value =	0.560629	0.00436 (*)	0.279544	0.262133	0.25738	0.00008 (**)
		non-induced	induced	non-induced	induced	non-induced	induced

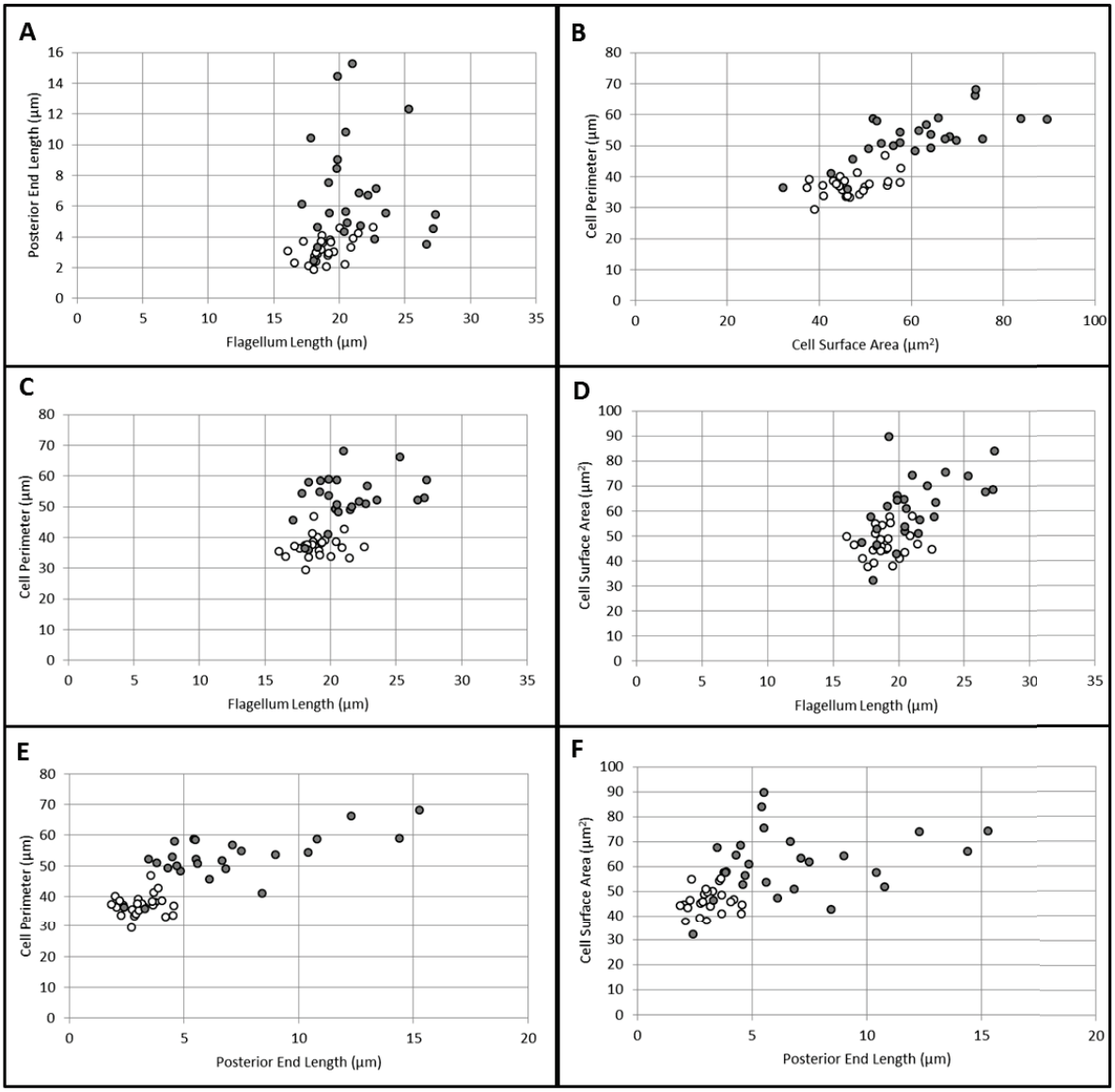


Figure 73: Cell morphometrics for TbCCDC19

(A-F) Scatter graphs for morphometric analysis. 25 cells from each population of 0 hours (white data points) and 72 hours post induction of RNAi (grey data points) were used for analysis. Axes are annotated separately for each plot (A-F)

5.4.8 Knockdown by RNAi of TbMNS1 affects the regulation of cell morphology

Measurement of the flagellum in induced (n=88) and non-induced (n=83) cells, revealed a statistically significant increase in flagellum length (P-value <0.0001**) (Figure 69). To continue investigating the effect knockdown of TbMNS1 had on *T. brucei,* other morphometric values were considered. Measurements were also made of the posterior end length, cell perimeter and 2D cell surface area (see section 2.6.5 Measurement of cell body dimensions, Figure 24) for non-induced (n=25) and induced cells (n=25).

The surface area (Figure 74; A) increased by 19.1% from an average of 44.56µm^2 in non-induced cells (n=25; light green) to an average of 53.07µm^2 in induced cells (n=25; dark blue). Figure 74 presents the measurements as a box and whisker plot and when evaluated using an unpaired *t* test there was a statistically significant difference (P-value = 0.0129*) between the two populations. As the surface area of cells increased after 72 hours of TbMNS1 knockdown, so did the perimeter of the cell body. There was a 37.1% increase in the perimeter of the cell body. This increased from an average of 38.42µm for non-induced cells (n=25; Figure 74; B, dark green) to an average of 52.66µm in the induced population (n=25; Figure 74; B, light green). This was a statistically significant increase in the average cell perimeter (P-value = < 0.0001**). The length of the posterior end of the cell body increased (Figure 74; C) from an average of 3.08µm in the non-induced cells (n=25; dark green) to an average length of 6.65µm in the induced cells (n=25; light green). This 115.9% elongation was assessed using an unpaired *t* test (P-value = < 0.0001**), which deemed the length increase to be statistically significant.

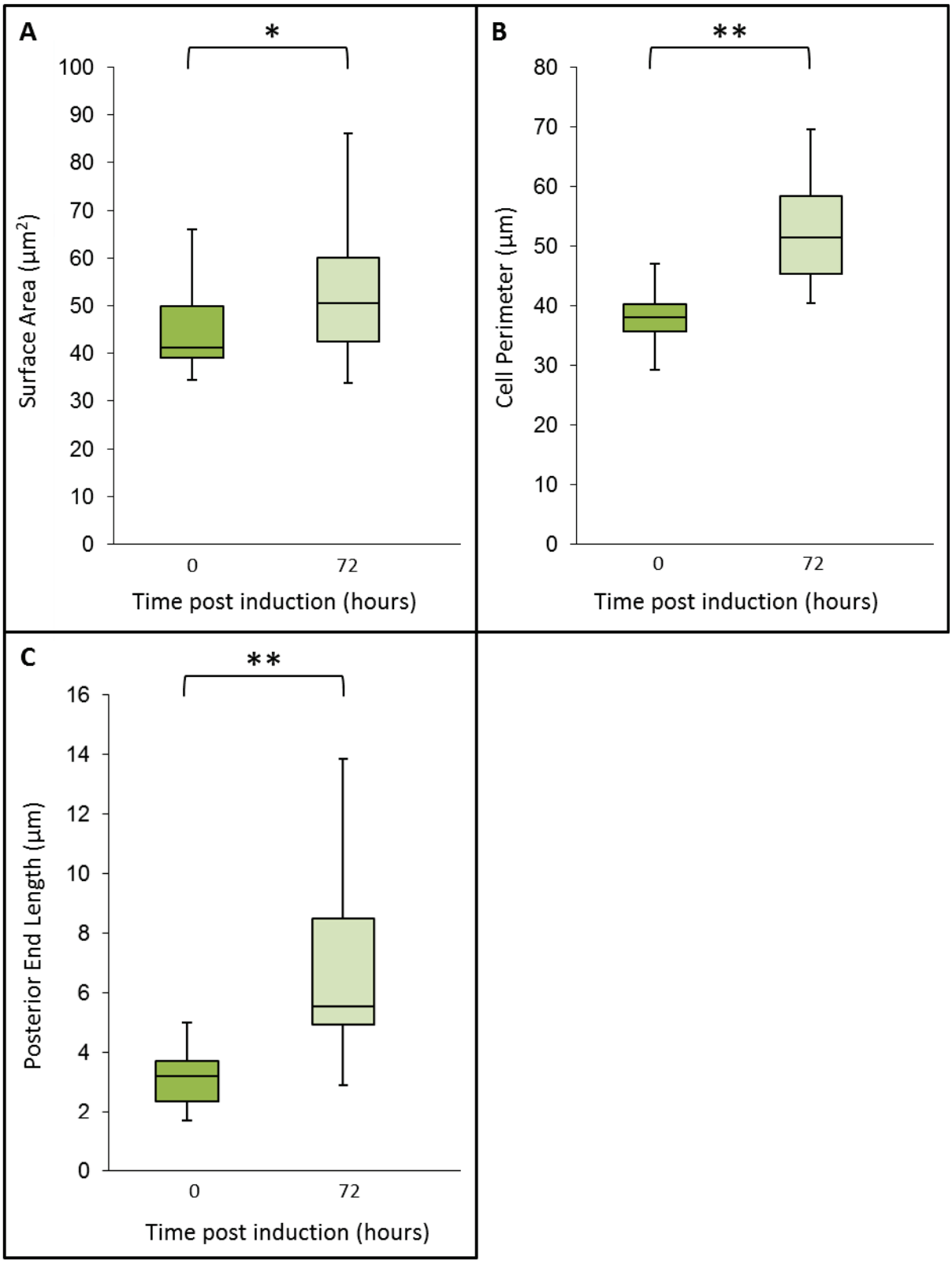


Figure 74: Cell measurements affected by TbMNS1 knockdown

Box and whisker plots to demonstrate the variation in cell dimensions of 1K1N1F cells for TbMNS1 knockdown. Populations analysed were TbMNS1 0 hours (n=25) and 72 hours (n=25). (A) 2D surface area (P-value = 0.0129*). (B) Cell body perimeter (P-value = <0.0001**). (C) Posterior end length (P-value = <0.0001**). Statistical significance was calculated using an unpaired *t* test through GraphPad (* and ** = statistically significant).

Once it had been established that these cellular dimensions increased in direct comparison of the non-induced to the induced cell population, I wanted to know if there was any correlation between the measurements. Variables were plotted against each other in all combinations to investigate if statistically significant relationships were evident.

When the length of the posterior end was compared to the length of the flagellum, data points for non-induced cells (n=25) grouped tightly together for TbMNS1 (Figure 75; A; white). In the non-induced there was a weak positive correlation (R= 0.2472), which was not statistically significant (P-value = 0.233528) between the two variables. After 72 hours the induced cell data points (n=25) were widely dispersed on the plot (Figure 75; A; grey) and had a weaker correlation (R=0.0626), which was not significant (P-value = 0.766263).

Due to the increase in the surface area of the cells, post knockdown of TbMNS1, it was expected that there would also be an increase in the cell perimeter. There was a statistically significant correlation between the 2D surface area and the perimeter of the cell body (R= 0.8292, P-value = < 0.00001**) in the non-induced cell population (Figure 75; B; white, n=25). The strength of this relationship dropped in the induced population (Figure 75; B; grey, n=25) although it was still a statistically significant positive correlation (R= 0.723, P-value = 0.000044**).

Analysing the relationship of flagellum length against cell perimeter (Figure 75; C) revealed that the non-induced population (n=25) had a mild positive correlation (Figure 75; C; white. R=0.4476, P-value = < 0.024858). In the induced population (n=25), the data points (Figure 75; C; grey) were more dispersed on the plot and therefore had a weaker correlation that was statistically not significant (R=0.3066, P-value = 0.136043).

Correlation between flagellum length and 2D surface area of the cell (Figure 75; D) was positive in the non-induced cell population (Figure 75; D; white, n=25). This positive relationship between the two variables decreased in the induced population (Figure 75; D; grey, n=25). The correlation coefficient of the non-induced cells (R=0.4578, P-value = 0.021382) was slightly higher to that of the induced cells (R= 0.2283, P-value = 0.272364) but neither were statistically significant.

Figure 75 (E) shows the relationship between the length of the cell posterior end and the cell perimeter. The data points from the non-induced cells (n=25; white) clustered together whilst the data points from the induced population (n=25; grey) were more dispersed on the x axis. The non-induced population showed a statistically significant positive correlation between the variable (R=0.5465, P-value = 0.004705*). Post TbMNS1 knockdown, it was apparent that there was a large increase in the posterior end length and a statistically significant, positive correlation with cell perimeter (R = 0.8301, P-value = < 0.00001**).

The relationship between the posterior end length and the cell 2D surface area (Figure 75; F) was positive in the non-induced cells (n=25; white) with R= 0.4786 (P-value = 0.15511). This positive correlation was stronger after knockdown of TbMNS1, the induced cells (n=25; grey) R = 0.5423 (P-value = 0.0.005102*), which was a significant relationship.

The statistical analysis of TbMNS1 morphometric analysis is summarised in Table 21, which shows the strength of the relationship between cell dimensions (R) and the P-value for that correlation.

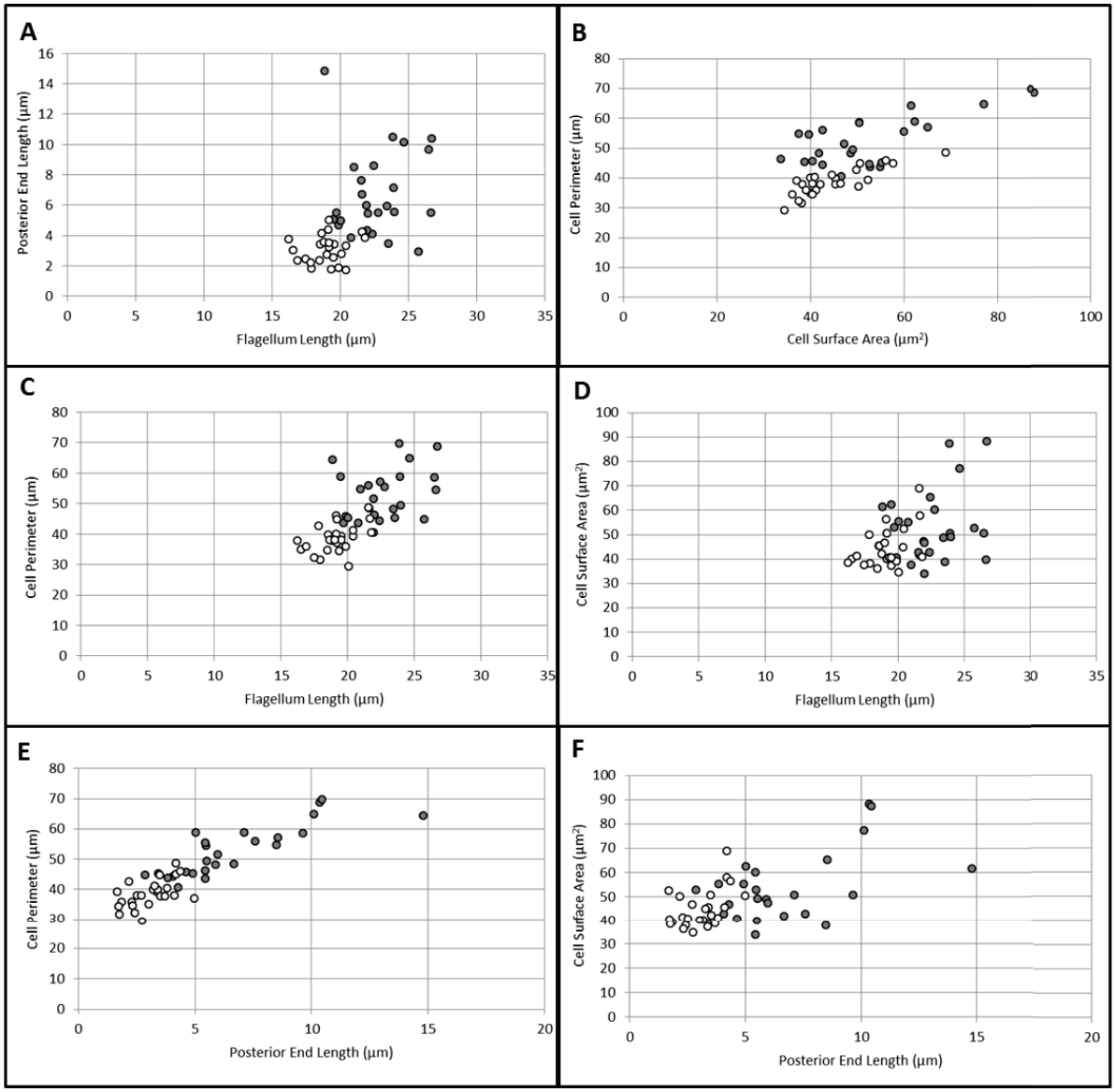


Figure 75: Cell morphometrics for TbMNS1

(A-F) Scatter graphs for morphometric analysis. 25 cells from each population of 0 hours (white data points) and 72 hours post induction of RNAi (grey data points) were used for analysis. Axes are annotated separately for each plot (A-F)

Table 21: Summary of TbMNS1 morphometric statistical analysis.

Collation of the Pearson correlation coefficients (R) and the P-values from non-induced and induced cell populations of the TbMNS1 RNAi cell line analysed in Figure 75.

TbMNS1		Flagellum length		Posterior end length		Perimeter	
Posterior end length	R =	0.2472	0.0626				
	P-value =	0.233528	0.766263				
Perimeter	R =	0.4476	0.3066	0.5465	0.8301		
	P-value =	0.024858	0.136043	0.004705 (*)	< 0.00001 (**)		
Surface area	R =	0.4578	0.2283	0.4786	0.5423	0.8292	0.723
	P-value =	0.021382	0.272364	0.15511	0.005102 (*)	< 0.00001 (**)	0.000044 (**)
		non-induced	induced	non-induced	induced	non-induced	induced

5.4.9 TPH domain-containing proteins are required for normal cell motility.

Knockdown of trypanosomal TPH domain-containing proteins affected flagellum length (Figure 69) and caused changes to cell body dimensions (Figure 70, Figure 74). It is common in trypanosome cell biology that perturbing a component of the axoneme will affect cell motility (Baron *et al.*, 2007a; Dawe *et al.*, 2007). Motility is an essential function of a cilium or flagellum. In mammals, if motility is impaired or paralysed this can lead to serious clinical phenotypes collectively known as ciliopathies (for review see (Baker and Beales, 2009))

To investigate the effect that separate knockdown of the three trypanosomal TPH domain-containing proteins would have on cell motility, parasites were analysed using live cell microscopy. Images were collected to screen for differences in motility between the non-induced and induced RNAi populations. *T. brucei* procyclic cells were seeded at 1 x 10^6 cells ml^{-1} and sequential images were captured for 15 x 1 minute sessions. Cell suspensions were deposited into disposable haemocytometers for image capture and each chamber was imaged twice at a different X and Y position on a non-heated stage. The image sets were processed in ImageJ using motility analysis scripts (Wheeler, 2012) and the average speed calculated for 100 cells from both the non-induced and induced populations of the three inducible RNAi cell lines.

To examine the effect knockdown of TbCCDC11 had on cell motility, 15 sets of 120 sequential static frames were analysed using motility analysis scripts (Wheeler, 2012) in ImageJ (NIH). After 100 tracks were detected, a stop was applied and excess data discarded. To increase accuracy, the motility analysis scripts discounted cells that were detectable for fewer than 5 seconds. For this functional analysis the aspect of interest was in the effect on propulsive motility therefore the variable analysed was speed of cell movement. The value spread for cell speed is represented as a histogram for non-induced (Figure 76; A) cells and induced cells (Figure 76; B).

The average speed of non-induced cells (n=100) was 3.65µm/second, which was reduced by 34.8% to 2.38µm/second after 48 hours of RNAi induction (n=100). Statistical analysis using an unpaired *t* test classed the difference in cell speed between the non-induced and the induced cell populations as statistically significant (P-value of <0.0001**). This meant

that ablation of TbCCDC11 had an effect on cell motility, specifically reducing cell speed. Tracking the cell movements demonstrated that non-induced cells (Figure 76; C) appeared to have straighter tracks than the population of cells 48 hours post induction of RNAi (Figure 76; D).

TbCCDC19 knockdown exhibited the most severe motility defect by 48hrs post RNAi induction with the average cell speed reducing by 58.4%. The non-induced population of cells (n=100, Figure 77; A) had an average speed of 3.41µm/second. Two days post induction of RNAi against TbCCDC19, the average cell speed had decreased to 1.42µm/second (n=100, Figure 77; B). Statistical analysis using an unpaired t test gave a P value of less than 0.0001. Trace analysis verified that propulsive motility was greatly reduced after TbCCDC19 knockdown and many cells exhibited shorter tracks (Figure 77; D) in comparison to non-induced cell tracks (Figure 77; C).

TbMNS1 also exhibited a decreased motility, which was milder than the phenotype quantified for TbCCDC11 or TbCCDC19 knockdown. At 48 hours post induction of RNAi the cells analysed had an average speed of 2.24µm/second (n=100, Figure 78; A) compared to 3.05µm/second in the non-induced population (n=100, Figure 78; B). Statistical analysis using an unpaired t test gave a P value of less than 0.0001**. When TbMNS1 was knocked down, average cell speed was reduced by 26.6%. Cell motility tracks for non-induced cells (Figure 78; C) appear longer and straighter than those for induced cells (Figure 78; D).

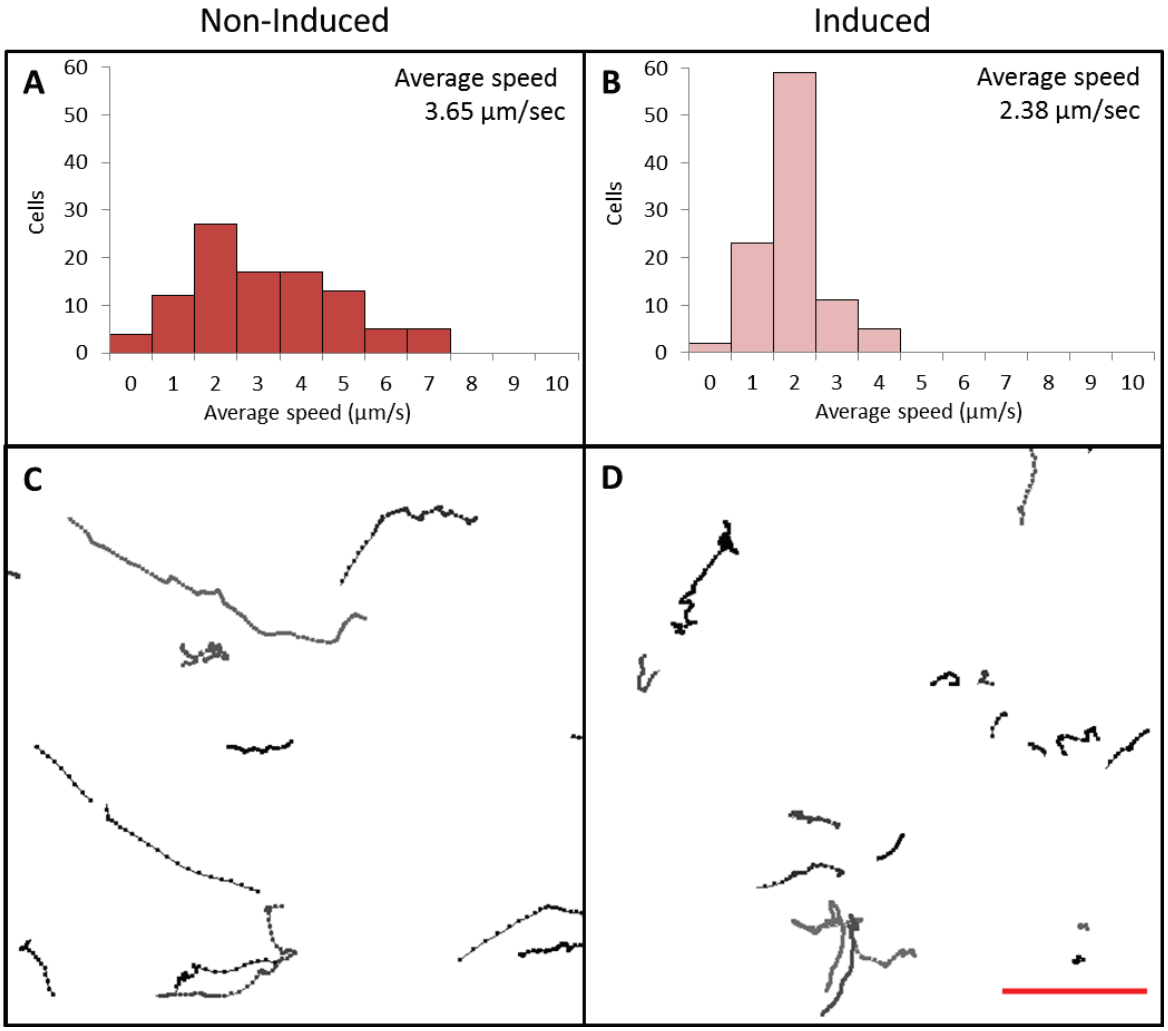


Figure 76: Knockdown of TbCCDC11 affects motility

Cell speed was calculated for 100 cells. Data were plotted on a histogram for non-induced cells (A, n=100) and induced cells (B, n=100). The average cell speed in the non-induced cell population was 3.65µm/second. This decreased to an average speed of 2.38µm/second in the inducted population. This difference in speed was statistically significant (P-value = < 0.0001**). The data from induced cells were recorded after 48 hours of RNAi induction targeted against TbCCDC11. Example movement tracks of non-induced (C) and induced (D) cells are shown from an area of 200µm^2 for each population (grey scale), number of tracks is arbitrary. Scale bar = 50µm.

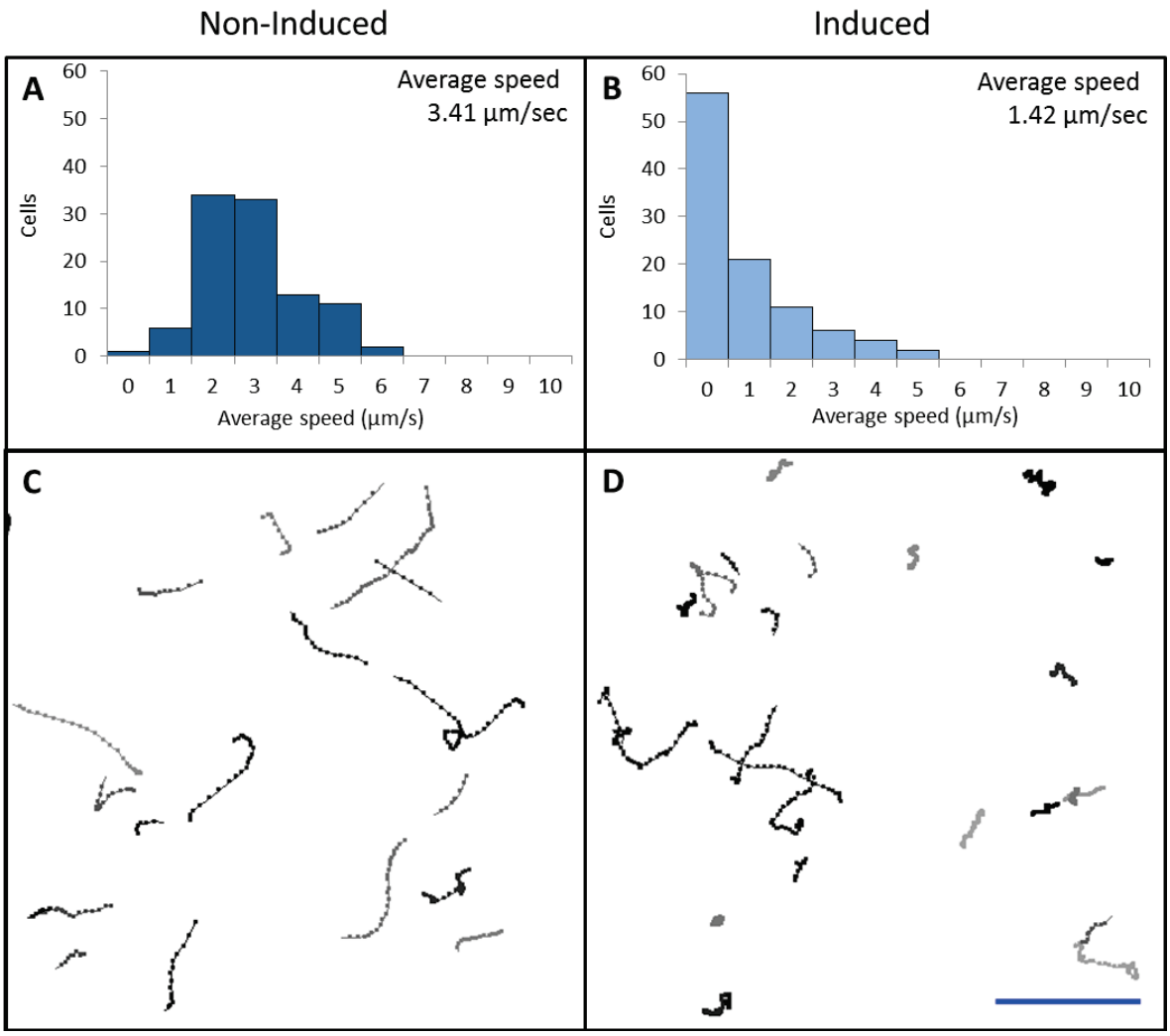


Figure 77: Knockdown of TbCCDC19 severely affects motility

Cell speed was calculated for 100 cells. Data were plotted on a histogram for non-induced cells (A, n=100) and induced cells (B, n=100). The average cell speed in the non-induced cell population was 3.41µm/second. This decreased to an average speed of 1.42µm/second in the inducted population. This difference in speed was statistically significant (P-value = < 0.0001**). The data from induced cells were recorded after 48 hours of RNAi induction targeted against TbCCDC19. Example movement tracks of non-induced (C) and induced (D) cells are shown from an area of 200µm^2 for each population (grey scale), number of tracks is arbitrary. Scale bar = 50µm.

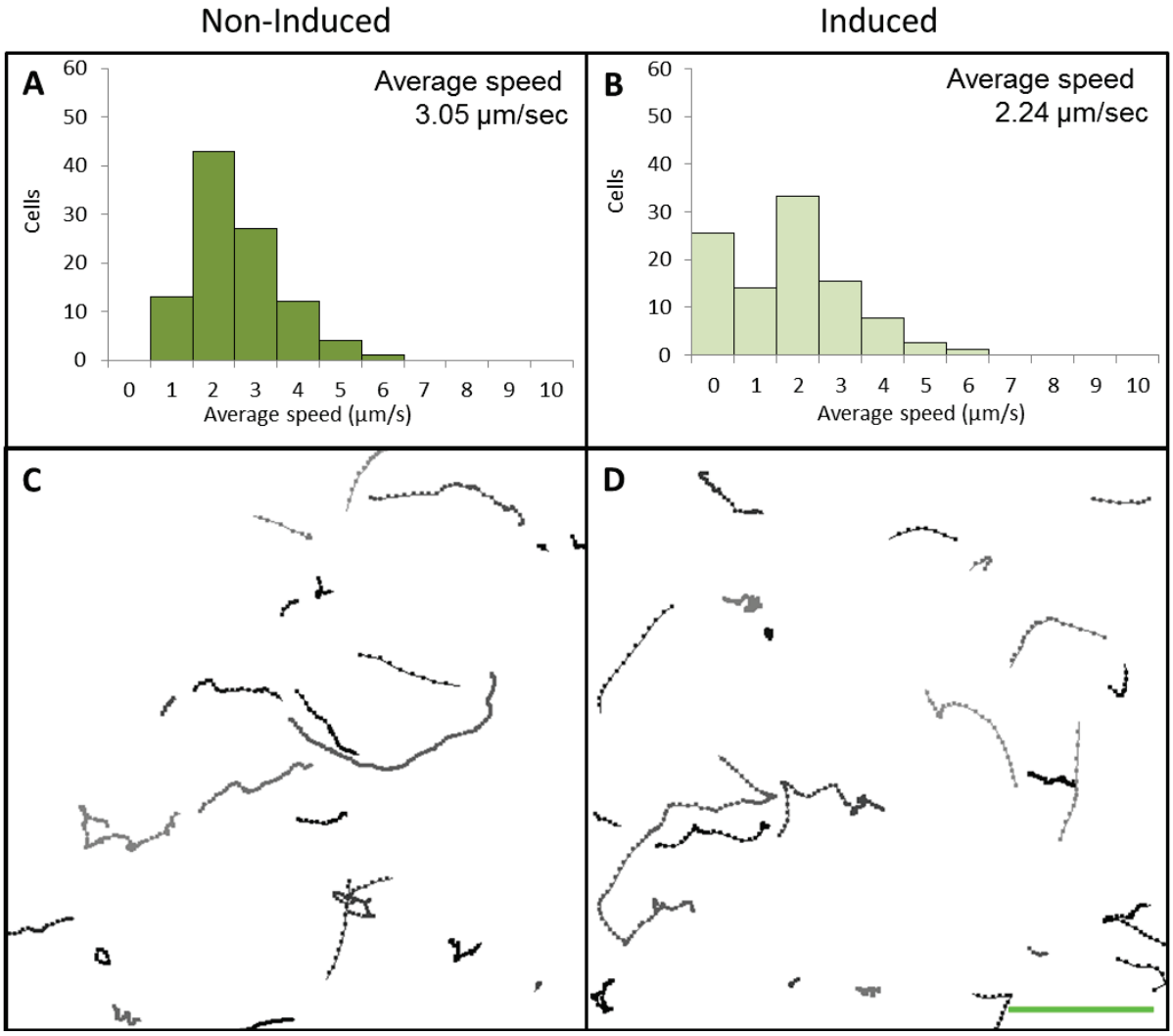


Figure 78: Knockdown of TbMNS1 affects motility

Cell speed was calculated for 100 cells. Data were plotted on a histogram for non-induced cells (A, n=100) and induced cells (B, n=100). The average cell speed in the non-induced cell population was 3.05µm/second. This decreased to an average speed of 2.24µm/second in the inducted population. This difference in speed was statistically significant (P-value = < 0.0001**). The data from induced cells were recorded after 48 hours of RNAi induction targeted against TbMNS1. Example movement tracks of non-induced (C) and induced (D) cells are shown from an area of 200µm^2 for each population (grey scale), number of tracks is arbitrary. Scale bar = 50µm.

5.4.10 Ectopically expressed YFP::TBCCDC19 localises to the flagellum

The pDex777 plasmid (Poon *et al.*, 2012) was used to ectopically express a copy of TbCCDC19 fused to YFP at the N terminus. The ectopic expression was induced by addition of doxycycline to the culture media. To assess the effect of ectopic YFP::TbCCDC19 on cell growth, a saw tooth growth chart was recorded (Figure 79; A). This compared the non-induced cell growth (Figure 79; A; black line) to induced cell growth (Figure 79; A; red line). To assess if the pDex777 vector was successfully inducing ectopic expression of YFP::TbCCDC19 a western blot was performed (Figure 79; B). An anti-GFP antibody was used to detect the YFP::TbCCDC19 fusion protein and L8C4, an antibody against the PFR, was used to check equal protein loading across lanes. Figure 79 (C) shows quantification of DNA-containing organelles through an induction of pDex777_YFP::TbCCDC19. Normal cell types fall into the 1K1N (blue), 2K1N (red) and 2K2N (green) categories. The 'others' category (purple) represents abnormal cell types including anuclete and multinucleate cells.

The linearised pDex777_YFP::TbCCDC19 construct inserts into a transcriptionally silent region of the *T. brucei* genome known as the 177bp repeats of the mini chromosomes. The pDex777 vector is an inducible system for ectopic protein expression; it is induced by the addition of doxycycline to the culture media. Therefore when the cell line is in a non-induced state the transfected construct is silent at its insertion site and no fusion protein visible (Figure 80, A). After the addition of doxycycline, the pDex777_YFP::TbCCDC19 construct is induced and the YFP::TbCCDC19 fusion protein localised to the flagellum (Figure 80, B).

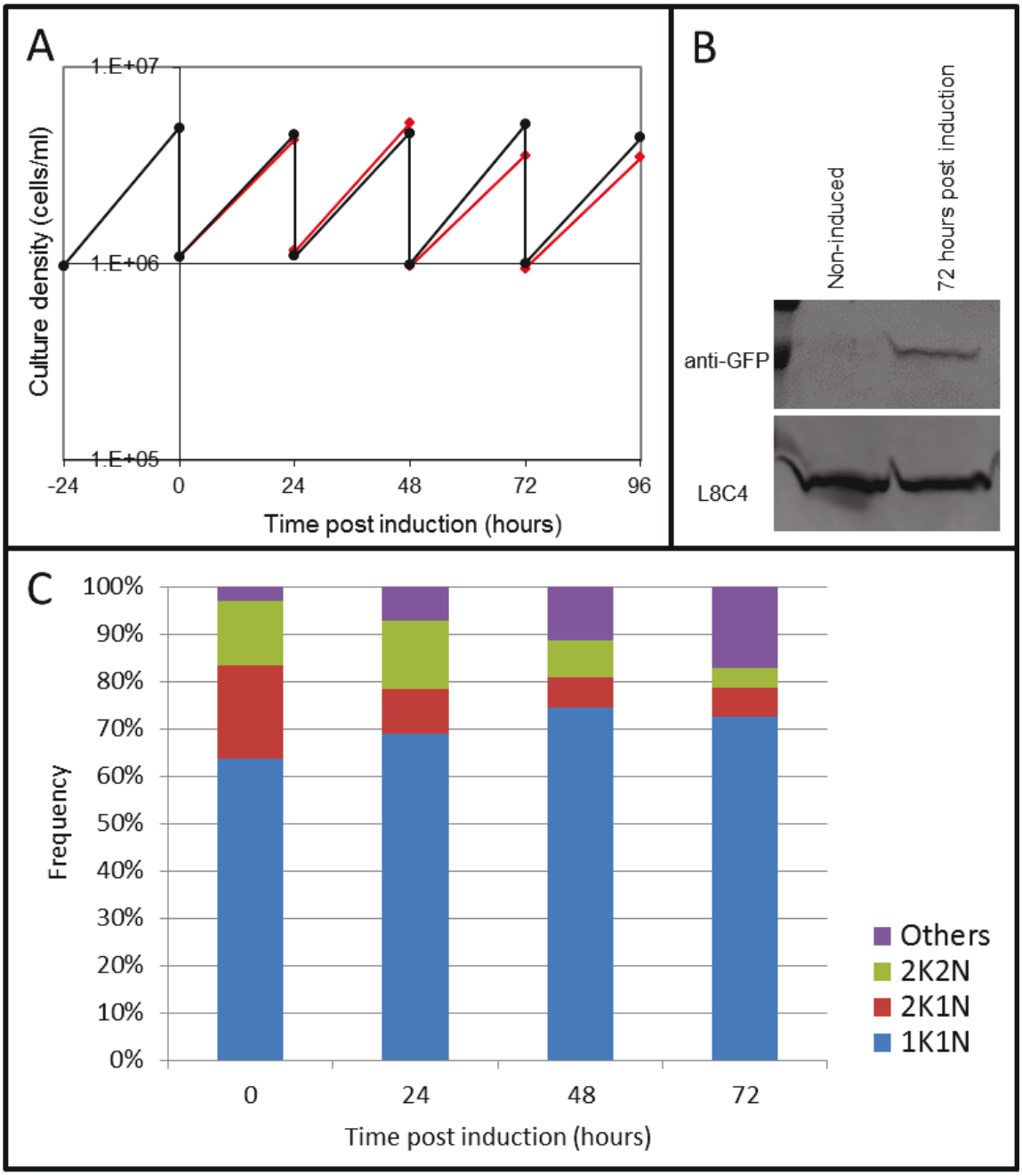


Figure 79: Ectopic expression of YFP::TbCCDC19 affects cell growth

(A) A saw tooth growth curve analysis comparing non-induced cell growth to induced culture cell growth (black and red, respectively). (B) Western blot was used to assess expression levels of YFP::TbCCDC19 using an anti-GFP antibody (top row). The PFR antibody, L8C4 was used as a loading control (bottom row). (C) Quantification of DNA containing organelles in non-induced (0 hours) and 24, 48 and 72 hours post induction (n=654, 548, 372, 702 respectively).

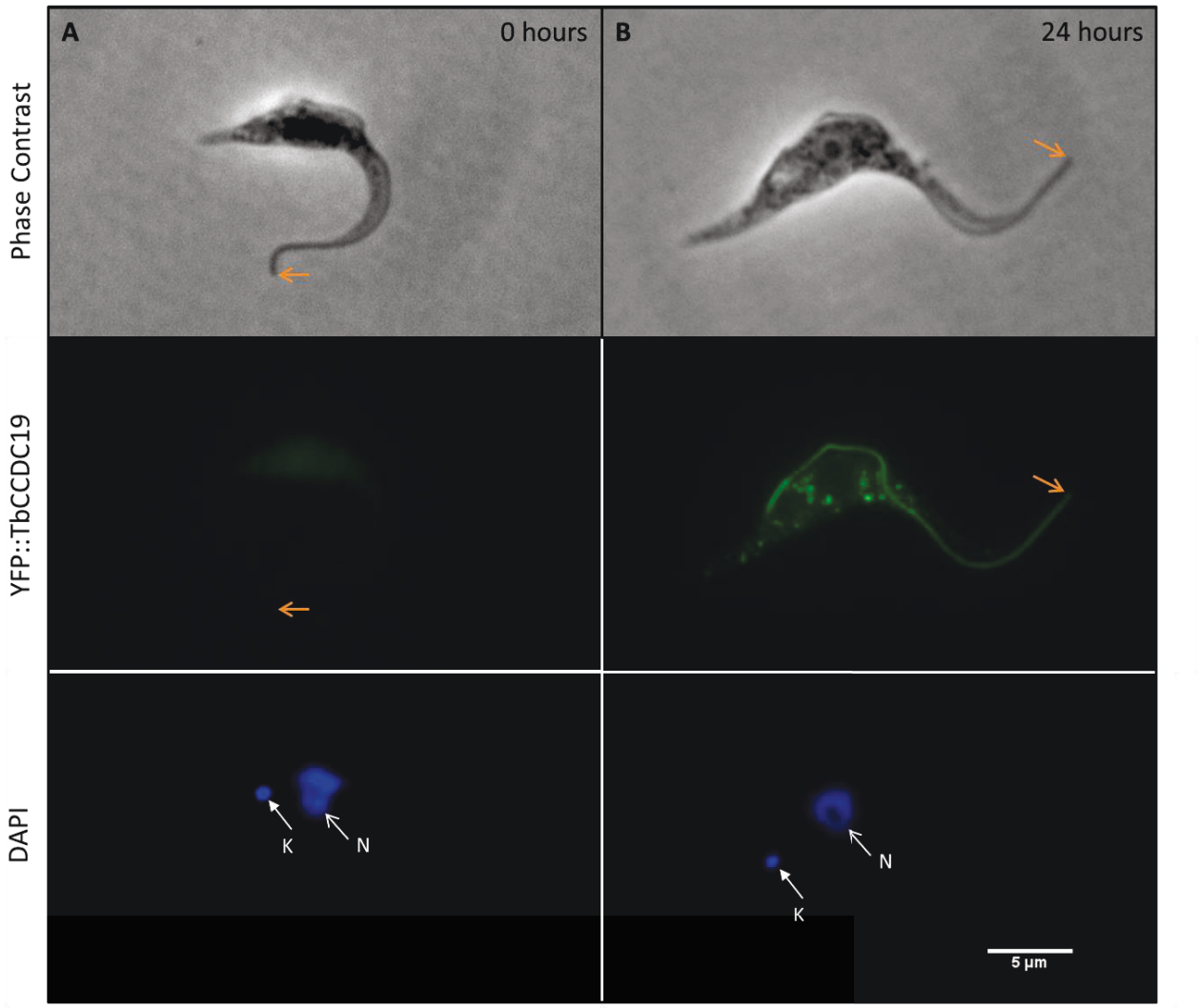


Figure 80: Exogenous expression of YFP::CCDC19 localises to the flagellum

1K1N cells from non-induced (0 hours; A) and 24 hours post induction (B). Prior to induction there is no YFP in the cell (A) at 24 hours post induction (B) there is a clear signal along the length of the flagellum, which extends to the distal tip (arrow). Scale bar 5µm.

5.4.11 TbCCDC19 is incorporated into the flagellum at the distal tip

When the doxycycline was added to the growth media to induce exogenous expression of YFP::TbCCDC19, the cells were in asynchronous growth. After induction, YFP::TbCCDC19 was first seen targeted to the flagellum after 6 hours, but only in a minority of cells. Figure 81 shows example *T. brucei* procyclic cells that were detergent extracted and fixed 12 hours post induction of YFP::TbCCDC19 expression. YFP::TbCCDC19 was localised at the distal tip of the flagellum (Figure 81; orange arrow) in cells with one flagellum and in cells that were assembling a new flagellum, YFP::TbCCDC19 localised to the new flagellum (Figure 81; pink arrow). This is evidence that TbCCDC19 is an axonemal protein and undergoes turnover at the distal tip of the established flagellum as well as being incorporated into a growing flagellum. Ectopic expression of YFP::TbCCDC19 also caused internal foci of YFP within the main cell body. There was no pattern to the localisation of the internal foci so my hypothesis is that it was excess YFP::TbCCDC19 being degraded in lysosomes and was therefore not a true localisation. To try and prevent the flagellum YFP signal being obscured by the internal YFP foci the cells were extracted by leaving the detergent on the slides for 5 minutes instead of 2 minutes. It is possible that the fixation of the cytoskeletons in methanol caused the remaining YFP protein to aggregate into the large YFP foci seen in Figure 81.

5.4.12 Ectopic expression of YFP::TbCCDC19 lead to flagellum detachment

When cells were examined by light microscopy and scanning electron microscopy (SEM) it was apparent that a sub population of cells that had YFP::TbCCDC19 ectopically expressed had detached flagella. This flagellum detachment was quantified (Figure 82) using the classifications of attached (turquoise), partially detached (purple) or fully detached (orange). In the non-induced population there were cells that were scored as having minor attachment errors, 6.19%, but by 72 hours post induction this had risen to 29.74% of the cells analysed with 12.42% having a fully detached flagellum.

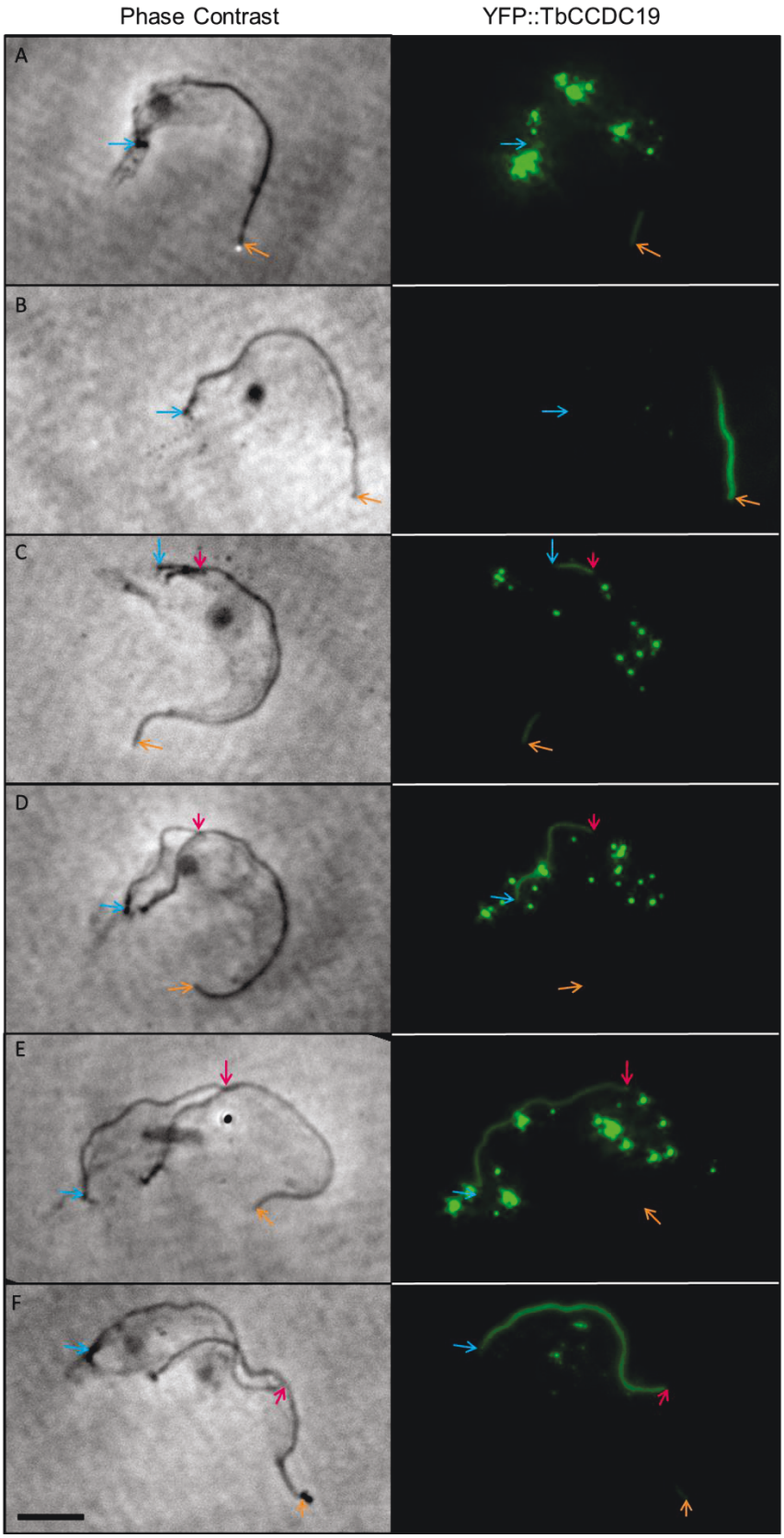


Figure 81: YFP::TbCCDC19 is incorporated into the flagellum at the distal end

Extracted cells at 12 hours post induction of YFP::TbCCDC19 from an exogenous locus. (A-B) 1K1N cells, (C) 2K1N cell with short NF, (D) 2K1N cell with long NF, (E) 2K cell with mitotic nucleus and long NF, (F) 2K2N cell with an almost full length NF. Blue arrow; proximal end of flagellum (A- B) or new flagellum (C-F), pink arrow; distal end of new flagellum, orange arrow; distal end of old flagellum. Scale bar = 5µm.

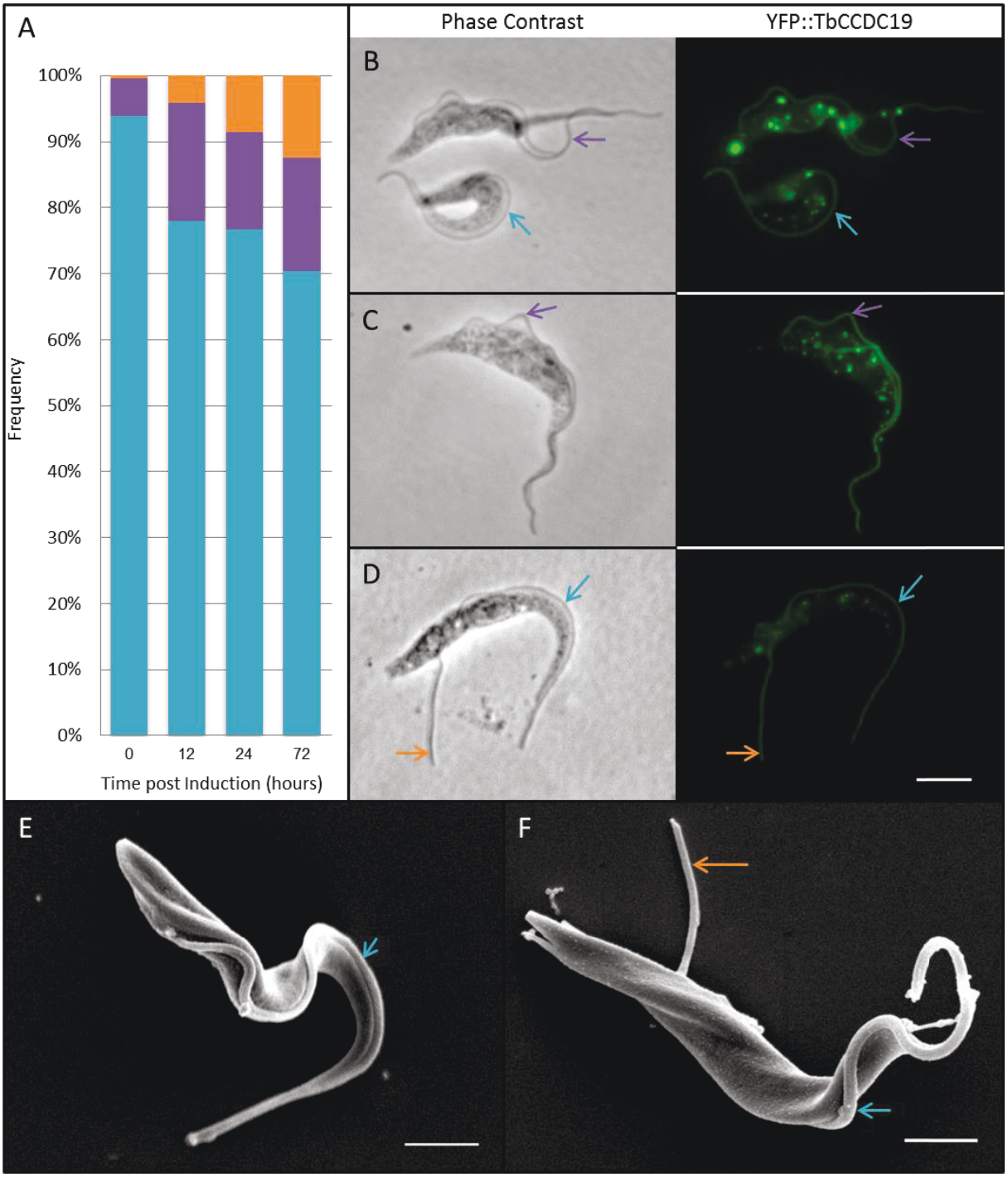


Figure 82: Quantification of flagellum detachment

(A) Degrees of flagellum detachment were classified as; totally detached (orange), partially detached/looped (purple), fully attached (turquoise). Cells were examined at 0, 12, 24 and 72 hours (n= 210, 213, 382 and 306 respectively). (B-D) Cells 24 hours post induction of ectopic YFP::TbCCDC19 expression. Scale bar = 5µm. Example cells with partial flagellum detachment (purple arrow, B and C). Example of flagellum attachment (turquoise arrow, B and D) and example of a flagellum, which is fully detached (orange arrow, D). Scanning electron micrograph of cell at 0 hours (E) with an attached flagellum (turquoise arrow). An SEM of a cell 48 hours post induction (F) with a detached flagellum (orange arrow) and an attached flagellum (turquoise arrow). Scale bars = 2µm.

5.5 Discussion

5.5.1 Localisation of TPH domain-containing proteins in *T. brucei*

Using endogenous expression systems the three TPH domain-containing proteins in *T. brucei* were localised to the flagellum. This was predicted because of the detection of TbCCDC11, TbCCDC19 and TbMNS1 in a proteomic analysis of the *T. brucei* flagellum (Broadhead *et al.*, 2006).

Work presented in this chapter shows that the three TPH domain-containing proteins in *T. brucei* were present in the growing new flagellum of cells and therefore TPH proteins are not incorporated post flagellum assembly as some cytoskeletal components are known to be (Vaughan *et al.*, 2006).

In this work TbCCDC19 was also localised to the flagellum using an ectopic expression system (Figure 80). None of these proteins had been localised in *T. brucei* before this study. The only TPH domain-containing proteins that have been previously localised are MNS1, which is localised to the flagellum in sperm of mice (Vadnais *et al.*, 2014; Zhou *et al.*, 2012) and CCDC11, which was localised to cilia in ciliated kidney epithelia of *Danio rerio* (Narasimhan *et al.*, 2014).

5.5.2 Ablation of TPH domain-containing proteins in *T. brucei* is not lethal

Individual knockdown of the three TPH domain-containing proteins in *T. brucei* by RNAi revealed that TbCCDC11 and TbCCDC19 are required for normal cell growth but knockdown of TbMN1 did not affect cell growth. These results are in agreement with the findings of a high throughput RNAi genome-wide screen (Alsford *et al.*, 2011), which did not find knockdown of any of the three TPH domain-containing proteins to be lethal in procyclic form *T. brucei* (Appendix 11). This is the first work on TPH domain-containing proteins in a uniflagellated model organism but in the mouse model, when MNS1 was knocked out the male mice were sterile due to abnormal sperm but the mice had normal left right asymmetry (Zhou *et al.*, 2012). Abnormal left right asymmetry in known as *situs inversus* and it is a common clinical phenotype of ciliopathies (Pennekamp *et al.*, 2015). Due to the published mouse knock out phenotype (Zhou *et al.*, 2012) I had expected a

growth phenotype when TbMNS1 was knocked down but it is important to remember that *T. brucei* is a single celled organism and is therefore very different from the multi cellular mouse model of disease. When DNA containing organelles were quantified the TbMNS1 sample at 72 hours post RNAi induction had the lowest percentage of 'other' abnormal cell types, 9.09% compared to 15.14% (TbCCDC11) and 15.98% (TbCCDC19). This results is surprising given that TbMNS1 RNAi had no effect on population growth (Figure 67; A), a small decrease in cell growth was expected due to the ~10% of abnormal cell types (Figure 67; B).

Future work could involve knockdown in duplicate or triplicate to assess any functional overlap of TPH domain-containing proteins in *T. brucei.* Combination knockdowns could yield more informative phenotypes, especially in light of MNS1 and CCDC11 having a common interacting protein in mice (Figure 60) (Lehti *et al.*, 2013).

5.5.3 Cell morphometric analysis

Knockdown of the three TPH domain-containing proteins in *T. brucei* by RNAi did not perturb flagellum assembly. In all three RNAi cell lines, new flagella were assembled and did not have any problems with flagellum attachment as no detached flagella were seen post induction of RNAi. Fully assembled flagella were evident by light microscopy. Comparison of flagella length between non-induced and induced cell populations concluded that there was an increase in flagellum length (Figure 69) due to ablation of each TPH domain-containing protein.

It is known that a smaller flagellum caused by knockdown of IFT components leads to a smaller trypanomastigote cell body, although the length of the posterior end is unchanged (Kohl *et al.*, 2003). In IFT mutants it is perhaps not surprising that lack of IFT trains causes a shorter flagellum to be assembled but it is interesting that the change in flagellum length only influences the cell body shape and length from the basal body to the anterior end. I was unable to find published material reporting a phenotype where the length of the flagellum increases in *T. brucei.* However, there are several known 'long flagellum mutants' in the *C. reinhardtii* field (for review see (Avasthi and Marshall, 2011; Wilson *et al.*, 2008).

The morphometric analysis of the RNAi cell phenotypes shows that there is a firm link between flagellum length and cell body size (2D surface area). There was also an increase in the length of the posterior end of the cell once each of the TPH domain-containing proteins were ablated (Figure 69).

In *T. brucei* there are several publications where protein knock down or overexpression has caused the common phenotype of extended posterior ends, also referred to as elongation mutants or the 'nozzle' phenotype (Bonhivers *et al.*, 2008b; Hammarton *et al.*, 2004; Hendriks *et al.*, 2001; Subota *et al.*, 2011). Although it may seem strange that knockdown of unrelated proteins can cause a common phenotype, it is not surprising in the context of the constraint of the subpellicular microtubule corset.

5.5.4 The role of TPH domain-containing proteins in propulsive motility

Despite there being lots of published work on the swimming speed and behaviour for bloodstream form trypomastigotes (Heddergott *et al.*, 2012; Uppaluri *et al.*, 2011), there is a lack of data published for procyclic form trypanomastigote motility. One published study recorded the speed of procyclic form *T. brucei* cells as 4.2µm/sec (Weisse *et al.*, 2012). The analysis of *T. brucei* procyclic form cell speed presented in this chapter (Figure 76, Figure 77 and Figure 78) was performed on data captured from cells that were imaged on a non-temperature controlled platform. It is important to highlight that Weisse and colleagues (2012) used a heated stage so it is expected that their recorded speed would be higher.

Independent ablation of all three TPH domain-containing proteins decreased the speed of procyclic *T. brucei.* Another phenotype observation was the increase in flagellum length (Figure 69) in the induced populations. Work in the biflagellated algae, *C. reinhardtii* has also found that an increase in flagellar length reduces cell speed (Khona *et al.*, 2013).

The results presented in this chapter indicate that all three TPH domain-containing proteins are involved in flagellum motility. These findings agree with published experimental data from CCDC11 and MNS1. The sperm of male -/- MNS1 mice have immotile sperm flagella (Zhou *et al.*, 2012) and knockdown of CCDC11 in zebrafish

disrupts the ciliary beat (Narasimhan *et al.*, 2014). Although nasal cilia from a human patient lacking CCDC11 exhibited normal ciliary beat (Narasimhan *et al.*, 2014).

5.5.5 Ectopic expression of TbCCDC19

Localisation of TbCCDC19 to the flagellum was confirmed using ectopic expression of YFP::TbCCDC19 (Figure 80), which also allowed identification of the pattern of incorporation of YFP::TbCCDC19 into the axoneme (Figure 81). It was known in *C. reinhardtii* that axoneme components were included at the distal tip of the flagellum (Marshall and Rosenbaum, 2001) as that is the location of the plus end of microtubules (Johnson and Rosenbaum, 1992). The polarity of axonemal microtubules in the *T. brucei* flagellum is the same (Figure 11) and therefore the distal tip is the growing end (plus end) but axonemal component incorporation has not been shown to date. In *T. brucei* only the pattern of PFR (Bastin *et al.*, 1999a) and FAZ (Sunter *et al.*, 2015) assembly have been investigated.

Ectopic expression of YFP::TbCCDC19 caused flagellum detachment from 12 hours post induction onwards. It is unlikely that solely expression of YFP caused this phenotype as flagellum detachment was not seen for endogenous expression of YFP::TbCCDC19. Expression of pDex777-GFP was not reported to cause detached flagella and the induced procyclic *T. brucei* were able to establish an infection with tsetse flies (Poon *et al.*, 2012), which indicates their motility was unaffected by the GFP. Without knowledge of TbCCDC19 interacting partners, it is hard to speculate why or how TbCCDC19 is involved in flagellum attachment.

When YFP::TbCCDC19 is ectopically expressed the YFP signal is seen within the cell body as punctate fluorescent aggregations as well as to the flagellum (Figure 80 and Figure 81). It is unlikely that the YFP signal within the cell body is a true localisation of TbCCDC19 because when the YFP-TbCCDC19 fusion protein is expressed at an endogenous level YFP::TbCCDC19 is localised to the flagellum and no intracellular YFP is seen. If the cell body pool of TbCCDC19 were true, the YFP signal would be present in a distinct pattern or in the same place in every cell but the position and sizes of the YFP aggregations within the cells varies. The cell body signal could be a result of the excess YFP-TbCCDC19 being

recycled by the lysosomal pathway but there is only one anti-YFP band detected by Western blot (Figure 79) so it does not appear that the YFP::TbCCDC19 is being cleaved into fragments for degradation. Therefore the intracellular YFP aggregations are likely to be an artefact of ectopic expression of TbCCDC19 rather than a true localisation.

5.5.6 CCDC11

In this chapter I have confirmed that TbCCDC11 is a component of the flagellum in PCF *T. brucei*. CCDC11 has been localised to the flagellum in one other model system; *D. rerio* (Narasimhan *et al.*, 2014). DrCCDC11 was localised to the motile cilia of kidney epithelial cells but was also localised to the basal body/centriole of cells in Kupffer's vesicle (Narasimhan *et al.*, 2014). Kupffer's vesicle is a specialised organ in *D. rerio,* which is essential during development for assigning left-right symmetry (Matsui and Bessho, 2012). It is functionally analogous to the node in mammalian embryogenesis (for review see (Amack, 2014)). This differential localisation depending on cell type in *D. rerio* is interesting and it will be important to evaluate CCDC11 localisation across cell types in other model multicellular organisms. After identifying CCDC11 as a ciliary component in zebrafish the authors identified a patient with total *situs inversus* who had a mutation in CCDC11 (patient OP-1069-II1)(Narasimhan *et al.*, 2014) (Appendix 12). CCDC11 also localised to the cilia of nasal epithelia in healthy humans but was absent from the cilia of patient OP-1069-II1 (Narasimhan *et al.*, 2014). Examination of patient's OP-1069-II1 nasal epithelial cilia by TEM revealed no obvious ultrastructural defects (Narasimhan *et al.*, 2014).

A clinical case study on two brothers; patient 2541 and patient 2535, reports a frame shift mutation in the gene coding for CCDC11 (Appendix 13) (Perles *et al.*, 2012). The mutation was within the TPH domain although the authors do not comment on this protein feature. Patient 2541 had digital clubbing and *situs inversus* of the abdomen and later died. Patient 2535 also had *situs inversus* but presented as externally healthy. Further examination of patient 2535 revealed normal cilia beat for nasal epithelial cells and cilia were ultrastructurally normal when examined by thin section TEM (Perles *et al.*, 2012). Sperm motility of patient 2535 was also normal (Perles *et al.*, 2012). The difference in

clinical symptoms between patients 2541 and 2535 is an example of mutational load and demonstrates the clinical spectrum of ciliopathy diseases.

5.5.7 CCDC19

The work presented in this thesis has directly localised a CCDC19 ortholog to the eukaryotic flagellum for the first time although CCDC19 has previously been detected in cilia/flagella proteomes (Table 18). CCDC19 is also known as nasopharyngeal epithelium specific gene 1 (NESG1) was first cloned in 1999 (Li et al., 1999) and found to be a prognosis factor for nasopharyngeal carcinoma (Liu et al., 2011a; Liu et al., 2011b) as it has a proposed function as a tumour suppressor (Liu et al., 2012; Liu et al., 2011b). The connection between cilia formation, cell cycle regulation and cancer is well known (Bettencourt-Dias et al., 2011; Hassounah et al., 2012) although it is not clear if this is the case for CCDC19. The exact mechanism of how HsCCDC19 is involved with nasopharyngeal carcinomas is unknown but Li *et al.,* (2011a) suggest it could be due to a cyclin-regulating function of HsCCDC19 that controls cell proliferation by inhibiting the cell cycle through driving expression of p21 and therefore the inhibition of cyclin A1 (Liu *et al.*, 2011a).

5.5.8 MNS1

As mentioned above, the homozygous mouse model for MNS1, -/-MNS1, exhibited abnormal sperm flagellum organisation (Zhou *et al.*, 2012) but -/-MNS1 was not lethal. The male mice were sterile, presumably due to short, immotile sperm flagella. Examination of the sperm flagellum by thin section TEM revealed disrupted architecture; the axonemal microtubule doublets are assembled but the doublets and outer dense fibres are disorganised (Zhou *et al.*, 2012). Further examination revealed that 63% of -/-MNS1 mice had semi or total *situs invertus*, likely due to abnormal motility of nodal cilia.

MNS1 binds with itself *in vitro*, which the authors propose is a feature of the coiled-coil domains (Zhou *et al.*, 2012). When MNS1 was expressed in murine fibroblast cells filaments were formed (Zhou *et al.*, 2012) reminiscent of overexpression of another coiled-coil protein, TbDIP13 (Price *et al.*, 2012). It is not known if CCDC19 or CCDC11 also

have the ability to form homodimers or even if the TPH domain-containing proteins could form heterodimers.

5.5.9 Future work

I have confirmed that the three TPH domain-containing proteins of *T. brucei* localise to the flagellum. To build on the significance of the results presented here it would be important to identify TbCCDC11-, TbCCDC19- and TbMNS1- interacting proteins. This could be achieved using Förster resonance energy transfer (FRET), bimolecular fluorescence complementation (BiFC) (Hu *et al.*, 2006) or a proximity labelling approach such as BioID (Roux *et al.*, 2012)

An approach like BiFC or BioID would be more appropriate to the size of the model system as the only FRET experiments performed so far on *T. brucei* have been in plate format, assessing a population rather than on individual cells (Wang *et al.*, 2013). The disadvantage of BiFC is that two suspecting interacting proteins would have to be known in order to carry out the experiment but with BioID you only need to know your 'bait' protein to identify interacting partners. BiFC would be a complementary method to confirm interactors that were identified through BioID and would help to identify the terminus that is involved in the interaction. Additionally, the BioID method has already been adapted to *T. brucei* (Morriswood *et al.*, 2013).

There are scattered examples in the literature of small scale interaction studies performed in *T. brucei* but nothing on the scale of the human interactome (Rual *et al.*, 2005). One such interaction study identified TbMNS1 and TbCCDC19 as possible interactors of TbDIP13 (Price *et al.*, 2012). Interestingly, RNAi mediated ablation of TbDIP13 in the BSF caused a reduction in TbCCDC19 detected in flagella extracts compared to the parental cell line (Price *et al.*, 2012).

The work presented in this chapter presented a preliminary investigation into the effect that TbCCD11/TbCCDC19/TbMNS1 knockdown has on cell morphology. I have shown, using light microscopy that individual knockdown of these three proteins results in elongation of the posterior end of the cell body but no attempt has been made to analyse

the effect on internal organelle quantity or distribution. Volumetric electron microscopy (EM) techniques such as serial block face SEM (SBF-SEM) or focused ion beam scanning electron microscopy (FIB-SEM) would enable whole cell phenotype analysis.

5.5.10 Conclusions

I have used endogenous expression to localise each of the three TPH domain-containing proteins conserved in *T. brucei* to the flagellum. I have used endogenous and exogenous expression and knockdown of these proteins to investigate their functional role within the flagellum of *T. brucei*. The effect on cell shape and motility proves that all three TPH domain-containing proteins are functional and important to maintain proper propulsive motility. This confirms the hypothesis that TPH domain-containing proteins within *T. brucei* are components of the flagellum and are important for motile flagella function.

6. Functional analysis of Centrin2 in *T. brucei*

6.1 Collaborations

Some of the work presented in this chapter was performed in collaboration with other researchers. I (SB) have performed the work presented in this chapter unless otherwise stated. Dr Katie Towers (Oxford Brookes University) designed the primers for, and cloned, the p2T7_TbCen2 RNAi vector. I transfected the p2T7_TbCen2 vector into the YFP::TbCen2 cell line and performed all experiments on those cells. I grew the procyclic cells and fixed the cell samples to produce resin blocks for standard TEM and SBF-SEM. I trimmed the resin blocks for TEM and collected the data. For SBFSEM of the wildtype PCF sample the resin block was sent to Dr Tobias Starborg (University of Manchester) for data collection. For SBFSEM of the TbCen2 sample, Robbie Crickley (Oxford Brookes University) trimmed the resin blocks and Dr Louise Hughes (Oxford Brookes University) operated the SBF-SEM to collect data. Dr Sam Dean (University of Oxford) provided the primers for TbHydin pPOT PCR, SB performed the PCR and subsequent experiments.

6.2 Introduction

There are 5 centrin proteins conserved in kinetoplastids (Appendix 14). This chapter is focussed on TbCen2 (Tb927.8.1080), specifically in its relation to basal body maturation and flagellum assembly in *T. brucei*. *T. brucei* is a well-established experimental model organism used to study basal body ultrastructure (Lacomble *et al.*, 2009; Lacomble *et al.*, 2010) and flagellum morphogenesis (Bastin *et al.*, 1999a; Kohl *et al.*, 1999).

This chapter describes the functional characterisation of TbCen2 using light and electron microscopy techniques. This approach identified that there is a fluctuation of TbCen2 associated with the mature basal body and the pro-basal body across the normal cell cycle. Ablation of TbCen2 disrupted flagellum attachment, cell morphology and cytokinesis. Thin section TEM and serial block face SEM were employed to investigate the ultrastructural changes to procyclic *T. brucei* when TbCen2 was knocked down.

6.3 Aims

This chapter focusses on the *T. brucei* centrin, TbCen2 (Tb927.8.1080) with the aim of understanding of the role that TbCen2 has to play in the duplication, segregation and maturation of basal bodies. The function of TbCen2 in procyclic form *T. brucei* was assessed using light and electron microscopy techniques.

The research aims of this work are:

- To determine the localisation of TbCen2 within procyclic *T. brucei*
- To characterise the function of TbCen2 through ablation of the protein

6.4 Results

6.4.1 YFP::TbCen2 localises to the basal bodies, bilobe and flagellum in *T. brucei*

To study the localisation of TbCen2 (Tb927.8.1080), a stable cell line was generated that constitutively expressed the YFP fusion protein, YFP::TbCen2 from one endogenous locus. This was achieved at the N terminus of TbCen2 using a long primer PCR based method for gene tagging known as pPOT (Dean *et al.*, 2015). See section 2.3.1 Primer design and plasmid construction.

YFP::TbCen2 localised to the flagellum, basal body and bilobe in 1K1N cells (Figure 35; A and E). This stage of the cell cycle was identified by the presence of a single nucleus (N, open white arrowhead) and a single flagellum, which is subtended by a single set of basal bodies (BB) and a single kinetoplast (K, closed white arrowhead). As the cell cycle continues the basal bodies duplicate and the kinetoplast DNA undergoes S-phase to produce a 2K1N cell (Figure 35; B). At this stage the cell only has one nucleus but the kinetoplast has divided and the basal bodies have segregated, with each mature basal body assembling a flagellum. The cell now has 2 flagella, known as the old flagellum (OF) and new flagellum (NF) (Figure 35; B and C). The new flagellum grows alongside the old flagellum (Briggs *et al.*, 2004b) and is always positioned more posterior to the old flagellum (Robinson *et al.*, 1995). The YFP::TbCen2 signal was present in the NF as it extended (Figure 35; B and C), which meant it is not a post assembly addition. The final stage in the cell cycle prior to cytokinesis is a 2K2N cell (Figure 35; C), following mitosis to

create two nuclei (open white arrowheads). The YFP::TbCen2 signal was present in both flagella (Figure 35; C).

The area of the 1K1N cell (Figure 35; A) marked out by the dotted box is displayed in the inset panels (Figure 35; D-F). The phase contrast panel (Figure 35; D) shows the cell had two basal bodies (a pair) and a single flagellum. The YFP::TbCen2 signal (Figure 35; E) was seen localising at the basal bodies, the bilobe (Figure 35; E, bracket) and the flagellum. The basal bodies were also immunolabelled with BBA4 (Woods *et al.*, 1989b), an antibody that recognises an unknown antigen on both basal bodies. BBA4 is a well characterised antibody within the trypanosome field (Kohl *et al.*, 2003; Morgan *et al.*, 2005) and labels both the mature basal body and the pro-basal body (Figure 35; F, arrows). There was a gap between the basal body fluorescence signal and the bilobe/flagellum signal of YFP::TbCen2 (Figure 35; E). This gap is due to an absence of TbCen2 from the transition zone of the flagellum.

In summary, fluorescence microscopy showed that YFP::TbCen2 localised to the flagellum, basal bodies and bilobe in procyclic *T. brucei* in detergent extracted cytoskeletons (Figure 35), therefore TbCen2 is confirmed as a cytoskeletal protein in *T. brucei*.

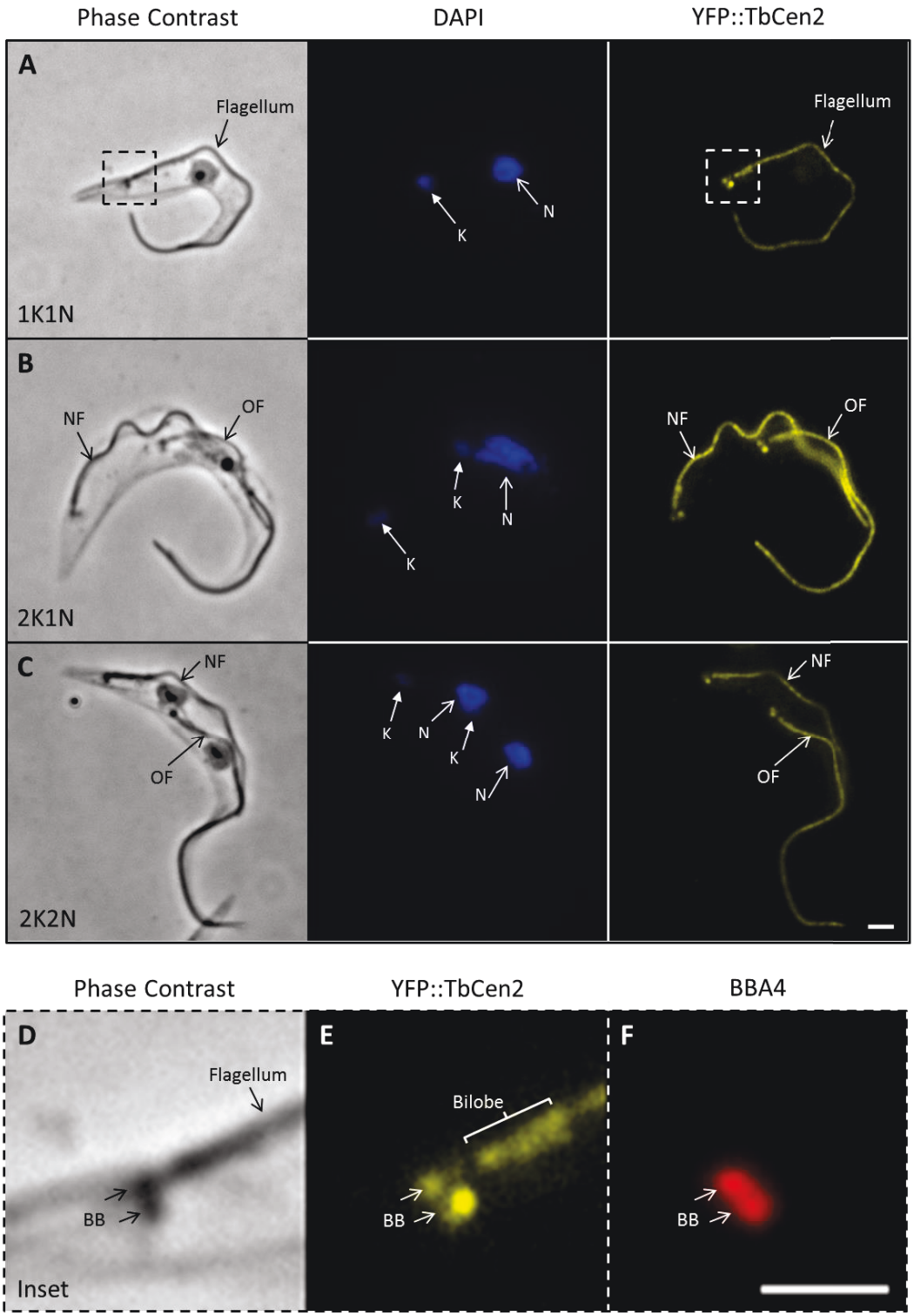


Figure 83: Localisation of YFP::TbCen2

TbCen2 was tagged at the N terminus with YFP (YFP::TbCen2; yellow) and expressed from the endogenous locus in procyclic *T. brucei*. Cells were detergent extracted, fixed and labelled with DAPI to stain DNA.

The YFP::TbCen2 signal can be seen localised to the flagellum in 1K1N (A), 2K1N (B) and 2K2N (C) cells. Insets (D-F) show the proximal end of the flagellum from the 1K1N cell (A). (D) Phase contrast image of the basal bodies and the flagellum. (E) YFP::TbCen2 signal, which is localised to the basal bodies (white arrows), the bilobe (bracket) and the flagellum. (F) Immunolabelling of the basal bodies with BBA4. NF; new flagellum, OF; old flagellum, BB; basal bodies. Scale bars = 2µm

6.4.2 Endogenous YFP::TbCen2 levels at the basal bodies fluctuate

Cells were fixed and immunolabelled with an anti-centrin antibody, 20H5 (Sanders and Salisbury, 1994). Different strength signals were observed when comparing the mature basal body and the pro-basal body of procyclic form cells (K. Towers, personal communication). Unfortunately, 20H5 labels more than one trypanosomal centrin (Absalon *et al.*, 2008; de Graffenried *et al.*, 2008) and it was not possible to determine why differential labelling of the basal bodies occurs.

In this study, when TbCen2 was tagged with YFP at the endogenous locus and therefore expressed under the native promoter, an uneven fluorescence signal was evident when cells were examined by light microscopy. To detect any pattern from this phenomenon, the occurrence of YFP::TbCen2 was visually classified and it was found that the pattern of YFP::TbCen2 fell into one of three categories (Figure 84). Initially, only cells with a single flagellum (1F) were considered. These cells only had one pair of basal bodies consisting of one mature basal body and one pro-basal body. 40.9% of the cell population had a signal of YFP::TbCen2 from both the mature basal body (MBB, closed headed arrow) and the pro-basal body (PBB, open headed arrow) that appeared equally bright in signal strength (Figure 84; A). 11.4% of the population had a bright YFP::TbCen2 signal on the PBB but a weaker, dim YFP::TbCen2 signal on the MBB (Figure 84; B). 47.7% of the population retained a bright YFP::TbCen2 signal on the PBB but no signal was visible from the MBB (Figure 84; C).

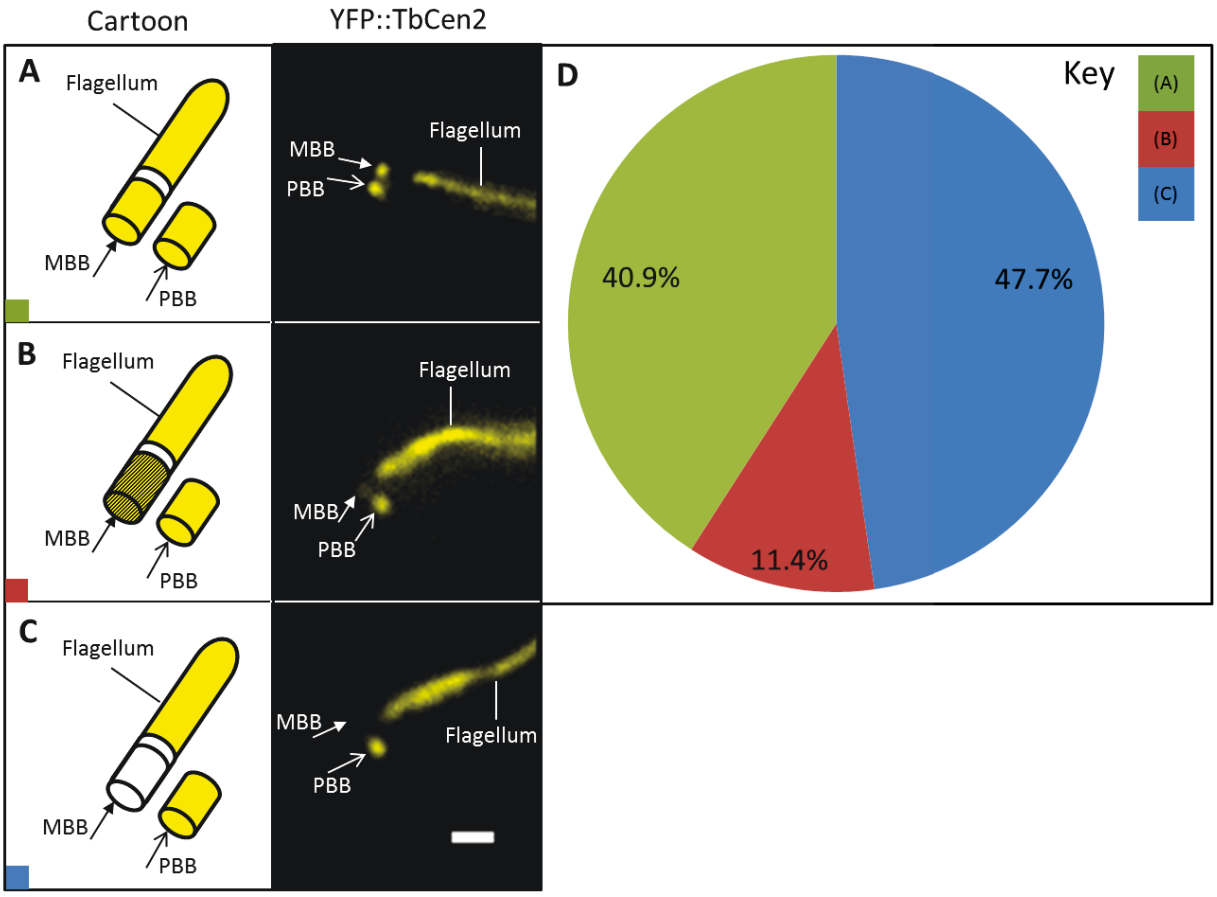


Figure 84: YFP::TbCen2 levels vary at the basal bodies in 1F cells

The signal intensity of YFP::TbCen2 was visually assessed in cells with a single flagellum and therefore one pair of basal bodies. Three patterns of YFP::TbCen2 fluorescence were seen, a cartoon model and example micrograph are given for each pattern. (A) Both the mature basal body (MBB) and the pro-basal body (PBB) were bright. (B) The MBB was dim but the PBB was bright. (C) The MBB was dark and not visible but the PBB was bright. In all three categories, the flagellum signal remained bright. (D) The occurrence of each pattern was counted in cells with a single flagellum (n=88) and values are expressed as a percentage of the population (A-C). Scale bar = 1µm.

To confirm any difference in signal, YFP::TbCen2 was imaged and the signal intensity was quantified in ImageJ (see section 2.6.3 Quantification of fluorescence intensity). The signal intensity of YFP::TbCen2 was measured at the mature basal body and the pro-basal body. For this analysis presented in this section, the YFP::TbCen2 signal was visualised and imaged directly without immunolabelling the cells. Images were collected of basal body pairs from 1F cells with one pair of basal bodies (Figure 85; A).

There was a statistically significant difference (P-value = < 0.0001***) in the fluorescence intensity of the mature basal body and the pro-basal body in cells (n=42) that only has one pair of basal bodies (Figure 85; B). A paired *t* test was used to assess statistical significance. The average signal for MBB was 0.99 a.u./μm^2 (Figure 85; B, red) compared to 2.12 a.u./μm^2 for PBB (Figure 85; B, orange). This is evidence for the level of YFP::TbCen2 being affected by the cell cycle stage or by basal body age.

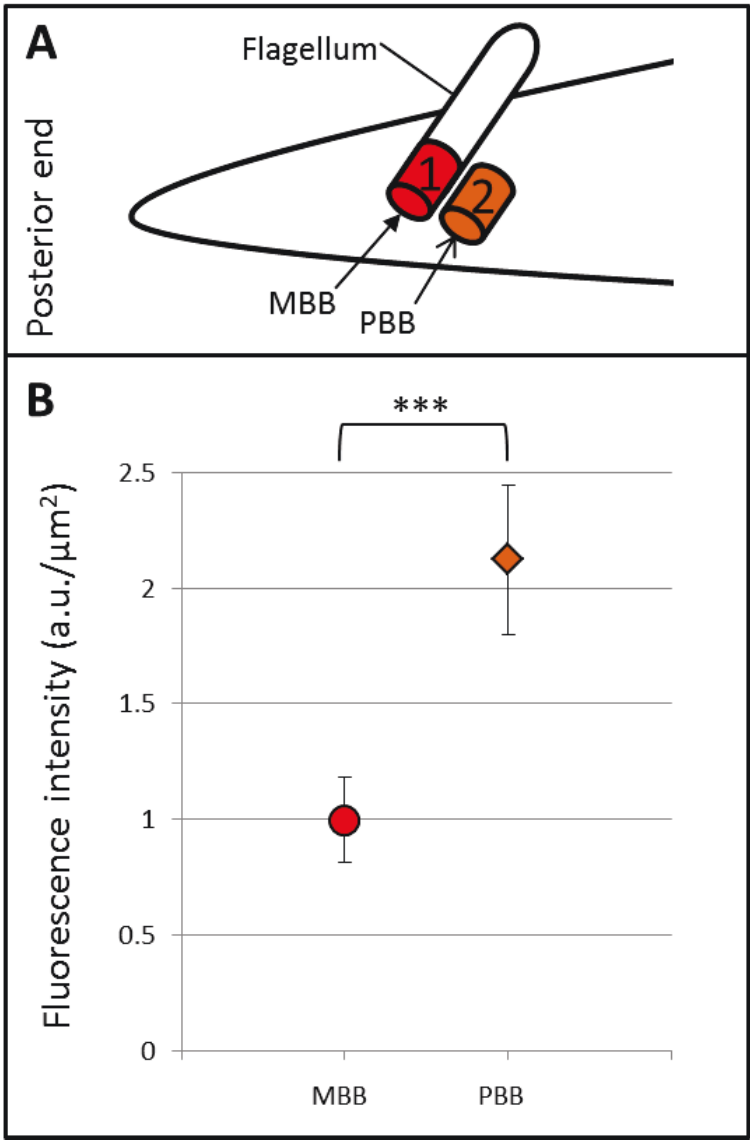


Figure 85: Quantification of YFP::TbCen2 signal at the basal bodies in 1F cells

The intensity of the endogenous YFP::TbCen2 fluorescence at the basal bodies was quantified in cells with 1 flagellum/ 2 basal bodies (A). Cells expressing YFP::TbCen2 were detergent extracted, fixed in -20^{0}C methanol and imaged on the same day. Data were collected as Z-stacks to eliminate focal plane bias and the sum signal from 10 optical sections was measured in each instance. Images were processed with equal settings. (B) The fluorescence intensity in arbitrary units per µm^2 is shown (mean ± s.d) for the mature basal body (MBB) and the pro-basal body (PBB), which was normalised against the background. A paired *t* test found a significant statistical difference (P-value = <0.0001***) between the signal intensity of the MBB and the PBB in cells with a single flagellum (n= 42). Error bars indicate standard deviation.

To assess if there was a link between YFP::TbCen2 signal and cell cycle stage the analysis was repeated with 2F cells that had two pairs of basal bodies; one pair of basal bodies associated with the new flagellum and one pair of basal bodies was associated with the old flagellum. Each pair of basal bodies consisted of one mature basal body and one pro-basal body. Images were collected of basal body pairs from 2F cells that contained two pairs of basal bodies (n=45) (Figure 86; A). The same sized ROI as before was used to quantify fluorescence signal intensity for all basal bodies. In the 2F cells, there was a statistically significant difference (P-value = < 0.0001***) in the fluorescence intensity of YFP::TbCen2 between the mature basal body (Figure 86; B, orange) and the pro-basal body (Figure 86; B, blue) associated with the new flagellum (NF). The NF is located more posteriorly in the cell than the old flagellum (OF). There was also a statistically significant difference (P-value = < 0.0001***) in the fluorescence intensity of YFP::TbCen2 between the mature basal body (Figure 86; B, red) and the pro-basal body (Figure 86; B, green) associated with the old flagellum (OF). A paired *t* test was used to assess statistical significance between MBB and PBB fluorescence intensities.

The average signal for a MBB associated with the NF was 1.45 a.u. /μm^2 (Figure 86; B, orange) compared to 1.33 a.u./μm^2 for MBB associated with the OF (Figure 86; B, red). A paired *t* test comparing the data of both MBBs determined there was no statistical difference (P-value = 0.1223). The same is true when the PBB of the NF and OF were compared against each other. The average signal for a PBB associated with the NF was 1.90 a.u. /μm^2 (Figure 86; B, blue) compared to 1.99 a.u. /μm^2 for PBB associated with the OF (Figure 86; B, green). A paired *t* test comparing the data of both PBBs determined there was no statistical difference (P-value = 0.2658). That is unsurprising as both of the PBB are formed in the same cell cycle and are therefore the same 'age'. When the MBB of the NF was paired with the PBB of the NF, the difference in fluorescence intensity was statistically significant (P-value = <0.0001***). This showed that even though the MBB of the NF (2, orange) had recently matured, there was still a significant difference between the associated YFP::TbCen2 of the MBB and the PBB. When the MBB of the OF was paired with the PBB of the OF, the difference in fluorescence intensity was statistically significant (P-value = <0.0001***). This was expected as the MBB of the OF is the oldest BB of the cell and the PBB of the OF is the youngest because it was formed in that cell cycle.

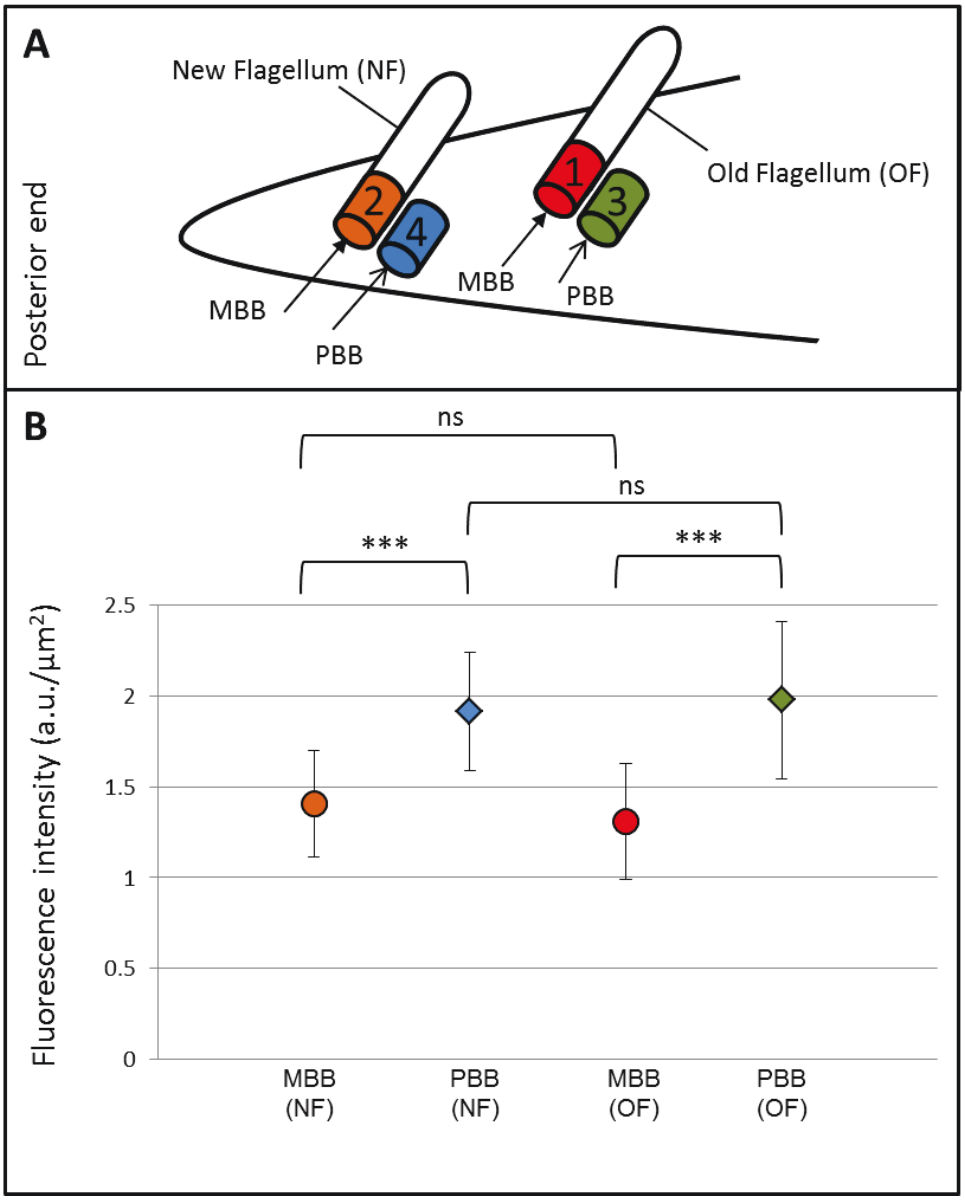


Figure 86: Quantification of YFP::TbCen2 signal at the basal bodies in 2F cells

The intensity of the endogenous YFP::TbCen2 fluorescence at the basal bodies was quantified in cells with 2 flagella/ 4 basal bodies (A). The new flagellum (NF) is more posterior than the old flagellum (OF). Cells expressing YFP::TbCen2 were detergent extracted and fixed in -20^{0}C methanol. Data were collected as Z-stacks to eliminate focal plane bias and the sum signal from 10 optical sections was measured in each instance. Images were processed with equal settings. (B) The fluorescence intensity in arbitrary units per µm^2 is shown (mean ± s.d) for the mature basal bodies (MBB, circles) of the new and old flagellum and the pro-basal bodies (PBB, diamonds) of the new and old flagellum. Fluorescence intensity readings were normalised against the background. A paired *t* test found a significant statistical difference (P-value = <0.0001***) between the signal intensity of the MBB (NF) and the PBB (NF) and a significant statistical difference (P-value = <0.0001***) between the signal intensity of the MBB (OF) and the PBB (OF). No statistically significant difference (ns) was found between the signal intensity of the MBB (NF) and the MBB (OF) (P-value = 0.1223) or between the signal intensity of the PBB (NF) and the PBB (OF) (P-value = 0.2658). Cells with 2 pairs of basal bodies were included in this analysis (n= 45). Error bars indicate standard deviation.

6.4.3 Flagellum YFP::TbCen2 levels do not fluctuate

YFP::TbCen2 localises to more than one organelle in procyclic *T. brucei* cells (Figure 83). It has been established that the quantity of YFP::TbCen2 that localised to the basal bodies varied between cells (Figure 85 and Figure 86). To determine if other pools of YFP::TbCen2 localisation also fluctuated, the fluorescence intensity of YFP::TbCen2 was measured at the flagellum. The signal intensity was compared between flagella of different ages to assess if there was a difference between the new and old flagella of duplicating cells and the single flagellum of non-duplicating cells (Figure 87).

Cells with one flagellum (1F), 1K1N cells, had an average YFP::TbCen2 fluorescence intensity of 1.80 a.u./μm^2 (n=42). Cells with two flagella (2F), 2K1N and 2K2N cells, had an average YFP::TbCen2 fluorescence intensity of 1.99 a.u./μm^2 from the new flagellum (NF, n=35) and 1.92 a.u./μm^2 from the old flagellum (OF, n=44). The YFP::TbCen2 incorporated into the flagellum is stable and is not affected by the cell cycle or cell age. This showed that the pool of TbCen2 that is present in the flagellum is distinct from other stores of TbCen2 within the cell and does not fluctuate through the cell cycle. Unpaired *t* tests between the flagellum types revealed that statistically there was no significant difference; 1F and NF_2F (P-value = 0.1754), 1F and OF_2F (P-value = 0.2883), NF_2F and OF_2F (P-value = 0.6112).

In conclusion, the signal intensity of YFP::TbCen2 did not differ between flagella.

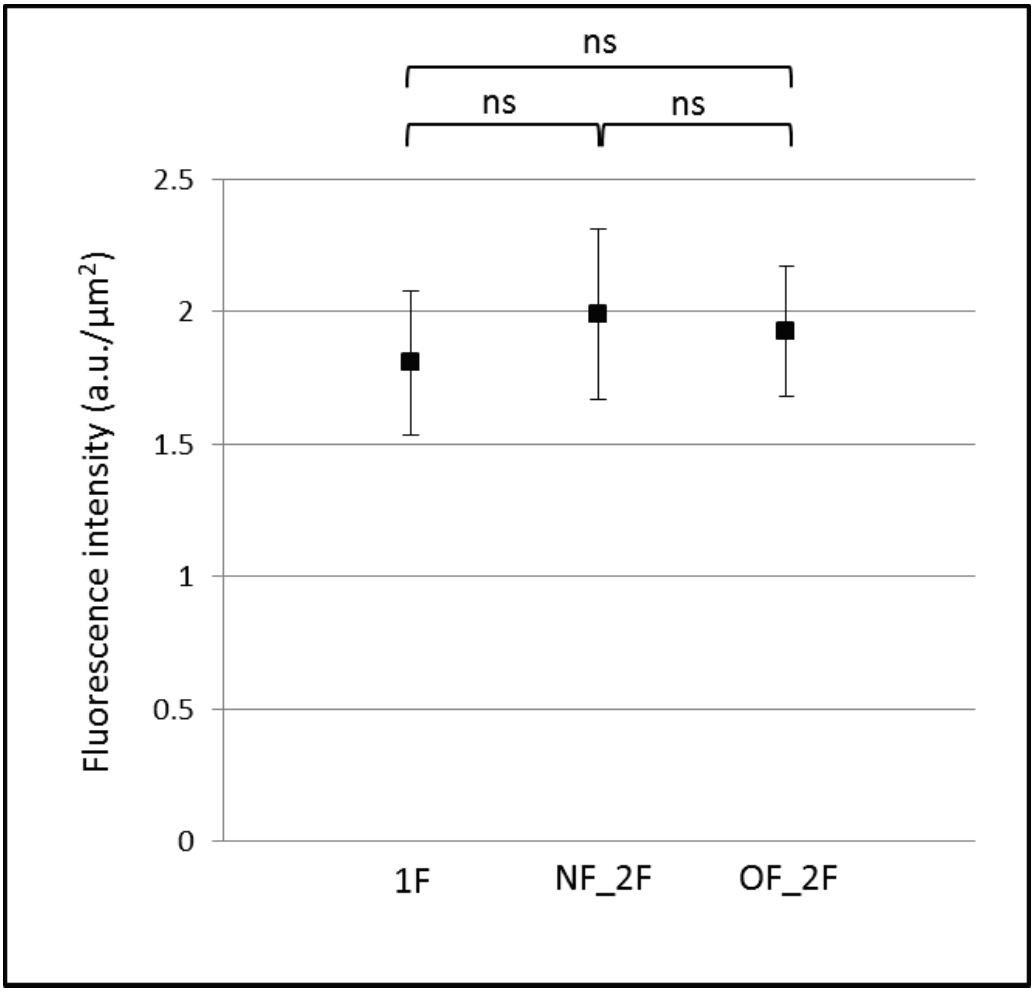


Figure 87: YFP::TbCen2 levels do not vary between flagella

The signal intensity of YFP::TbCen2 was quantified from the single flagellum of 1F cells (n=42), the new flagellum of 2F cells (NF_2F, n= 35) and the old flagellum of 2F cells (OF_2F, n=44). Graph shows mean ± s.d. Error bars indicate standard deviation. The differences in signal intensity (a.u./μm^2) were not statistically different between any of the three of flagellum categories. Unpaired *t* tests between; 1F and NF_2F (P-value = 0.1754), 1F and OF_2F (P-value = 0.2883), NF_2F and OF_2F (P-value = 0.6112).

6.4.4 A link between YFP::TbCen2 levels at the basal bodies and the cell cycle

Once it had been established that there was a difference in YFP::TbCen2 fluorescence intensity between the PBBs and the MBB, an investigation began to find a mechanism for how this may occur.

To investigate any correlation between YFP::TbCen2 fluorescence intensity and basal body maturity/age in *T. brucei*, the length of the new flagellum was measured to allow arrangement of cells in cell cycle order (Figure 88). See section '2.6.4 Measurement of flagellum length'. The relationship between the length of the NF and the fluorescence intensity of YFP::TbCen2 at the MBB associated with the NF (Figure 88; A, orange) had a positive correlation (R = 0.349), which was not statistically significant (P-value = 0.02362). The relationship between the length of the NF and the fluorescence intensity of YFP::TbCen2 at the old MBB associated with the old flagellum (Figure 88; B, red) had a positive correlation (R = 0.453), which was statistically significant (P-value = 0.00257*). The relationship between the length of the NF and the fluorescence intensity of YFP::TbCen2 at the PBB associated with the NF (Figure 88; C, blue) had a weak positive correlation (R = 0.043), which was not statistically significant (P-value = 0.78444). The relationship between the length of the NF and the fluorescence intensity of YFP::TbCen2 at the PBB associated with the OF (Figure 88; D, green) had a positive correlation (R = 0.2315), which was not statistically significant (P-value = 0.1401).

In summary, the levels of YFP::TbCen2 varied between pro-basal bodies and mature basal bodies. The level of YFP::TbCen2 detected at pro-basal bodies was consistently more intense than that detected at mature basal bodies. The level of YFP::TbCen2 measured at the flagellum was also consistently strong and similar to the level at the pro-basal bodies.
A positive correlation was found between the length of the new flagellum and the YFP::TbCen2 signal at the MBB associated with the new flagellum (the newest MBB) and the PBB associated with the old flagellum, although those positive correlations were not statistically significant. A statistically significant positive correlation was found between the length of the new flagellum and the fluorescence intensity of YFP::TbCen2 at the MBB associated with the old flagellum, the oldest MBB. Therefore there is a link between basal body age and YFP::TbCen2 fluorescence intensity.

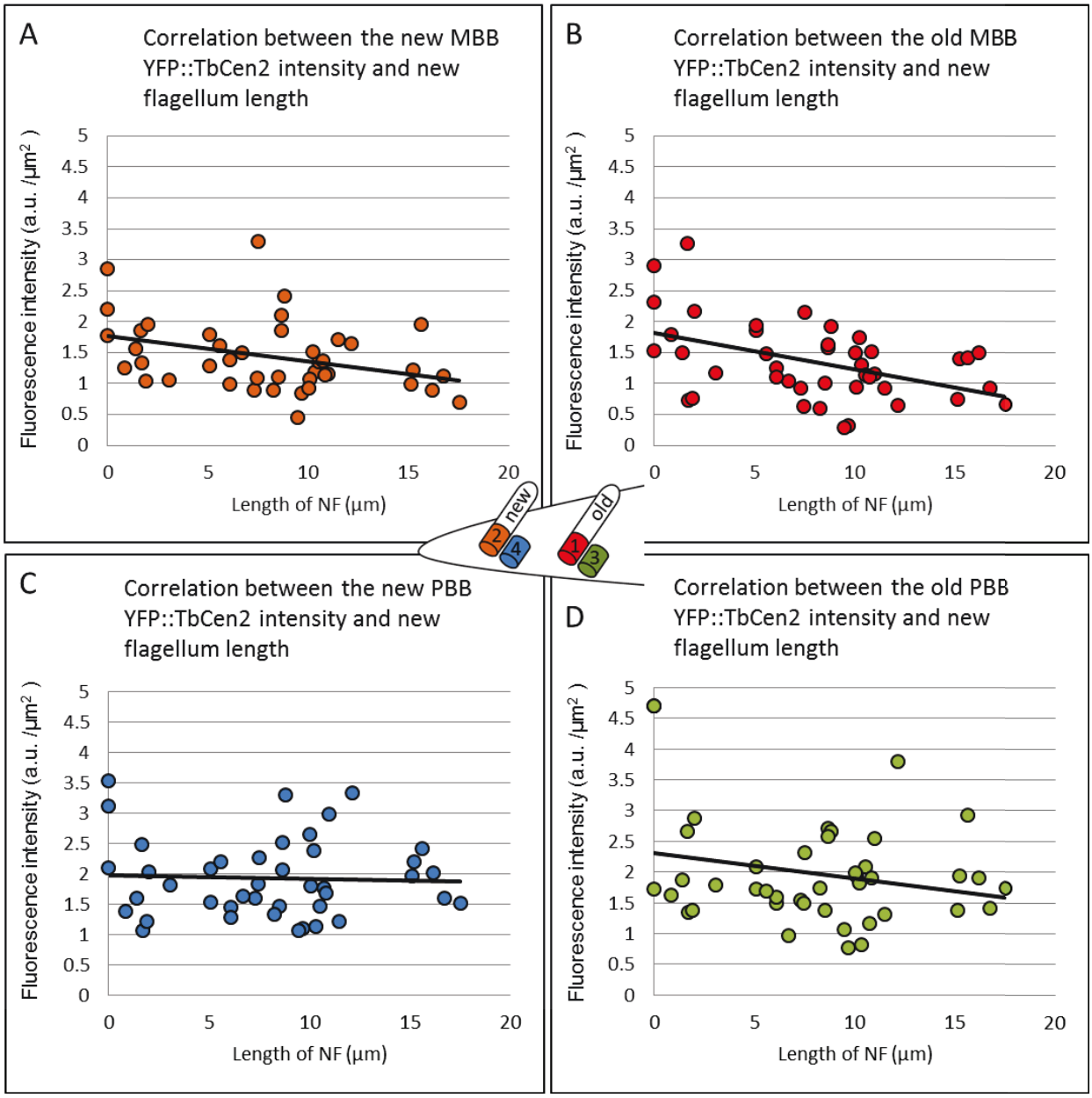


Figure 88: Correlation of YFP::TbCen2 signal at the basal bodies with new flagellum length

The length of the new flagellum was measured and plotted against the intensity of YFP::TbCen2 fluorescence at; (A) the newly matured MBB (orange) associated with the new flagellum (NF), (B) The old MBB (red) associated with the old flagellum, (C) the PBB (blue) associated with the new flagellum (NF), (D) The PBB (green) associated with the old flagellum. Flagellum length was measured using NeuronJ (Meijering *et al.*, 2004) (n= 42). A trendline is included on the graph (black).

6.4.5 Knockdown of TbCen2 affects procyclic cell growth

The assess the function of TbCen2 an inducible p2T7-177 vector against TbCen2 was transfected into the YFP::TbCen2 cell line to create a stable YFP::TbCen2/p2T7_TbCen2 cell line. To investigate any effect of TbCen2 knockdown on the cell cycle, the DNA containing organelles were quantified at various stages through the TbCen2 RNAi induction (Figure 89; A). In the non-induced population (0 hours) of TbCen2 3% of cells had an abnormal configuration of kinetoplasts (K) and nuclei (N). These abnormal cell types were inclusive of 1K2N (Figure 89; A, purple), 1K0N or zoid cells (Figure 89; A, turquoise), multinucleated cells (Figure 89; A, orange) and 'other' (Figure 89; A, skyblue). The 'other' (skyblue) category contained abnormal cell types that were not prominent enough to have a distinct category.

After 48 hours of TbCen2 knockdown, the abnormal cell types had increased from 3.04% to 27.7% of the cell population. By 96 hours the abnormal cell types made up the majority of the population at 71.5%. The increase in DNA-containing organelle copy numbers was reflected in a decrease of observed normal cells. In the non-induced population (0 hours) 79.2% of cells are 1K1N but by 96 hours post induction this had dropped to 18.9% of the observed population. The segment (orange) of the pie chart that represents multinucleated (MN) cells is highlighted by a black circle. In the non-induced population the MN segment is small, representing 0.47% of the population. However, by 96 hours of TbCen2 knockdown, the majority of cells in the population are MN with 53.4% (Figure 89; A, orange). (See Figure 92 and Figure 91 for examples).

Knockdown of YFP::TbCen2 was confirmed by electrophoresis and western blotting of whole cell lysates (Figure 89; B), see section '2.5 Methods: Protein analysis'. The L8C4 probed blot had a band in each lane, which confirmed that cell protein was present. Left to right (Figure 89; B), the first lane contained parental SmOx P9 cells (Poon *et al.*, 2012), which do not contain a YFP tag for any protein. This was confirmed as there was no band on the anti-GFP probed blot. The second lane contained protein from the YFP::TbCen2/p2T7_TbCen2 cell line where the RNAi had not been induced against TbCen2. The anti-GFP antibody detected a band corresponding to the endogenous fusion protein, YFP::TbCen2. The third and final lane, contained protein from YFP::TbCen2/p2T7_TbCen2 cells that had been treated with doxycycline for 48 hours. This

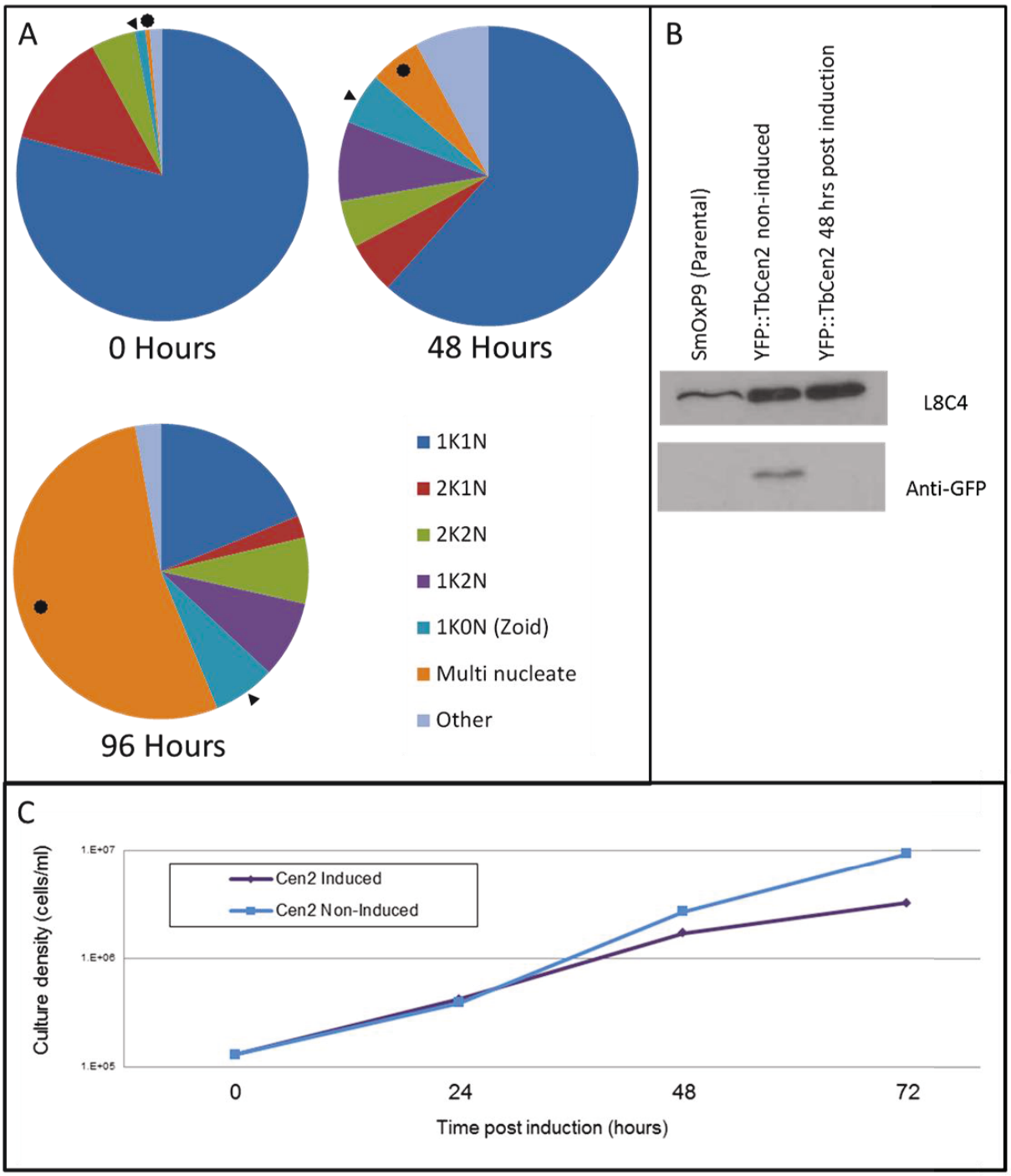


Figure 89: RNAi ablation of TbCen2 affects cell growth

(A) Quantification of DNA containing organelles for non-induced cells, cells that have been induced for 48 hours and cells that have been induced for 96 hours (n = 428, 538 and 249 respectively). Key for pie charts is shown on the right. The earliest abnormal phenotypes that emerge are 1K0N (zoid; turquoise) and 1K2N cells. At later stages of induction the cell population contained many multinucleated cells (orange). (B) Effective RNAi mediated knockdown of TbCen2 evaluated by western blotting against the YFP::TbCen2 fusion protein using an anti-GFP antibody. The L8C4 antibody was used as a loading control. (C) Continuous growth analysis of non-induced (blue) and induced (purple) populations of YFP::TbCen2/p2T7_TbCen2.

lane did not have a band detected by the anti-GFP. This meant that the RNAi targeted against TbCen2 produced effective knockdown of TbCen2 by 48 hours post induction.

Cell growth was assessed for cell populations post induction of RNAi and compared to the growth of the non-induced cell population. A continuous growth curve of YFP::TbCen2/p2T7_TbCen2 cell line non-induced population (Figure 89; C, blue line) and induced population (Figure 89; C, purple line), revealed that knockdown of TbCen2 affected the growth of procyclic form cells but was not lethal by 72 hours.

6.4.6 Knockdown of TbCen2 does not affect basal body duplication

In wildtype *T. brucei* cells, one pair of basal bodies is associated with each flagellum. This can be verified by immunolabelling with an anti-basal body antibody such as BBA4 (Woods *et al.*, 1989b). In a 1K1N cell that has one flagellum, the cell has one pair of basal bodies, which is seen in non-induced p2T7_TbCen2 cells (Figure 90; A). A 2K2N cell has two flagella and therefore has two sets of basal bodies (Figure 90; B). After p2T7_TbCen2 was induced the number of flagella and basal bodies per cell was observed.

When p2T7_TbCen2 was induced, cells were detergent extracted, fixed and immunolabelled with the basal body antibody, BBA4 (Woods *et al.*, 1989b). See section '2.6.2 Immunofluorescence' for details. BBA4 immunolabelling revealed that cells had a normal ratio of basal bodies to flagella (2BB:1F). After TbCen2 knockdown, even when cells had an abnormal quantity of kinetoplasts (K), nuclei (N) or flagella (F); each flagellum still had one pair of basal bodies. When basal body immunolabelling by BBA4 was quantified, 90.7% of cells (n=59) had the correct number of BBA4 labelled basal bodies by 96 hours post induction of p2T7_TbCen2. When TbCen2 was depleted, after 72 hours the cells still had a BB:F of 2:1 (Figure 90; C and D). As reported in section 6.4.3, knockdown of TbCen2 caused produced cells with an abnormal number of DNA containing organelles. Figure 90 (C) shows a cell that is 1K2N, which is irregular. However, the cell has a single flagellum and associated with it, a single pair of basal bodies (BBA4 channel; Inset). Another example from 72 hours post induction is shown (Figure 90; D), which is 2K4N. The cell has two flagella, each with a pair of basal bodies associated with it (BBA4 channel; inset). By 96 hours post induction of p2T7_TbCen2 there is still one pair of basal

bodies per flagellum (Figure 90; E and F). A cell with 1 kinetoplast but two flagella (Figure 90; E) is shown. The new flagellum (NF) is detached from the cell body but the old flagellum (OF) is attached as normal. The cell is 1K2N. It is abnormal that this cell only has one kinetoplast because the NF is quite long and therefore the kinetoplasts have usually separated at this point in the cell cycle. The mutant cell has two flagella and each flagellum correctly has one associated pair of basal bodies (BBA4 channel; inset). A cell that is undergoing cytokinesis (Figure 90; F) has a mis-positioned nucleus, which is more posterior to the flagellum. Each flagellum has a pair of basal bodies associated with it (BBA4 channel; inset).

Therefore, cells that contain an abnormal number of DNA-containing organelles maintain the correct number of basal bodies in relation to the number of flagella. This suggests basal body duplication is not affected by TbCen2 knockdown.

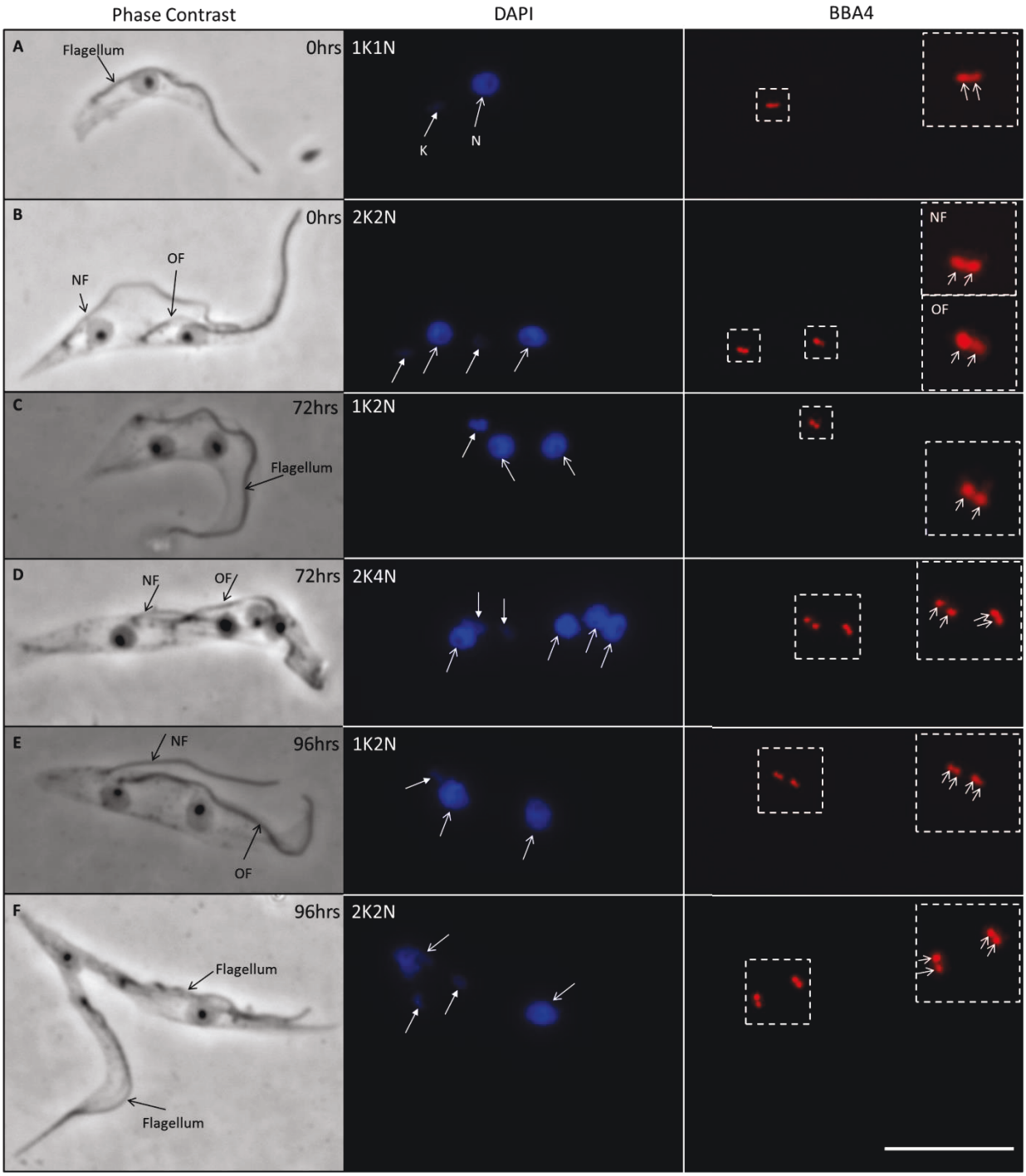


Figure 90: Knockdown of TbCen2 does not affect basal body duplication

Cells were detergent extracted, fixed, labelled with DAPI and immunolabelled with BBA4. (A) a non-induced cell (0 hours) with one kinetoplast (K; solid white arrowhead) and one nucleus (N; open white arrowhead). (B) a non-induced 2K2N cell (0 hours). (C) 72 hours post induction of p2T7_TbCen2 cell that is 1K2N. (D) 72 hours post induction of p2T7_TbCen2 cell that has 2K4N. (E) 96 hours post induction of p2T7_TbCen2 cell that is 1K2N and the NF is detached. (F) 96 hours post induction of p2T7_TbCen2 cell that has two kinetoplasts and two nuclei that are mispositioned within the cell. BBA4 insets are enlarged 100% (A-C) and 50% (D-F). NF; new flagellum, OF; old flagellum, BB; basal bodies. Scale bars = 10µm

6.4.7 Knockdown of TbCen2 disrupts flagellum attachment but does not affect flagellum assembly

In the non-induced population of p2T7_TbCen2 cells, the cells had attached flagella (Figure 91; A, red arrows). By 96 hours post induction, cells with detached flagella were observed in the slide preparations (Figure 91; B and D, orange arrows). Free flagella were also observed (Figure 91; C, orange arrows). By 96 hours of TbCen2 ablation 66.5% of cells had fully or partially detached flagella (n=221).

In conclusion, flagella are assembled in the absence of TbCen2 but ablation of TbCen2 compromises flagellum attachment.

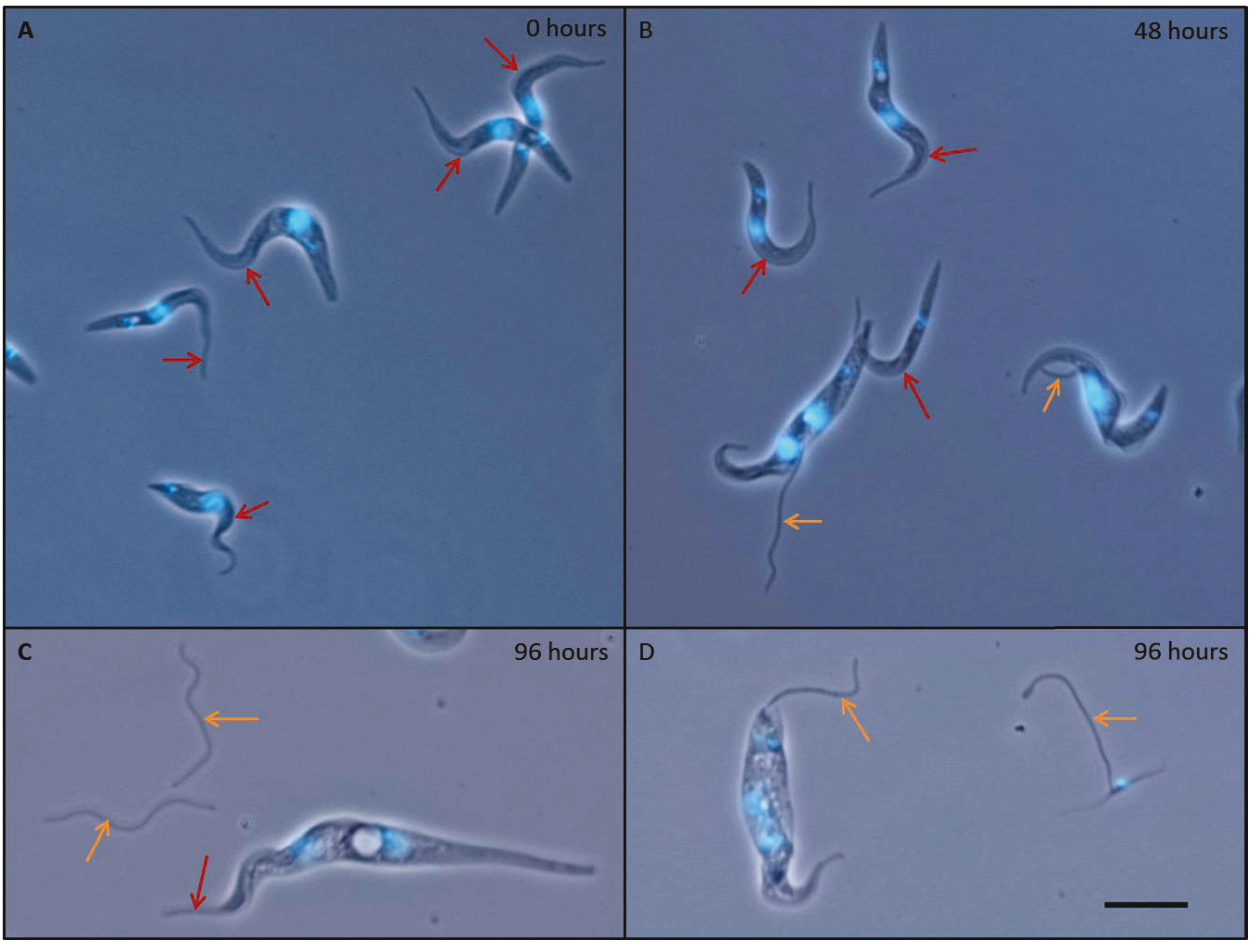


Figure 91: Knockdown of TbCen2 leads to detached flagella

Overview for non-induced cells (A, 0 hours) and induced cells (B, 48 hours post induction) and (C and D, 96 hours post induction). Cells with an attached flagellum (red arrow) and cells that have a detached flagellum (orange arrow) are shown. Phase contrast and DAPI channels are merged in all panels. Scale bar = 10µm

6.4.8 Knockdown of TbCen2 affects cell morphology

After TbCen2 knockdown, abnormal cell shapes were observed in suspension culture. To qualify and quantify the variation of cell morphologies seen, cells were fixed and labelled with DAPI. Due to the constrictive nature of the subpellicular corset, when the regulation of cell body shape is disrupted, there are a limited number of irregular forms that can be observed. Using previous morphology classifications of *T. brucei* (Allen *et al.*, 2003), categories were devised for quantification of abnormal cell shapes seen when TbCen2 was knocked down (Figure 92). The categories used were; normal, teardrop, fat (multi-nucleate) and extended posterior end. Cells analysed were from non-induced (0 hours, n= 129) and induced (96 hours, n=161) populations. Within the non-induced population abnormal cell morphologies were observed although they were present in small quantities at 10.1% of cells observed. By 96 hours post induction, 93.2% of the population was made up of abnormally shaped cells. Example cells from the 96 hours post induction population (Figure 92) showed normally formed cells (Figure 92; A), teardrop shaped (Figure 92; B), fat/multi-nucleate (Figure 92; C) and a cell with an extended posterior end (Figure 92; D). These example cells (Figure 92; A-D) demonstrate the importance of analysing cells by morphology as well as by DNA-containing organelle copy number. Figure 92 (B) is an abnormally shaped 'teardrop' cell but has two kinetoplasts (K, closed white arrow head) and two nuclei (N, open white arrow head), by only counting the DNA-containing organelles within that particular cell, it would be classed as a normal configuration. The same applied to the example cell shown for the extended posterior end category (Figure 92; D). The posterior end of the cell body is 10-fold longer than it should be but the cell contains a single kinetoplast and a single nucleus, making it a 1K1N cell.

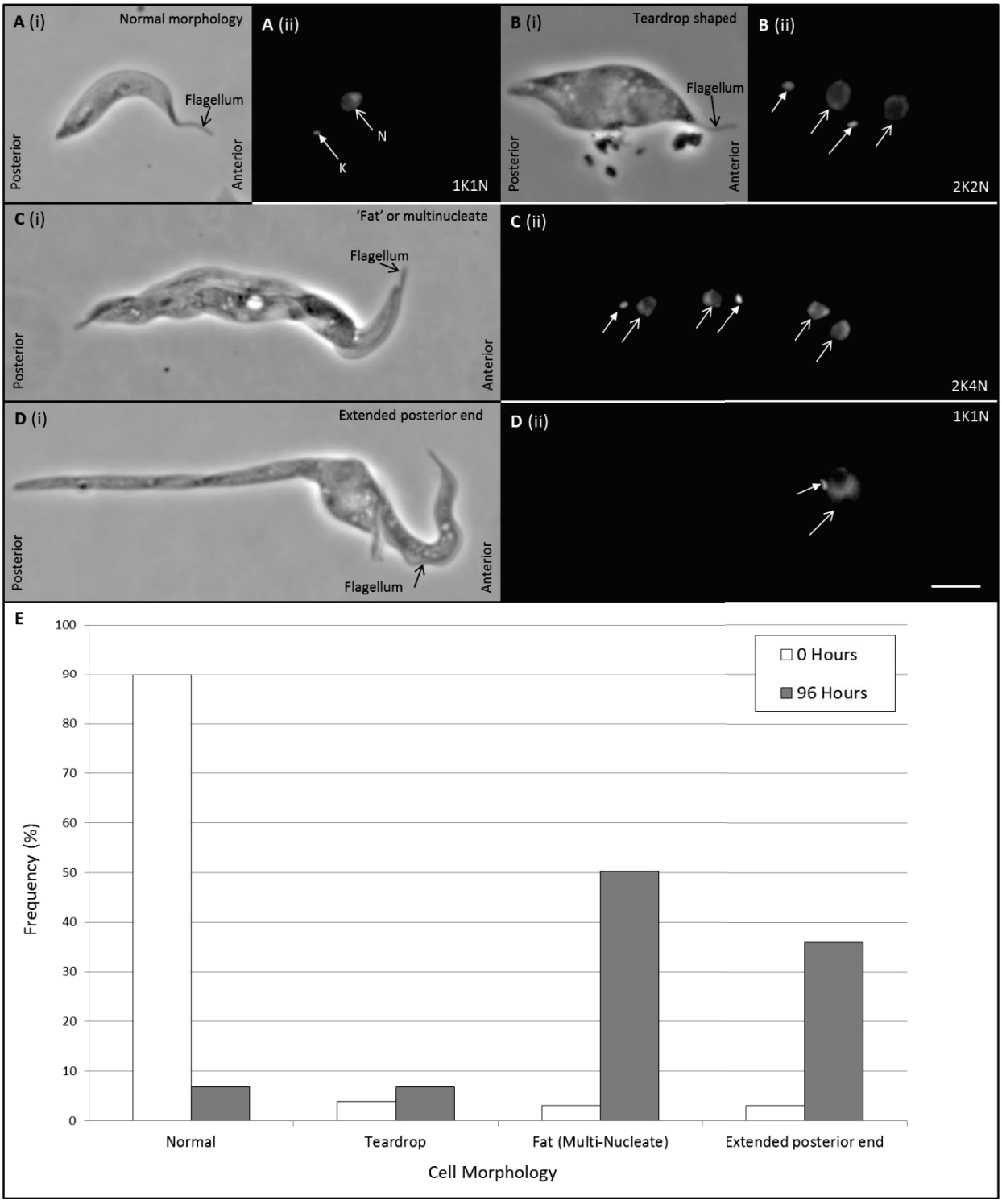


Figure 92: Knockdown of TbCen2 affects cell morphology

Example cells for each morphology category with a phase contrast view (i) and DNA-containing organelles stained with DAPI (ii). (A) A normal shaped cell present in the culture 96 hours post induction of RNAi targeted against TbCen2. The cell still has a single kinetoplast (K; closed white arrow head) and a single nucleus (N; open white arrow head). (B) A teardrop shaped cell with a normal configuration of kinetoplasts and nuclei. (C) A cell that is classed as 'fat' or 'multi-nucleate' has 2 kinetoplasts and 4 nuclei contained within the large but well-proportioned cell body. (D) A cell body that has an elongated posterior end. The kinetoplast and nuclei are too close together. Scale bar = 5µm. (E) Quantification of each cell type for non-induced cells (n=129, white, 0 hours) and induced cells (n= 161, grey, 96 hours).

6.4.9 Characterisation of normal cell morphology by SBF-SEM

This work is the first volumetric EM study of whole cell procyclic form *T. brucei.* Characterisation of the wildtype procyclic cell will allow meaningful comparison of mutant cell lines. This type of whole cell volumetric reconstructions will allow phenotype analysis that has been restricted by conventional microscopy techniques. Procyclic cells at different stages of the cell cycle were isolated from the data and segmented (Figure 93). See section 2.7.5 SBF-SEM data analysis. The cells were chosen (i- vi) based on recognisable, landmark morphological changes in the procyclic cell cycle. There are other methods to identify procyclic cell cycle stages based on subtle ultrastructural changes to organelles or biochemical changes but that was beyond the scope of the SBF-SEM data set.

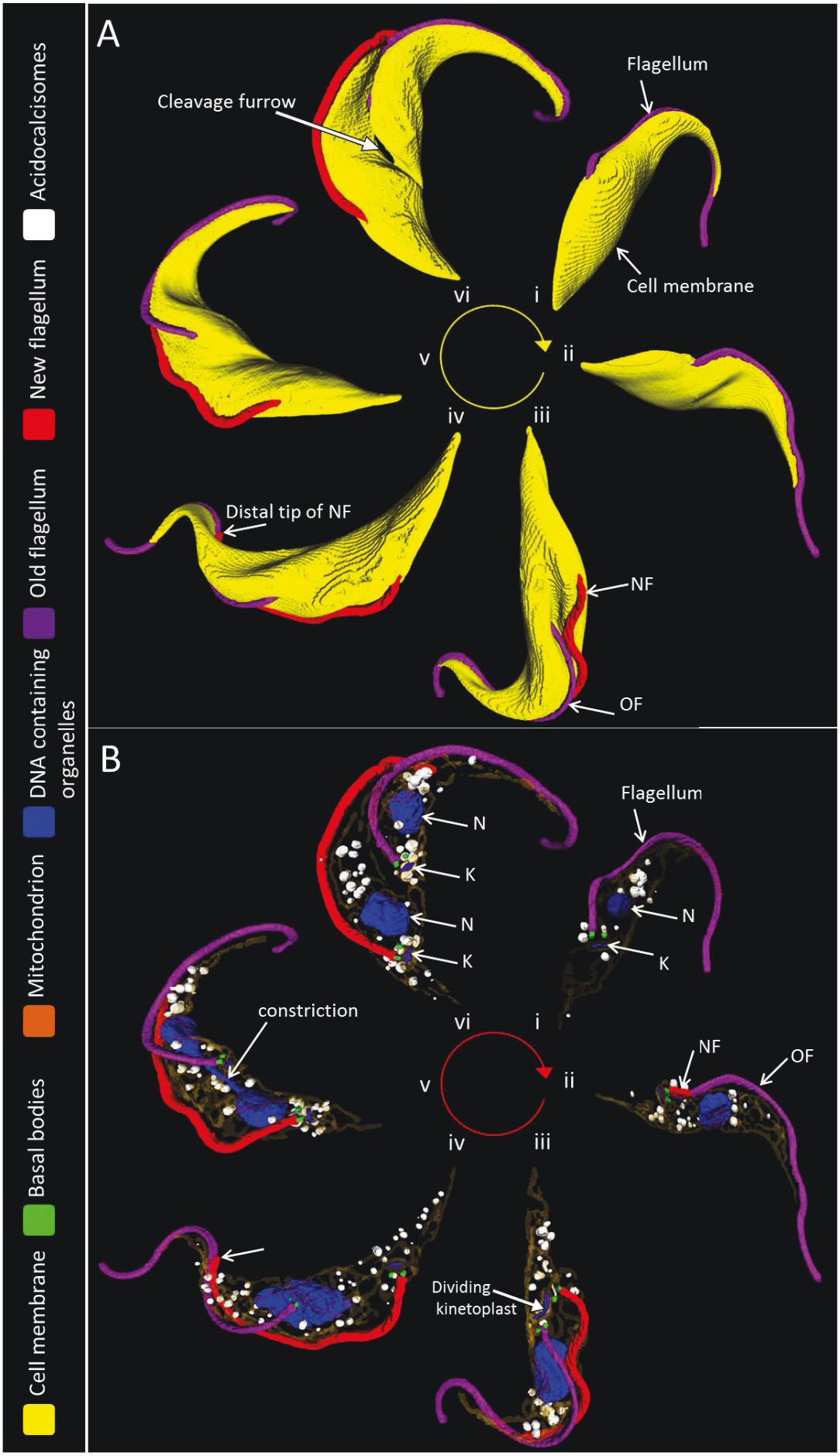
Procyclic cells are arranged clockwise through the cell cycle (Figure 93; A) with the posterior end of the cell pointing towards the middle of the circle, a colour key is included on the left. The new flagellum (red) is positioned nearer the posterior end of the cell compared to the old flagellum (purple) (Robinson *et al.*, 1995) (Figure 93; A, iii-vi). The last cell cycle stage portrayed in this work (Figure 93; A, vi) is a cell prior to abscission of the two daughter cells. The cleavage furrow is evident (Figure 93; A, vi, arrow). Panel B (Figure 93) shows the same cells as panel A (Figure 93), except the cell membrane (yellow) is not shown to allow visualisation of the internal organelles. By conventional SEM cells (i) and (ii) would have appeared to be the same cell cycle stage, as when viewed externally, each cell had a single flagellum (Figure 93; A). However, when the cells were reconstructed it became apparent that cell (ii) had a short new flagellum (Figure 93; B, ii, red), which had not yet exited the flagellar pocket. This demonstrates the power of whole cell SBF-SEM data.

A G1 stage cell (Figure 93; i), has one kinetoplast and one nucleus (DNA containing organelles, blue) and therefore is also referred to as a 1K1N cell. A video of the SBF-SEM data and 1K1N cell model is included as Appendix 16. There is a single flagellum (purple) with a single pair of basal bodies (green). The cell has a single, elaborate mitochondrion (dark orange/brown), which occupies space along the length of the cell body. Most of the organelles within *T. brucei* are single copy; exceptions include ribosomes, glycosomes and acidocalcisomes. In this dataset the acidocalcisomes were modelled (Figure 93; B, white)

to allow perspective of multi-copy organelle distribution. From the resolution of this data set, the first landmark change in the cell cycle is the extension of a new flagellum (Figure 93; B, ii, red), which extends on the left hand side of the old flagellum (purple) as the cell is viewed posterior to anterior. The distal tip of the new flagellum is attached to the old flagellum (Briggs *et al.*, 2004b; Moreira-Leite *et al.*, 2001). As the new flagellum extends, the kinetoplast divides (Figure 93; B, iii) to produce a 2K1N cell (Figure 93; B, iv) with two kinetoplasts and one nucleus, which is pre-mitotic. As the nucleus divides there is a constriction around the mitotic spindle (Ogbadoyi *et al.*, 2000) forming a bridge between the daughter nuclei (Figure 93; B, v). When karyokinesis is complete the cell has two nuclei and is also known as 2K2N (Figure 93; B, vi). Following this stage the cell will complete cytokinesis and produce two 1K1N daughter cells. A video of the SBF-SEM data and rendered model of an example cell with a dividing kinetoplast is included as Appendix 17.

Figure 93: Analysis of procyclic form *T. brucei* by SBF-SEM

Reconstructed surfaces from segmented SBF-SEM data of example cells displaying morphological landmarks used to determine cell cycle stage in *T. brucei* procyclic form cells. (I) A cell in G1. (II) The new flagellum (NF, red) begins to assemble. (III) The basal bodies (green) have duplicated and the kinetoplast is dividing, the cell still has a single nucleus. (IV) The cell has two distinct kinetoplasts and a single nucleus that is elongated prior to karyokinesis. (V) Nuclear mitosis. (VI) A cell prior to abscission that has two kinetoplasts and two nuclei undergoing cytokinesis to produce two daughter cells with correct organelle copy number.



(Figure 93; figure legend on previous page)

6.4.10 Characterisation of mutant cell morphology by SBF-SEM

Abnormal cell morphology was observed and quantified by light microscopy following knockdown of TbCen2 (Figure 92). To allow characterisation of abnormal cell morphology in three dimensions, p2T7_TbCen2 whole cells were fixed after 72 hours of doxycycline treatment that induced RNAi targeted against TbCen2. The first identifiable stage in the procyclic cell cycle is a 1K1N cell. Figure 94 shows a wildtype 1K1N cell (A – same cell as Figure 93, i) that has normal cell body morphology and a single flagellum. After TbCen2 knockdown, cells were seen with an extended posterior end (Figure 94; B), which were also 1K1N cells with a single flagellum.

The posterior end of the cell was defined as the distance from the mature basal body (MBB, green) to the posterior end of the cell body membrane. The posterior end of the wildtype cell and the mutant cell (Figure 94; C and D, respectively), were measured. The posterior end of the wildtype procyclic cell was 3.91µm (Figure 94; C), only the MBB is shown on the model. The posterior end of the TbCen2 knockdown cell was 8.54 µm (Figure 94; D), only the MBB is shown on the model.

Knockdown of TbCen2 causes flagellum detachment (Figure 91) and this was also seen in the SBF-SEM data. Figure 95 shows two views of the same cell (A and B) to demonstrate that the new flagellum (NF, red) has assembled although it is detached and the distal tip of the NF did not remain attached to the OF (Figure 95; A). The NF shared the same flagellar pocket (FP) as the old flagellum (OF, purple). Nuclear division has occurred and therefore in the normal cell cycle, the two flagella would exit from two separate flagellar pockets (for example see Figure 93; vi). Although the NF (red) has rotated to be on the left of the OF (Figure 95; B) so loss of TbCen2 does not affect basal body rotation.

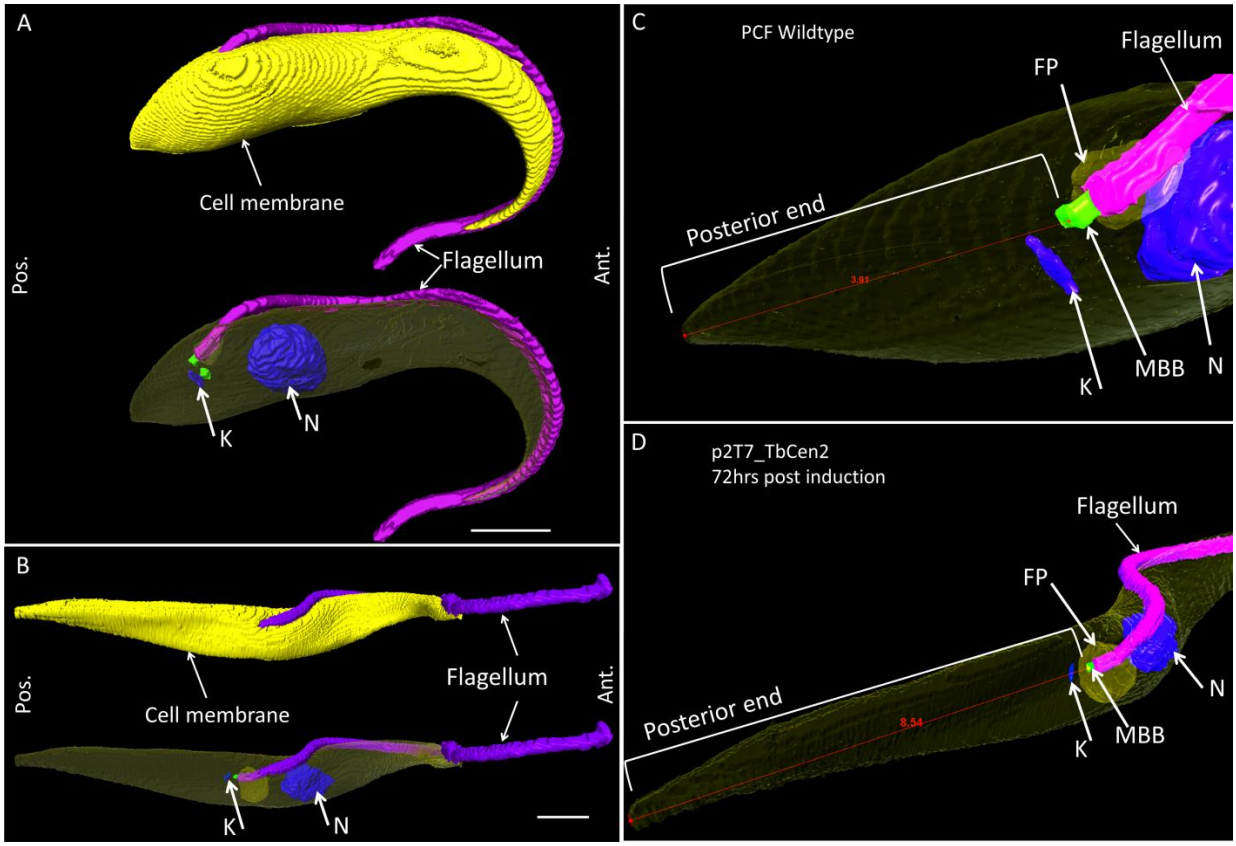


Figure 94: SBF-SEM investigation of abnormal 1K1N cell morphology as a result of TbCen2 knockdown

Segmentation and rendering of SBF-SEM data. (A) A wildtype 1K1N cell (same as Figure 93; i). (B) p2T7_TbCen2 (72 hours post induction) 1K1N cell with an elongated posterior end. (C) Rotated close up of posterior end of cell in (A), posterior end is 3.91µm long. (D) Rotated close up of posterior end of cell in (B), posterior end is 8.54µm in length. Scale bars 2µm (A and B).

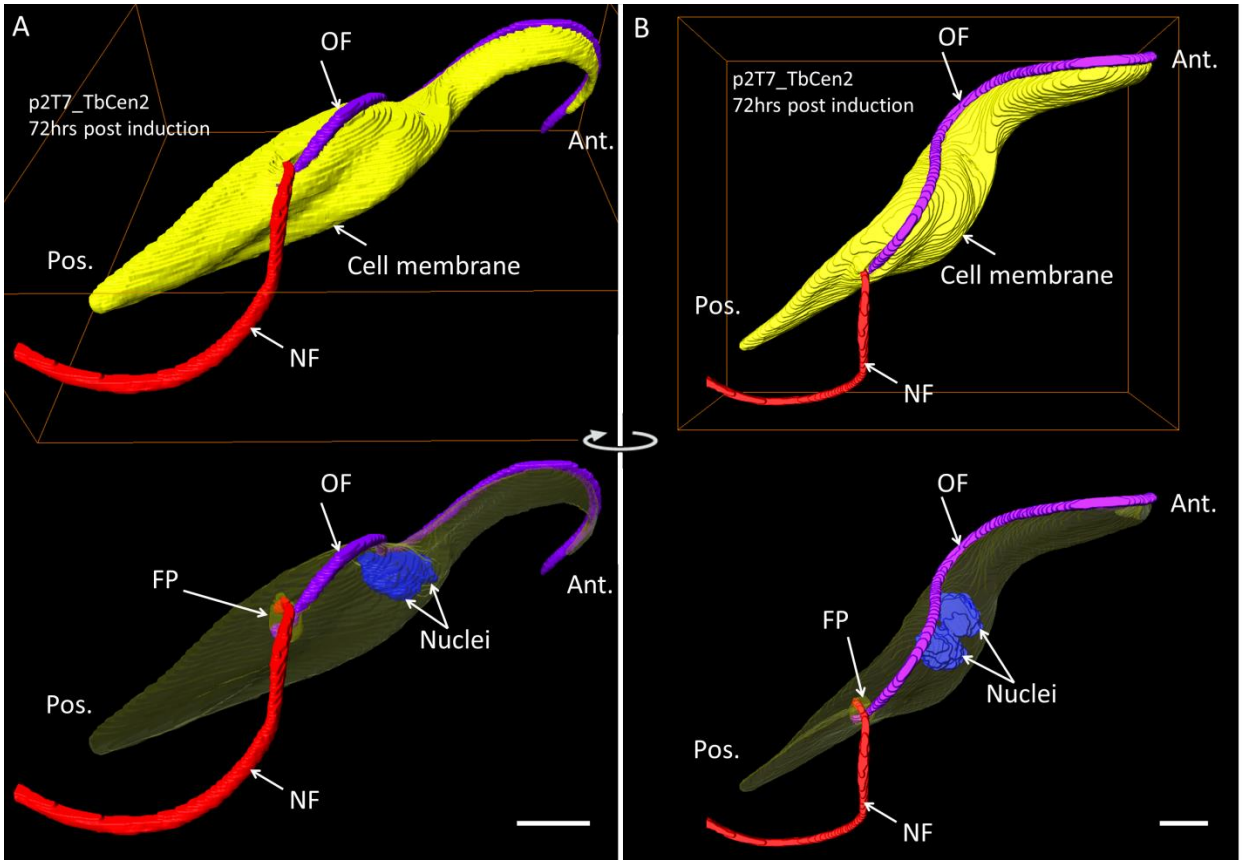


Figure 95: SBF-SEM investigation of abnormal cell morphology as a result of TbCen2 knockdown

Segmentation and rendering of SBF-SEM data collected 72 hours post ablation of TbCen2. (A) A large cell with a detached new flagellum (NF, red), an attached old flagellum (OF; purple) and abnormal positioning of the two nuclei (N, blue). (B) The same cell as (A) from a different angle to show a clear view of the flagellar pocket (FP) with the NF on the left of the OF. Scale bars 2µm.

Figure 96 shows the rendered models of a wildtype PCF 2K2N cell (A) and an abnormal 2K2N cell from the induced population of p2T7_TbCen2 (B). Although both cells possess two kinetoplasts (K) and two nuclei (N), there is abnormal organelle positioning. Specifically, the kinetoplasts and the nuclei are further apart in the p2T7_TbCen2 cell (Figure 96; F) compared to the wildtype (Figure 96; C). In the wildtype cell, the distance between the kinetoplast associated with the NF and the nucleus was 1.34µm (C1) and the distance between the kinetoplast associated with the OF and the nucleus was 1.50µm (C2). These distances are increased after knockdown of TbCen2 (Figure 96; F) where the distance between the kinetoplast associated with the NF and the nucleus was 4.17µm (F1) and the distance between the kinetoplast associated with the OF and the nucleus was 5.99µm (F2). Despite the increased gap between the kinetoplast and the nucleus, the distance between the kinetoplast and the mature basal body (MBB, green) (K←→MBB) remained constant (Figure 96; insets, C and F). Only the MBB is displayed to avoid confusion with PBB. In the wildtype cell, the distance between the kinetoplast and the MBB associated with the NF (red) was 0.35µm and the distance between the kinetoplast and the MBB associated with the OF (purple) was 0.33µm (Figure 96; insets, C). After TbCen2 knockdown, the distance between the kinetoplast and the MBB associated with the NF (red) was 0.37µm and the distance between the kinetoplast and the MBB associated with the OF (purple) was 0.38µm (Figure 96; insets, F). This meant that the tripartite attachment complex (Ogbadoyi *et al.*, 2003) between the MBB and the kinetoplast was unaffected by the absence of TbCen2. In both cells the NF is positioned correctly behind the OF but knockdown of TbCen2 caused detachment of the NF (Figure 96; D and E), the OF was attached.

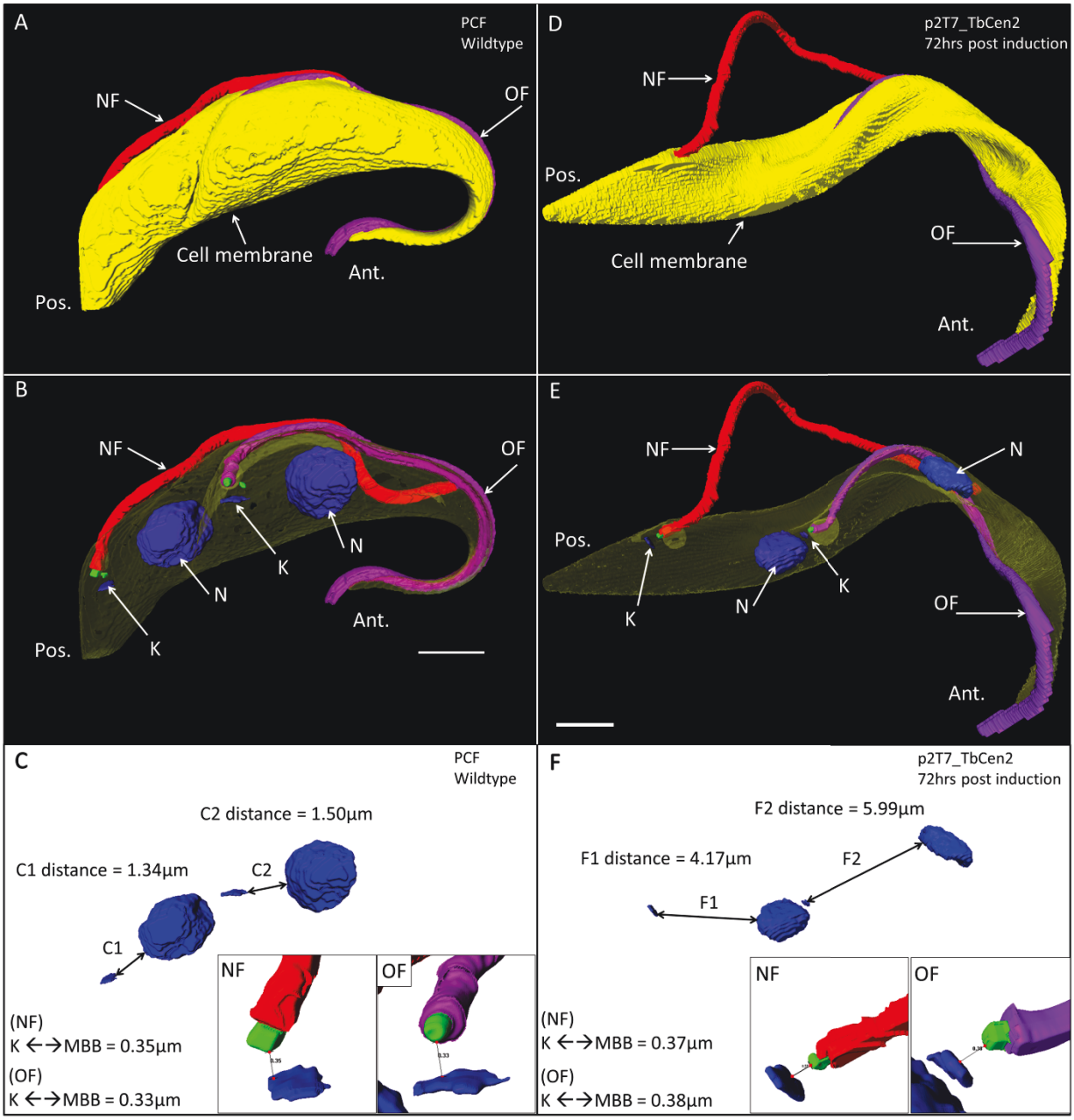


Figure 96: SBF-SEM investigation of abnormal 2K2N cell morphology as a result of TbCen2 knockdown Segmentation and rendering of SBF-SEM data. (A-C) A wildtype PCF 2K2N cell. (D-F) p2T7_TbCen2 (72 hours post induction) 2K2N cell with a detached new flagellum (NF; red) and attached old flagellum (OF). (A and D) external views of solid cell membrane and flagella. (B and E) transparent cell membrane allows view of internal organelles. (C and F) isolated view of segmented kinetoplasts (K) and nuclei (N) with distances measured (double ended arrow). Insets (C and F), MBB ←→ K distances for NF (red) and OF (purple). Scale bars 2µm.

6.4.11 Ablation of TbCen2 causes an ultrastructural defect in the axoneme

After RNAi induction targeted against TbCen2, it was noted that cells still built a flagellum, despite it being detached in some cells (Figure 91). The canonical eukaryotic microtubule axoneme consists of 9 outer doublet microtubules and 2 microtubules in the centre, known as the central pair of microtubules – or C1 and C2. Examination of cytoskeletal axonemes by TEM, as described in section '2.7.2 Transmission electron microscopy', revealed that knockdown of TbCen2 lead to the absence of a ventral portion of the C2 microtubule. Figure 97 demonstrates that when TbCen2 is depleted, after 48 hours, a ventral section of the C2 microtubule is absent (Figure 97; B and D). It was not possible to identify which specific protofilaments were missing, the cartoon (Figure 97; B) is just a representation of the partial C2 microtubule. This incomplete C2 protofilament arrangement was not seen in the axoneme of non-induced cells (Figure 97; A and C). The C1 microtubule appeared intact and morphologically normal in mature axonemes with a PFR.

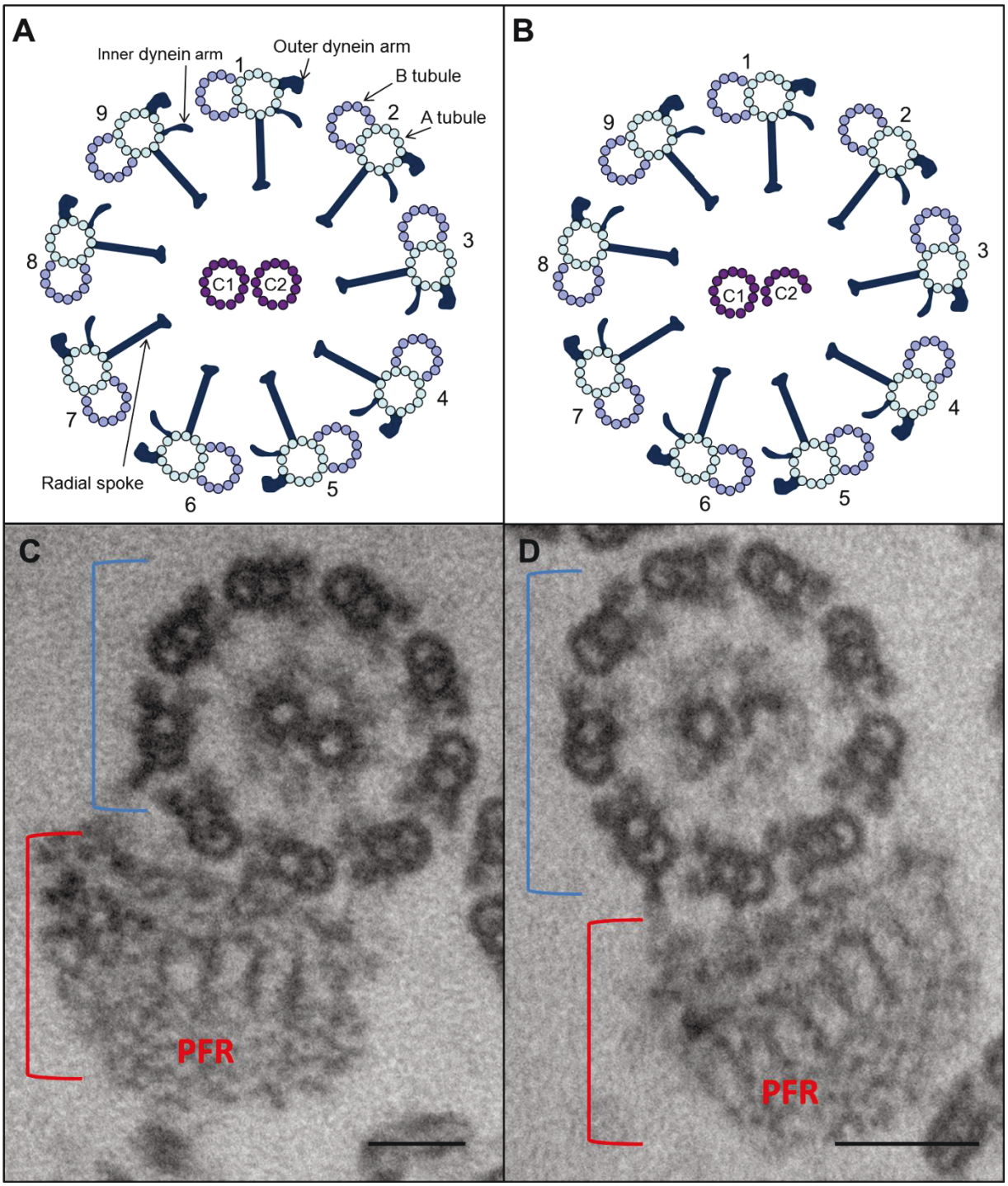


Figure 97: Ablation of TbCen2 causes a defect in the C2 microtubule of the axoneme

(A) A cartoon to show the normal ultrastructure of the axoneme in *T. brucei* (C) A micrograph of the flagellum, showing the PFR (red bracket) and the axoneme (blue bracket) from a non-induced cell that has complete C1 and C2 microtubules. (B) A cartoon to show the aberrant ultrastructure of the axoneme observed in TbCen2 RNAi cells. (D) A micrograph of the flagellum, showing the PFR (red bracket) and the axoneme (blue bracket) from a cell 48 hours post knockdown of TbCen2 where the C1 microtubule is complete but the ventral portion of the C2 microtubule is absent. Cells were detergent extracted, fixed and resin embedded for observation of the cytoskeleton by transmission electron microscopy (TEM). Axonemes (C and D) are not elliptically corrected. View of axonemes is proximal to distal, with the microtubule doublets arranged in a clockwise direction. Scale bar = 100nm

TEM images of axonemes were elliptically corrected, see section '2.7.3 Axoneme analysis', to create a superimposed 9 fold rotation of the outer doublet microtubules and associated structures. This revealed the same pattern for non-induced cells (Figure 98; A) and induced cells (Figure 98; C). This meant knockdown of TbCen2 did not affect the assembly or structure of the outer doublet microtubules, the inner dynein arms, the outer dynein arms or radial spokes. When the elliptically correct axonemes were directly averaged together (no 9 fold rotation), the average Z-projection from axonemes of non-induced cells showed a common, complete C2 microtubule (Figure 98; B, red arrow). It was apparent that the incomplete C2 microtubule was a common phenotype of induced cell axonemes with the Z-projection showing a consistent absence of the ventral section of the C2 microtubule (Figure 98; D, red arrow).

TbCen2 has been localised to the flagellum previously (Ikeda and de Graffenried, 2012), however its function is unknown and there is are ultrastructural localisation data. Although TbCen2 has not been directly localised to the C2 microtubules, the TbCen2 RNAi phenotype implies that TbCen2 is necessary to assemble/stabilise the ventral section of the C2 microtubule.

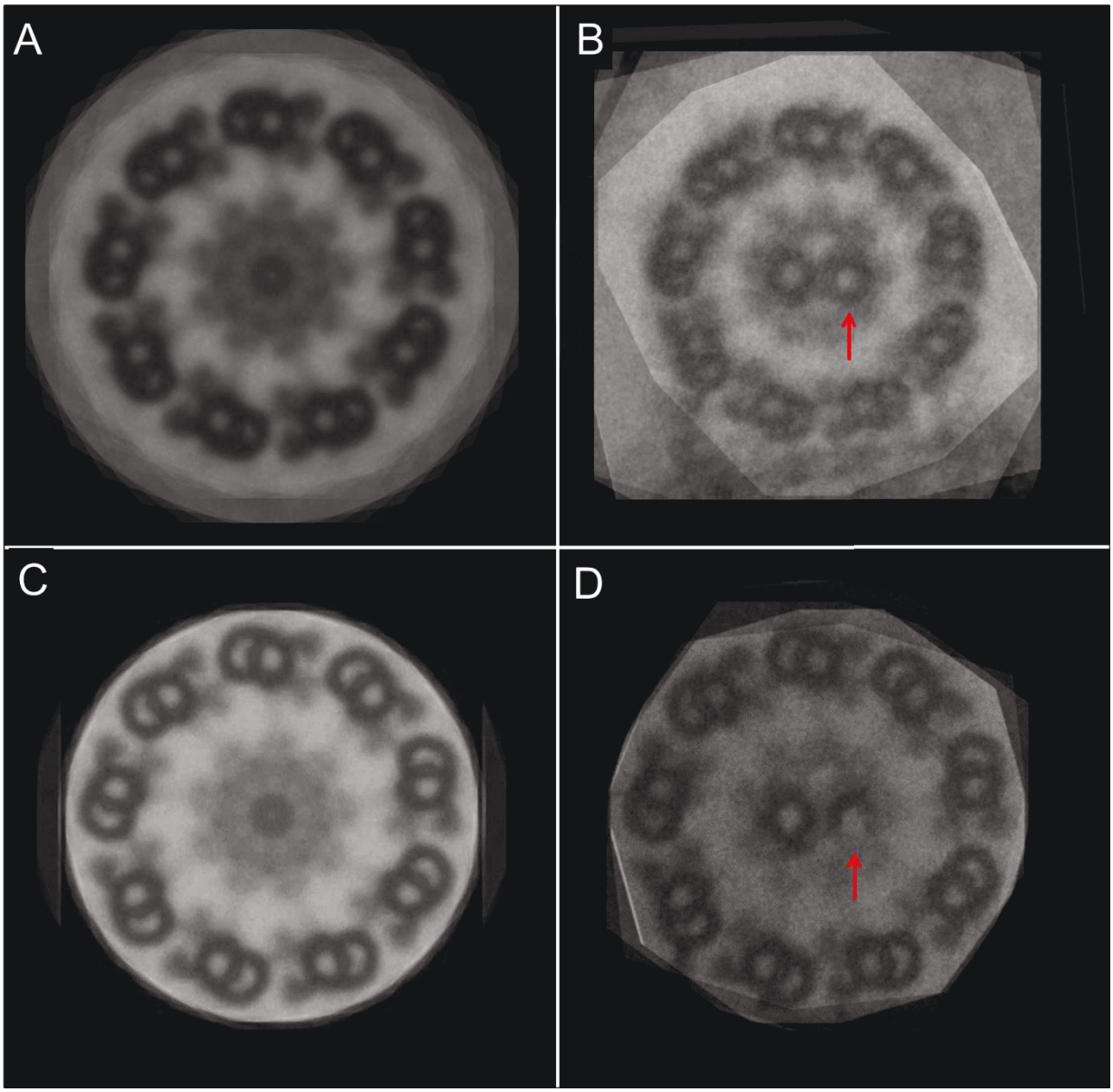


Figure 98: Ablation of TbCen2 does not affect the outer microtubule doublets of the axoneme

(A) Axonemes from non-induced cells, which have been elliptically corrected and rotated (n=5), which showed radial spokes, inner and outer dynein arms. (B) Elliptically corrected axonemes (n=5) were stacked and projected to show the presence of complete C2 microtubule (red arrow). (C) Axonemes from cells 48 hours post knockdown of TbCen2, which have been elliptically corrected and rotated (n=5) were stacked and projected to highlight conserved ultrastructure. (D) Elliptically corrected axonemes (n=5) were stacked and projected to confirm the common absence of the ventral portion of the C2 microtubule (red arrow). View of axonemes is proximal to distal, with the microtubule doublets arranged in a clockwise direction.

6.4.12 Ablation of TbCen2 does not disrupt TbHydin localisation

Due to the partial C2 phenotype caused by knockdown of TbCen2 (Figure 97) the hypothesis was that loss of the ventral portion of the C2 microtubule would therefore displace the projection associated with the C2 microtubule, the 2b projection. There is no published work on the central pair projections in *T. brucei* but in *C. reinhardtii* the 2b projection is known to contain Hydin. Therefore the hypothesis was that loss of the ventral portion of the C2 microtubule would displace TbHydin. To investigate this, a stable cell line was generated that constitutively expressed YFP::TbHydin (Tb927.6.3150) from one endogenous locus using the pPOT method (Dean *et al.*, 2015) to fuse YFP to the N terminus of TbHydin. Fluorescence microscopy showed that YFP::TbHydin localised to the flagellum in whole cells (Figure 99) and detergent extracted cytoskeletons (Figure 100; A and Appendix 18). See section '2.6.1 Light microscopy'.

To investigate if TbCen2 ablation affected YFP::TbHydin localisation, the p2T7-177 vector against TbCen2 was transfected into the YFP::TbHydin cell line to create a stable YFP::TbHydin/p2T7_TbCen2 cell line. The reciprocal cell line (YFP::TbCen2/p2T7_TbHydin) was not generated.

When TbCen2 was knocked down, there was no effect on the localisation of YFP::TbHydin (Figure 100) by light microscopy. After 48 hours of RNAi induced against TbCen2, the YFP::TbHydin was still observed at the flagellum (Figure 100). Knockdown of TbCen2 was not confirmed for experiments performed on the YFP::TbHydin/p2T7_TbCen2 cell line but after the addition of doxycycline to the culture media, the phenotype of the cells was the same as the YFP::TbCen2/p2T7_TbCen2 cell (see section 6.4.8 Knockdown of TbCen2 affects cell morphology). Examples shown include a cell with an elongated posterior end (Figure 100; B), a zoid cell that also has a detached flagellum (Figure 100; C) and a cell that has a detached new flagellum (NF) but the old flagellum (OF) is attached to the cell body.
In summary, by light microscopy, knockdown of TbCen2 had no effect on the localisation of YFP::TbHydin.

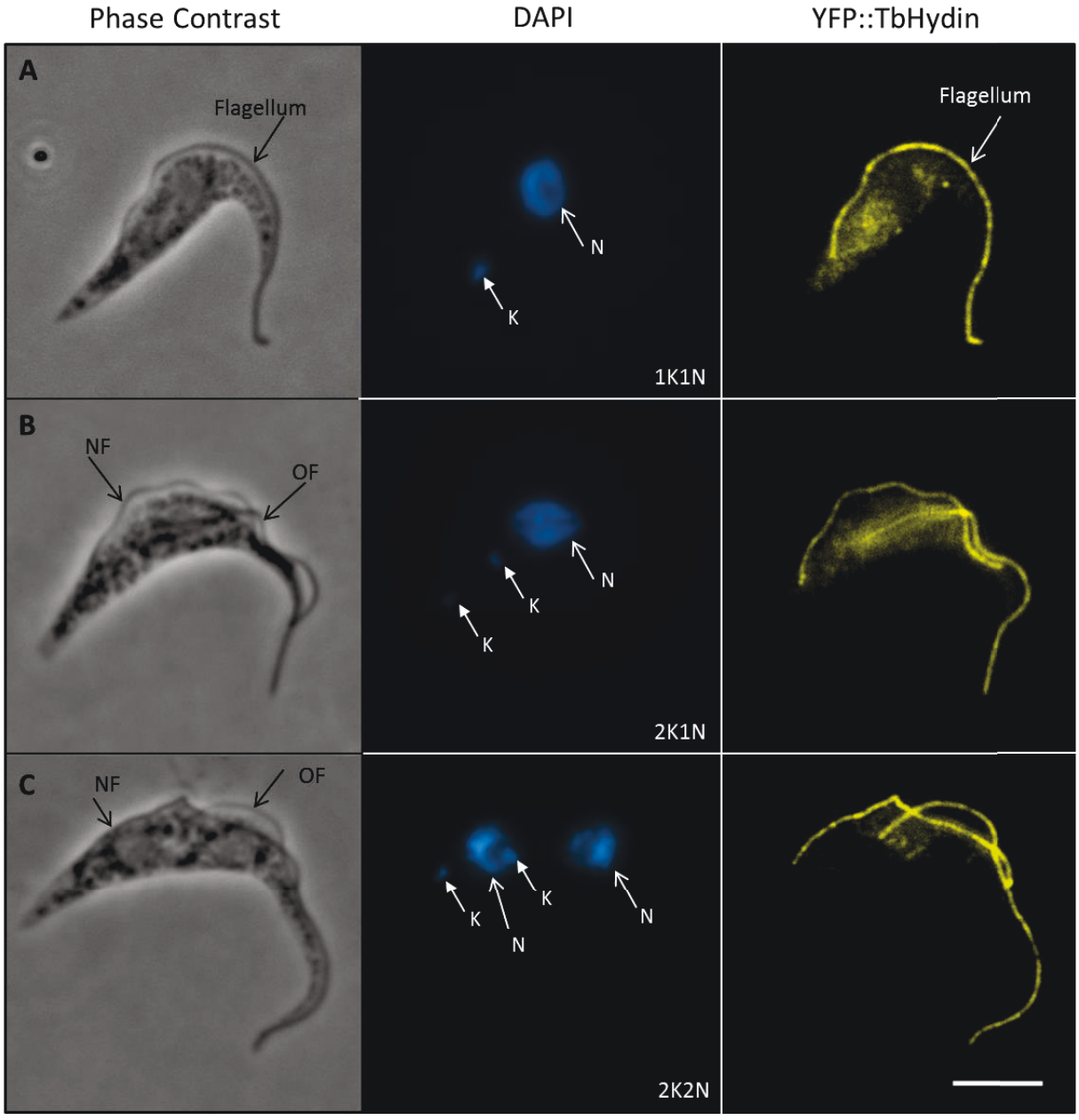


Figure 99: Localisation of YFP::TbHydin

TbHydin was tagged at the N terminus with YFP and expressed from the endogenous locus in procyclic *T. brucei*. Cells were fixed and labelled with DAPI to stain DNA. The YFP::TbHydin signal can be seen localised to the flagellum in 1K1N (A), 2K1N (B) and 2K2N (C) cells. NF; new flagellum, OF; old flagellum, K; kinetoplast, N; nucleus. Scale bar = 5µm

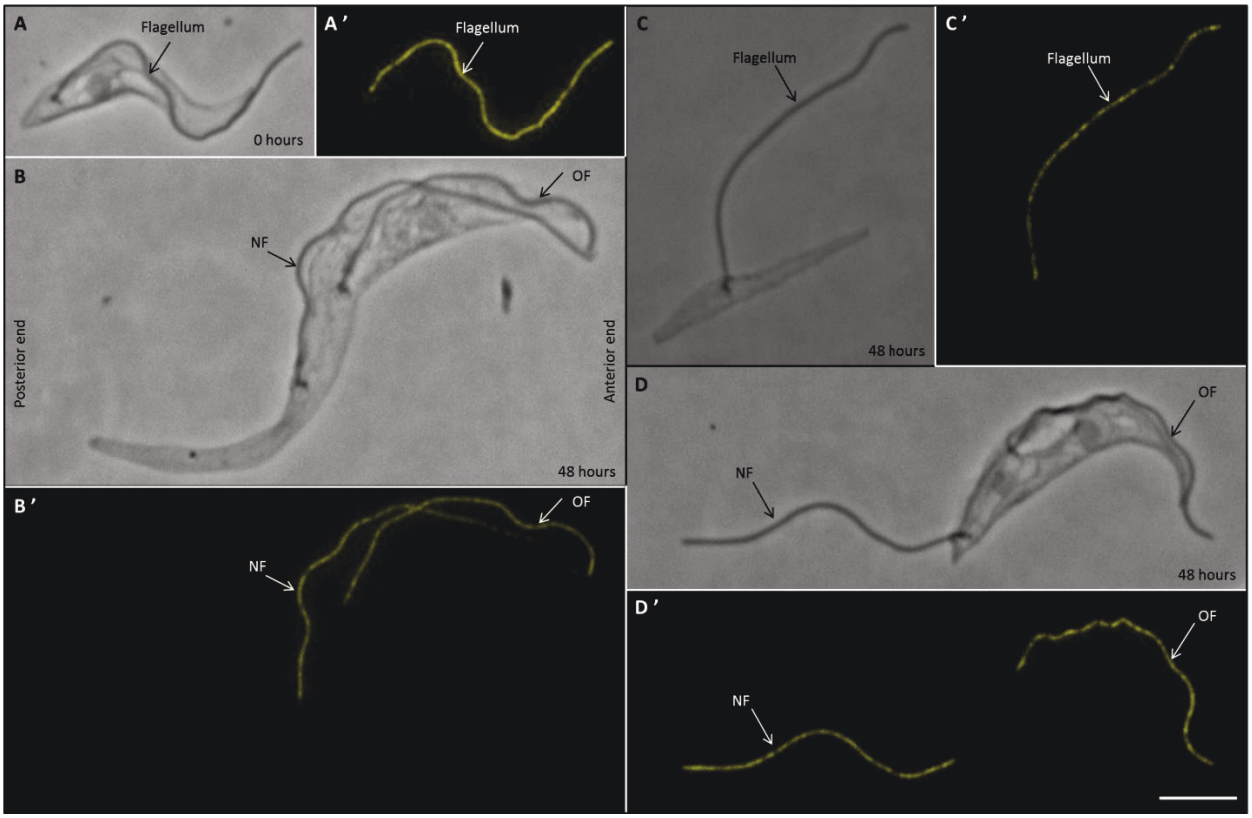


Figure 100: Knockdown of TbCen2 does not affect the localisation of YFP::TbHydin

YFP::TbHydin/p2T7_TbCen2 cells were induced. Cells were detergent extracted and fixed after 48 hours of TbCen2 knockdown. (A) Non-induced cell with YFP::TbHydin localised to the flagellum (A'), which was attached to the cell body. After TbCen2 knockdown the morphology of cells was altered but YFP::TbHydin remained localised to the flagellum in induced cells (B to D). (B) a cell with a long posterior end. (C) A zoid cell that has a detached flagellum. (D) A cell that has a detached NF and an attached OF. Scale bar = 5µm NF; new flagellum, OF; old flagellum.

6.5 Discussion

6.5.1 Localisation of TbCen2

TbCen2 is reported as localising to the flagellum of *T. brucei* (Ikeda and de Graffenried, 2012) although not in all publications (He *et al.*, 2005; Wang *et al.*, 2012). The flagellar localisation of TbCen2 is not surprising because the pan-centrin antibody, 20H5 (Sanders and Salisbury, 1994), is known to localise to the flagellum of *T. brucei* (Absalon *et al.*, 2008; de Graffenried *et al.*, 2008). In addition to flagellum localisation, TbCen2 localises to the basal bodies (He *et al.*, 2005) and the bilobe (de Graffenried *et al.*, 2013; Wang *et al.*, 2012). The bilobe is a cytoskeletal structure of unknown function, which has only been described in *T. brucei* (Esson *et al.*, 2012) (see section 1.6.9 Bilobe). The bilobe is located at the neck region of the flagellar pocket. TbCen2 and TbCen4 are among the few proteins known to localise at the bilobe (Morriswood *et al.*, 2013; Wang *et al.*, 2012). TbCen3 (Tb927.10.8710) localises to the axoneme, specifically the inner dynein arms (Wei *et al.*, 2014).

Although TbCen2 has been localised to the flagellum of *T. brucei* previously (Ikeda and de Graffenried, 2012) it was not known what the function of TbCen2 was or where it may localise at an ultrastructural level within the flagellum. Work presented in this chapter shows that YFP::TbCen2 expressed under the unknown endogenous promoter, localises to the flagellum in addition to the basal body and bilobe (Figure 83). However, the flagellum YFP::TbCen2 signal is not continuous from the basal body YFP::TbCen2 signal; there is a gap that implies YFP::TbCen2 is excluded from the transition zone (Figure 83; E). Therefore, the hypothesis is that the flagellum pool of TbCen2 is associated with the central pair apparatus (Figure 101; D), although this is not experimentally confirmed. (This is discussed in more detail in section 6.5.6 Central pair microtubule defect).

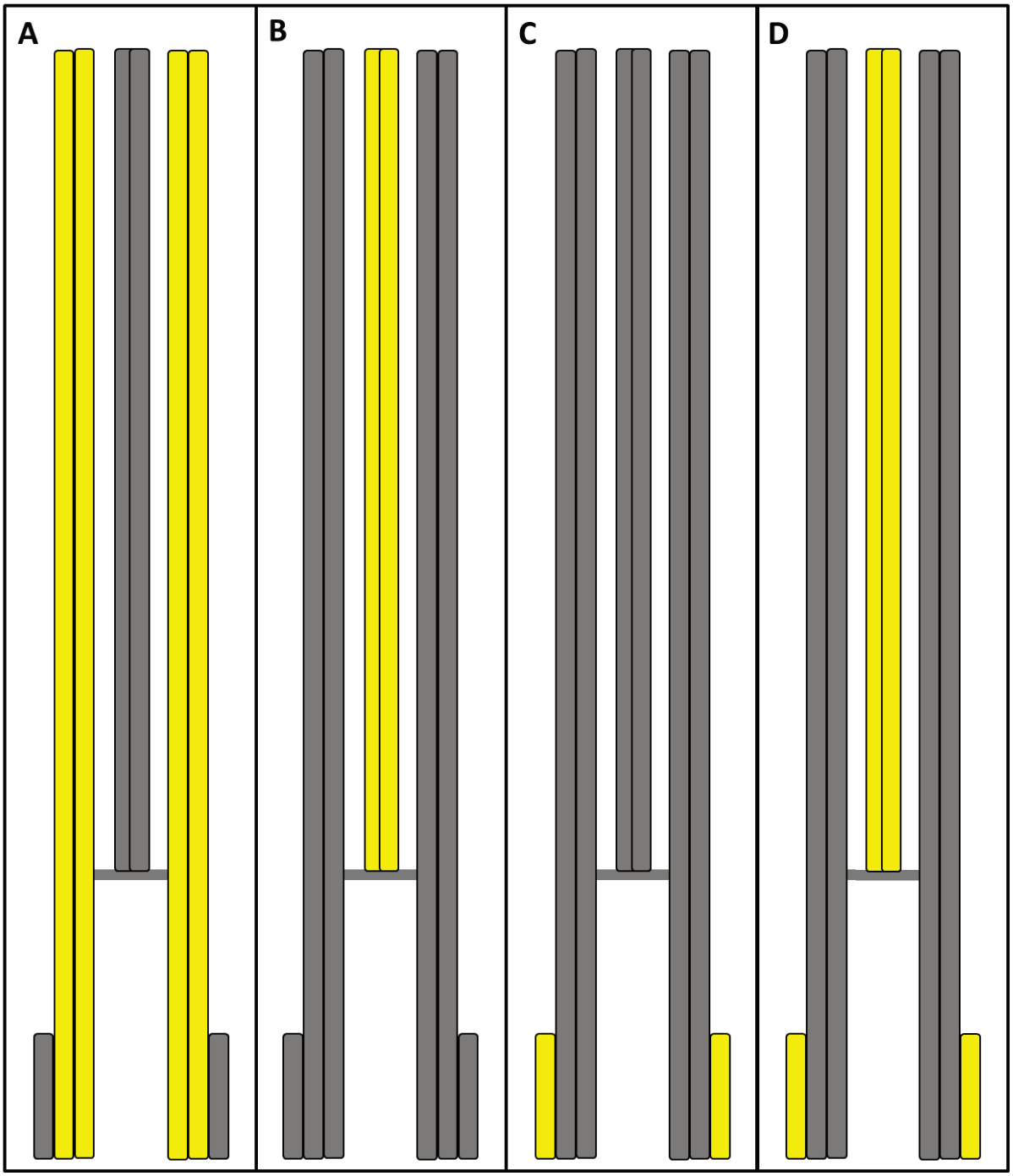


Figure 101: Schematic of protein localisation

A simplified illustration of YFP::protein localisations within the flagellum. YFP is represented in areas highlighted yellow, grey areas represent nil fluorescence. (A) YFP signal localised to the outer doublet microtubules from the proximal end of the basal body to the distal tip of the axoneme. (B) YFP signal localised to the central pair microtubules, which are nucleated from the basal plate. (C) YFP signal localised to the basal body. (D) YFP::TbCen2 signal at the basal body and central pair microtubules. PFR and bilobe structures not shown. Diagram not to scale.

6.5.2 Fluctuation of TbCen2 at the basal bodies

TbCen2 is present in both basal bodies (He *et al.*, 2005; Wang *et al.*, 2012). Previously, it was declared that the expression of TbCen2 was constant throughout the cell cycle (Wang *et al.*, 2012). It is important to note that this conclusion was reached from observations made of immunolabelled cells (anti-BB2) (He *et al.*, 2005) and immunoblots (Wang *et al.*, 2012), therefore any subtle difference in signal intensity could have been masked by saturation of antibody binding sites. The authors found competitive binding between TbCen2 and TbCen4 (Wang *et al.*, 2012); When TbCen4 was overexpressed, TbCen2 signal intensity on the bilobe decreased and when TbCen4 was knocked down, TbCen2 signal intensity on the bilobe increased (Wang *et al.*, 2012). It has been previously documented that RNAi targeted against TbCen1 did not affect expression of TbCen2 and *vice versa* (He *et al.*, 2005)

In this chapter, it was shown that levels of YFP::TbCen2 were not equal between the MBB and PBB of procyclic *T. brucei* with three types of signal patterns being documented in cells with a single flagellum (Figure 84). The proposed a model for how three different YFP::TbCen2 signal patterns could occur (Figure 102) is supported by the normal *T. brucei* cell cycle (Figure 12). The advantage of using *T. brucei* as an experimental model for basal body morphogenesis is that within the *T. brucei* cell it is relatively easy to classify which basal body is the oldest or most mature and which basal body is the youngest or pro-basal body. The flagellum remains intact through the cell cycle, unlike mammalian cells, and is identifiable as the new or old flagellum by its cellular position. In the model (Figure 102), the original cell 'Z' has a MBB (1) and a PBB (2). The specific age of the MBB was unknown but it is known that the PBB was formed during the previous cell cycle (Sherwin and Gull, 1989). During the subsequent cell cycle the PBB (2) will mature into a MBB, extend a new flagellum and form a new PBB (4). The original MBB (1) remains mature but also forms a new PBB (3). When this cell undergoes cytokinesis (dotted line), each daughter cell (X and Y) will inherit one pair of basal bodies. Each pair from the mother cell is different and there is no method of differentiating between the two post cytokinesis; one daughter cell 'X' inherits a pair of basal bodies where the MBB is a newly matured basal body with a PBB (2, 4). The other daughter cell 'Y' will inherit a pair of basal bodies where the MBB could be several cell cycles old but also has a PBB formed in the previous cell cycle (1, 3). Therefore three types of 1F cell can occur; Z, X or Y, each with one pair of

basal bodies. This normal cell cycle could account for why three patterns of BB YFP::TbCen2 fluorescence are seen (Figure 84). Although this analysis was indicative of a difference between the YFP::TbCen2 signal of the MBB and PBB, it is important to remember that this visual classification is subjective analysis.

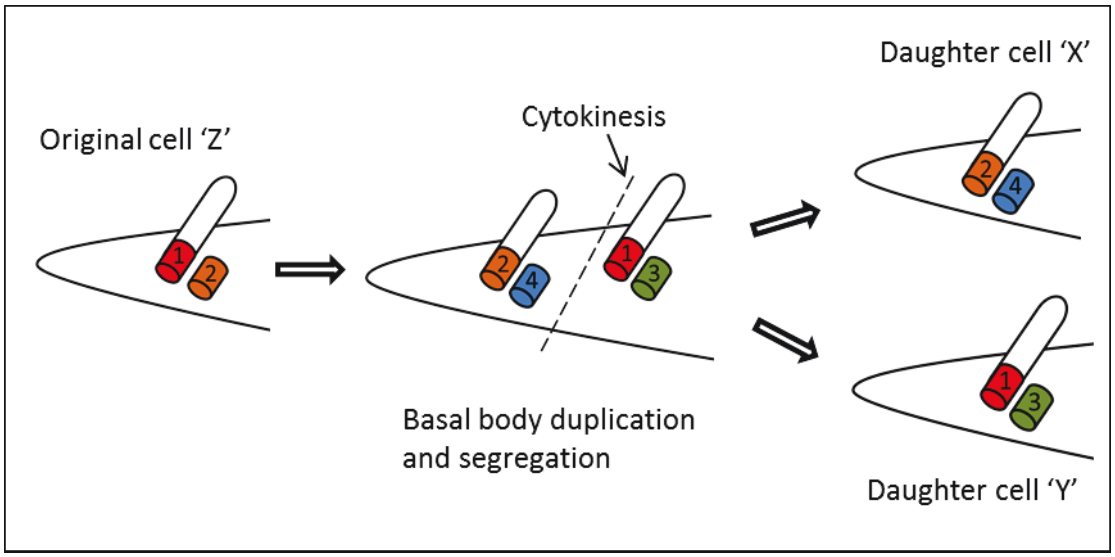


Figure 102: Model of basal body inheritance in *T. brucei*

A cartoon model of three cells with a single flagellum (Z, X, Y) that can occur in culture. There is no mechanism to evaluate how old the oldest mature basal body is (1) but the relative age of other basal bodies (2, 3 and 4) can be established in a cell with 2 flagella. Once cytokinesis has been completed the basal bodies of the two daughter cells, X and Y, are morphologically indistinguishable.

Using the length of the new flagellum as a marker of cell cycle progression, the YFP::TbCen2 signal intensity was analysed for all 4 basal bodies in a 2F cell (Figure 88). In the mature basal bodies there was a decrease in the YFP::TbCen2 signal intensity through the cell cycle. This analysis shows that there is a link between basal body age and YFP::TbCen2 signal intensity although the regulatory mechanism is unknown.

It is important to clarify that there are two tiers of basal body maturity/age within eukaryotic biology:

A. The mature basal body (mother) is older than the pro-basal body (daughter).

B. There is a difference in age between mature basal bodies of different cells

The first tier of basal body maturity (A), the idea of the mature and pro-basal body being of different ages/maturity, is one that is fairly well characterised and easy to study *in vitro* as a pair of centrioles/basal bodies are physically attached in close proximity to each other (Paintrand *et al.*, 1992). There are proteins that are specific to the mature centriole/basal body such as cenexin/ODF2 (Chen and Megraw, 2013; Lange and Gull, 1995) and ninein (Ou *et al.*, 2002) as well as ultrastructural differences between the MBB and PBB such as distal appendages, a transition zone and nucleating an axoneme. In some systems the mature and pro-basal bodies are referred to as mother and daughter basal body, which implies a generation of inheritance. The second tier of basal body maturation (B) is considered over a population rather than within a single cell and is equivalent to grandmother and great grandmother basal bodies.

The concept of basal body maturity has existed for many years (Vorobjev and Chentsov Yu, 1982). However, only recently has data emerged to prove that mature basal bodies are not equal to one another. A recent example describes that daughter cells inheriting a pair of basal bodies each and therefore could be equal, are not. One of the daughter cells builds a cilium before the other (Anderson and Stearns, 2009). The daughter cell that builds the cilium first is the one that inherited the oldest mature basal body (Anderson and Stearns, 2009). This has been attributed to the mature basal body retaining a portion of the cilia membrane across cytokinesis (Paridaen *et al.*, 2013).

In other model systems SAS-4, SAS-6, cenexin/ODF2, ninein and Cep164 have been identified as markers of centriole/basal body maturation (Chen and Megraw, 2013; Jord *et al.*, 2014; Lange and Gull, 1995; Novak *et al.*, 2014; Ou *et al.*, 2002; Schmidt *et al.*, 2012). Cenexin/ODF2 and ninein are not conserved in *T. brucei* but SAS-4, SAS-6 and Cep164 are conserved and could be studied in conjunction with TbCen2.

The tubulin family appear to have a key role in centriole maturation on both levels; ε is preferentially localised to the mature basal body (Chang and Stearns, 2000) and when δ tubulin is mutated in *C. reinhardtii* the basal bodies require an extra cell cycle to mature and assemble a flagellum (Dutcher and Trabuco, 1998).

T. brucei represents an excellent model organism to study basal body biogenesis because In *T. brucei* basal bodies are identifiable by light microscopy as being mature or pro-basal bodies according to their position; only mature basal bodies nucleate an axoneme; pro-basal bodies do not. In cells extending a second flagellum, prior to cell division, the 'new' flagellum will be positioned more posterior to the 'old' flagellum (Robinson *et al.*, 1995; Sherwin and Gull, 1989) (Figure 12) thus allowing identification of the oldest MBB and newest MBB with their respective PBBs. It is important to point out that in mammalian cells, the primary cilium is disassembled prior to the cell re-entering the cell cycle (Figure 4). The basal bodies of *C. reinhardtii* are also identifiable as mature or pro-basal bodies when the cell is flagellated (Boyd *et al.*, 2011), they also act as centrioles during mitosis (Coss, 1974) when the cell has undergone deflagellation. Therefore *T. brucei* is the best model organisms to use for investigating eukaryotic basal body maturity and inheritance.

An elegant set of experiments in mouse fibroblasts addressed the concept of basal body maturity using pulsed expression of tagged α tubulin to identify which basal body had been assembled pre or post expression of the tagged α tubulin (Anderson and Stearns, 2009). This approach could be utilised in *T. brucei* to trace which basal body was the newest mature post cytokinesis. A different approach to study cell fate in relation to centriole/basal body maturation is to use a photoswitchable fluorophore. This approach was used in *D. melanogaster* to track the inheritance pattern of centrosomes between daughter cells (Januschke *et al.*, 2011).

6.5.3 The function of TbCen2 in basal body biogenesis

Centrin orthologs have been localised to the MTOC of many cell types including the basal body of *C. reinhardtii* (Salisbury et al., 1988), the mammalian centriole (Errabolu *et al.*, 1994) and the spindle pole body of yeast (Spang *et al.*, 1993). Therefore it is not surprising that TbCen2 also localises at an MTOC of *T. brucei*, the basal body. Although the specific function of TbCen2 at the basal bodies has not been identified. The role of TbCen2 in basal body duplication and segregation has been disputed in the literature for many years. It was thought that depletion of TbCen2 inhibited basal body duplication (He *et al.*, 2005) but more recently it was published that TbCen2 had no involvement in basal body

duplication and segregation (Selvapandiyan *et al.*, 2012). The work presented in this thesis is in agreement with Selvapandiyan *et al.,* (2012).

Each flagellum of wild type *T. brucei* should have two basal bodies at the proximal end, the mature basal body, which nucleates the flagellum and the pro-basal body, which is connected to the mature basal body (Lacomble *et al.*, 2009). The pro-basal body does not nucleate a flagellum (Sherwin and Gull, 1989). Although the mature basal body and the pro-basal body both have the same triplet microtubule structure, there are ultrastructural and antigenic differences. Some antibodies will immunolabel both of the basal bodies such as BBA4 (Woodward *et al.*, 1995) whereas others may only immunolabel one basal body. An example of this is YL1/2 (Kilmartin *et al.*, 1982), which recognises an antigen specific to the mature basal body and not the pro-basal body (Stephan *et al.*, 2007).

It was published that depletion of TbCen2 inhibited basal body duplication (He *et al.*, 2005) as the authors only observed a single basal body after TbCen2 knockdown. However, those observations were made from cells that had been immunolabelled with YL1/2, which only recognises a single basal body; the mature basal body. This means the labelling pattern observed was normal. In this chapter cells were immunolabelled with BBA4, which labels both the mature and pro-basal body. The correct formation of two basal bodies (one pair) was observed in PCF cells after TbCen2 knockdown (Figure 90) and therefore TbCen2 is not essential for basal body duplication or segregation in PCF cells.

If the kinetoplastid resource database, TriTrypDB, is searched using the defined centrin-binding repeat (CBR) (Kilmartin, 2003) (see section 1.3.2 Centrin interacting proteins) then eight proteins are detected. Experimental data are not available on any of these 8 potential centrin-interacting proteins. Interacting and regulatory proteins of TbCen2 may hold the answer to the role of TbCen2 is basal body biogenesis. In *Paramecium*, PtCen2 recruits a FOP protein, PtFOR20 to the transition zone where it assists in basal body docking (Aubusson-Fleury *et al.*, 2012). An interaction between TbCen2 and TbFOR20 (Tb927.11.3.90) has not been investigated. The *T. brucei* ortholog of POC5 (Tb927.10.7600) would be an additional candidate to investigate in relation to TbCen2 expression levels as the human ortholog of POC5 is known to be essential for basal body assembly and is recruited in a cell cycle stage dependant manner (Azimzadeh *et al.*, 2009).

6.5.4 The function of TbCen2 in the flagellum

In a wildtype PCF *T. brucei* cell, YFP::TbCen2 localises to the new flagellum as it is assembled (Figure 83; B and C) but it is not clear what the function of TbCen2 is as ablation does not perturb flagellum assembly, although the assembled flagella are detached (Figure 91). It has previously been shown that PCF cells have a detached flagellum after TbCen2 knockdown (Selvapandiyan *et al.*, 2012). However, the flagellum detachment was not quantified by Selvapandiyan and colleagues (2012). Flagellum detachment was also reported when TbCen1 was knocked down (Selvapandiyan *et al.*, 2012), which is unexpected as TbCen1 is not known to localise to the flagellum (He *et al.*, 2005).

This study found no evidence of a basal body duplication defect when TbCen2 was ablated (Figure 90). After induction of RNAi against TbCen2, cells with more than one kinetoplast are observed in the population (Figure 89; A). This evidence indicates a kinetoplast segregation defect. It is unlikely that there is a total failure of kinetoplast duplication due to cells being observed with more than one kinetoplast but a failure of proper kinetoplast segregation prior to cytokinesis would lead to daughter cells with an aberrant number of DNA-containing organelles, which is a phenotype that is observed (Figure 89; A). Mechanistically, it is probable that the detached flagellum is the cause of inefficient kinetoplast segregation. New flagellum motility and rotation around the old flagellum is the machinery which divides the flagellar pocket (Lacomble *et al.*, 2010) and segregates the kinetoplast due to the kDNA being physically linked to the basal bodies via the TAC (Figure 13; B)(Gluenz *et al.*, 2011; Robinson and Gull, 1991). The migration of the new flagellum to a more posterior position occurs at the same point that the nabelschnur is resolved; completing kinetoplast segregation. Therefore if cells have a detached flagellum due to TbCen2 ablation, the cell no longer has the mechanical ability to separate the kinetoplast units.

TbCen2 has been shown to be a phosphorylation target of TbPLK (Yu *et al.*, 2012). When the phosphorylation site of TbCen2 is mutated, flagellum detachment and aberrant cytokinesis are seen (de Graffenried *et al.*, 2013). Morphologically, this is an overlap with the phenotype of TbCen2 knockdown (chapter 6) but de Graffenried and colleagues did

not investigate the effect of TbCen2 knockdown on TbPLK localisation (de Graffenried *et al.*, 2013).

6.5.5 Role of TbCen2 in regulating cell morphology

Knockdown of TbCen2 was not lethal in procyclic form *T. brucei* by 72 hours post induction. This decline in growth rate after TbCen2 knockdown has been documented before (Selvapandiyan *et al.*, 2012) and (Alsford *et al.*, 2011), (Appendix 15). Knockdown of TbCen2 has been reported as causing pleomorphic cells (Selvapandiyan *et al.*, 2012) but the exact morphology of this phenotype has not been described. Ablation of TbCen2 affected the duplication and segregation of DNA-Containing organelles (Figure 89) and (Selvapandiyan *et al.*, 2012). This excess of internal organelles caused a disruption to the normal morphology of the procyclic trypanomastigote form. The work presented in this chapter has used light and electron microscopy techniques to characterise the change in morphology following TbCen2 ablation.

The morphology of wildtype procyclic *T. brucei* cells is well characterised by light microscopy, standard transmission electron microscopy and electron tomography (Gluenz *et al.*, 2011; Lacomble *et al.*, 2009; Lacomble *et al.*, 2010; Ogbadoyi *et al.*, 2000; Robinson *et al.*, 1995; Sherwin and Gull, 1989; Vickerman *et al.*, 1988). Recently there have been major advances in volumetric electron microscopy (EM) techniques such as focused ion beam scanning electron microscopy (FIB-SEM) and serial block face SEM (SBF-SEM). SBF-SEM was first conceived in 1981 (Leighton, 1981) but was made commercially viable in 2004 (Denk and Horstmann, 2004). The principle of the technique is that there is an automated ultra-microtome internal to the SEM chamber that allows repeated sectioning and imaging of the resin block face (for reviews see (Hughes *et al.*, 2014; Peddie and Collinson, 2014)). The technique is still novel with less than 100 papers published using SBF-SEM in the field of biology; there are more papers within the field of material sciences. To date, there are two studies published on *T. brucei* utilising SBF-SEM; one focussed on a groove structure specific to the bloodstream form (Hughes *et al.*, 2013) and one that examined the flagellar pocket of both procyclic and bloodstream form cells (Demmel *et al.*, 2014).

It is thought that the rotation of the newly matured basal body and new flagellum around the old mature basal body and old flagellum is the method for pocket division (Lacomble *et al.*, 2010). However, in an example cell reconstruction (Figure 95; B) the new flagellum (red) has successfully rotated to be on the left of the old flagellum so loss of TbCen2 does not affect basal body rotation but the pocket has not divided and the new flagellum has not repositioned to be more posterior. This is interesting because TbCen2 is a target for phosphorylation by TbPLK (de Graffenried *et al.*, 2013; Yu *et al.*, 2012). When TbPLK is modified so that the ATP-binding site is inhibited, basal body rotation is prevented but the new flagellum does still extend (Lozano-Nunez *et al.*, 2013). This highlights the intrinsic link between TbPLK and TbCen2 in flagellum biogenesis (de Graffenried *et al.*, 2013; Lozano-Nunez *et al.*, 2013).

6.5.6 Central pair microtubule defect

Previously, a brief examination of procyclic *T. brucei* cells by TEM observed that TbCen2 RNAi cells are multinucleated (Selvapandiyan *et al.*, 2012) but no in depth study was made of the axoneme ultrastructure. To investigate if the assembled flagellum was ultrastructurally normal cells were detergent extracted for observation of the cytoskeleton by TEM (Figure 97).

Following ablation of TbCen2, examination of axonemes by TEM revealed that there was a defect in the C2 microtubule of the central pair complex (Figure 97). A ventral portion of unknown protofilaments was absent from the C2 microtubule. There are two hypotheses for how this occurs:

1) The C2 microtubule was not fully assembled

2) The C2 microtubule was fully assembled but was unstable and therefore fragmented

When a microtubule is assembled from a protofilament sheet into a tubule, there is a ‘seam’ of α to β tubulin interactions (Figure 1). The seam is the weakest point of a microtubule (Simon and Salmon, 1990). In the central pair, the seam of the C2 microtubule is between protofilament 1 and 13 (Figure 104; A), adjacent to an inter-microtubule bridge. Although it was not possible to determine the exact protofilaments

that were absent in the partial C2 microtubule phenotype, it is likely that the missing section starts at the seam.

If the C2 was not properly assembled along the whole length of the flagellum, it was expected that the observation of a partial C2 microtubule by electron microscopy would be more frequent. If this hypothesis is correct the C2 microtubule was assembled in full but destabilised by the absence of TbCen2 and therefore the ventral portion would only be absent at the unstable points of the microtubule (like unzipping a zip) (Figure 103). Although I have not directly localised TbCen2 to the C2 microtubule, the knockdown phenotype of TbCen2 implies that TbCen2 is necessary to stabilise the ventral section of the C2 microtubule and/or support the C2b projection.

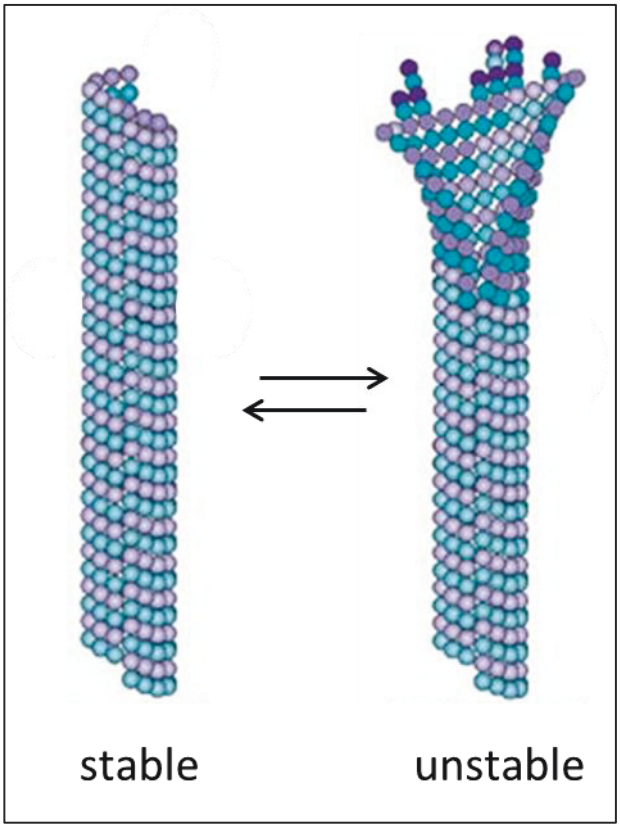


Figure 103: Microtubule stability

A proposed model for TbCen2 loss based on the dynamic instability model of microtubule dynamics. Adapted from (Conde and Caceres, 2009).

The central pair apparatus of *C. reinhardtii* and *T. brucei* are known to have a slightly different abilities; the central pair of *C. reinhardtii* rotates when the flagellum beats (Mitchell and Nakatsugawa, 2004; Omoto *et al.*, 1999) and the central pair of *T. brucei*

does not (Gadelha *et al.*, 2006). Central pair rotation was first observed in the ciliate, *P. tetraurelia* (Omoto and Kung, 1980).

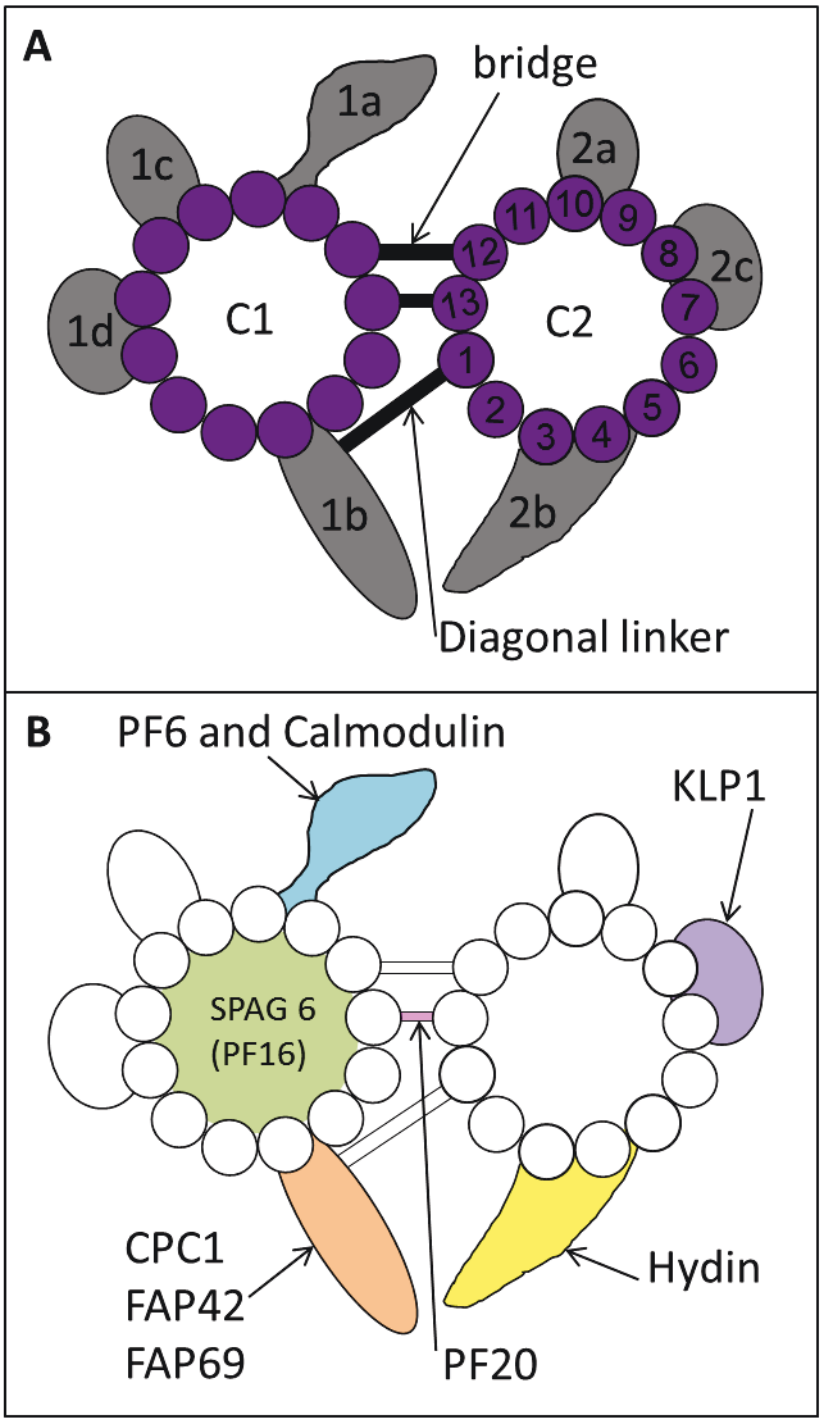


Figure 104: Composition of the central pair microtubules in *C. reinhardtii*

(A) Schematic diagram of the central pair complex ultrastructure, including C1 and C2 microtubules with asymmetric projections. Projection annotation based on (Wargo and Smith, 2003) (B) A colour map of the protein components of the central pair that have been identified. Localisation data for PF6 (Mitchell and Sale, 1999; Wargo *et al.*, 2005), calmodulin (Wargo *et al.*, 2005), KLP1 (Bernstein *et al.*, 1994), SPAG6/PF16

(Smith and Lefebvre, 1996), hydin (Lechtreck and Witman, 2007), PF20 (Smith and Lefebvre, 1997a), CPC1 (Mitchell and Sale, 1999), FAP42 and FAP69 (Mitchell, 2009).

There is a lack of published information about the central pair microtubule complex in the *T. brucei* axoneme. Most of the knowledge about the central pair apparatus of a eukaryotic axoneme has been derived from work on *C. reinhardtii* (Lechtreck *et al.*, 2013; Lechtreck and Witman, 2007). Figure 104 shows the known ultrastructural configuration of the central pair apparatus with associated projections in *C. reinhardtii* (Figure 104; A) and the protein components identified so far (Figure 104; B), excluding α and β tubulin of the microtubules. Proteomic analysis of isolated central pair apparatus predicts there are at least 23 distinct proteins (Dutcher *et al.*, 1984). This experiment should be repeated in the post genomic age.

In *C. reinhardtii* calmodulin is part of the central pair C1 microtubule complex (Wargo *et al.*, 2005) (Figure 104; B). Calmodulin is conserved in *T. brucei*, however, knockdown of disrupts the PFR (Ginger *et al.*, 2013) and no effect on the central pair apparatus was reported. Several isoforms of calmodulin are conserved in *T. brucei* (Berriman *et al.*, 2005; Ginger *et al.*, 2013) and therefore it is possible that a different isoform has a conserved role in the C1 complex of *T. brucei*.

The only experimentally confirmed component of the *T. brucei* central pair is Hydin (Dawe *et al.*, 2007), which was first identified in *C. reinhardtii* (Lechtreck and Witman, 2007). The C2b projection is known to be comprised of Hydin in *C. reinhardtii* (Lechtreck and Witman, 2007), mice (Lechtreck *et al.*, 2008) and humans (Olbrich *et al.*, 2012). Hydin has not been localised by electron microscopy within *T. brucei*. Knockdown of Hydin in *T. brucei* causes the central pair to become displaced (Dawe *et al.*, 2007), inferring that the role of TbHydin is to stabilise the central pair microtubules. For this reason TbHydin was chosen for localisation studies in the p2T7_TbCen2 cell line. However, ablation of TbCen2 did not disrupt the localisation of YFP::TbHydin when cells were examined by light microscopy. One possibility is that TbHydin may be lost when TbCen2 is knocked down but YFP::TbHydin may remain, if it has the extra YFP to 'wedge' it in. This could only be answered with an anti-hydin antibody to immunolabel native TbHydin after TbCen2 knockdown.

Another approach would be to knock down γ tubulin in the same cell line as YFP::TbCen2. This would allow a conclusion to be drawn on whether TbCen2 is a component of the central pair apparatus as γ tubulin is essential for central pair microtubule nucleation (McKean *et al.*, 2003).

Although a partial microtubule phenotype is unusual there is precedent for it. The *C. reinhardtii unc1* mutant strain has a similar phenotype to the results presented here for TbCen2 knockdown, with up to ¾ of the C2 microtubule being absent (Mitchell and Smith, 2009). The pf16 mutant strain has also been commented to 'affect the stability of the tubulin lattice' although no further description of the phenotype is provided (Dutcher *et al.*, 1984).

6.5.7 Conclusions

I have used endogenous expression to localise TbCen2 to the basal bodies, bilobe and flagellum. I have identified a variation in the level of YFP::TbCen2 at the basal bodies through the cell cycle of procyclic *T. brucei,* which suggests a link between basal body maturity and TbCen2. Prior to this study, TbCen2 had not been linked to the central pair or associated projections but the YFP::TbCen2 localisation result suggests it is present in the central pair region. Ablation of TbCen2 highlights a previously unidentified role of TbCen2 in maintenance of the central pair apparatus.

7. General Discussion

This thesis presents a bioinformatics screen, which identified 26 potential novel flagellum proteins within *T. brucei* (Table 14). The bioinformatics analysis of 8 proteins; TbCCDC96, TbCCDC113, TbRib72, TbDRC5, TbWDR92, TbCCDC19, TbCCDC11 and TbMNS1 is presented in chapters 3 and 4. Four of the original 26 proteins were localised to the flagellum (TbCCDC19, TbCCDC96, TbRib72 and TbDRC5) and an additional 3 proteins were localised to the flagellum (TbCCDC113, TbCCDC11 and TbMNS1) that were not on the original candidate list but deemed important by association. TbCen2 was also localised to the flagellum, bilobe and basal bodies.

Prior to the work presented in chapter 4, the family of TPH domain-containing proteins had not been identified. Bioinformatics analysis revealed that the TPH domain is a widely conserved eukaryotic domain with three proteins members present in *T. brucei*. The localisation and functionality of these three TPH domain-containing proteins in *T. brucei* was investigated in chapter 5. This work demonstrated that all three TPH domain-containing proteins localised, under the endogenous promoter, to the flagellum in procyclic form cells. Another discovery was that depletion of each TPH domain-containing protein individually caused an increase in flagellum length and elongation of the posterior end. This physical alteration of the cell shape and flagellum length decreased the motility of procyclic trypanomastigotes. The sub-cellular localisation of many trypanosomal flagellum proteins has yet to be pinpointed on an intra-flagellum scale. It will be interesting to identify where in the flagellum the TPH domain-containing proteins localise; if they are components of the central pair complex or associated with the axonemal microtubule doublets.

The work presented in chapter 6 provides new insights into fluctuating TbCen2 protein levels associated with the cell cycle and indicates a link between TbCen2 and basal body maturity. When TbCen2 was knocked down by RNAi, the C2 phenotype provides the first partial microtubule mutant in *T. brucei.* TbCen2 is important for normal cell growth, flagellum attachment, stability of the central pair apparatus and cytokinesis in procyclic form *T. brucei*.

References

Abrahamsen, M. S. *et al.* (2004). Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304 (5669), pp.441-445.

Absalon, S. *et al.* (2008). Intraflagellar transport and functional analysis of genes required for flagellum formation in trypanosomes. *Mol Biol Cell* 19 (3), pp.929-944.

Adams, C. M., Danks, D. M. and Campbell, P. E. (1974). Comments upon the classification of infantile polycystic diseases of the liver and kidney, based upon three-dimensional reconstruction of the liver. *J Med Genet* 11 (3), pp.234-243.

Adams, M. D. *et al.* (2000). The genome sequence of *Drosophila melanogaster*. *Science* 287 (5461), pp.2185-2195.

Adams, N. A., Awadein, A. and Toma, H. S. (2007). The retinal ciliopathies. *Ophthalmic Genet* 28 (3), pp.113-125.

Afzelius, B. A. (1998). Immotile cilia syndrome: past, present, and prospects for the future. *Thorax* 53 (10), pp.894-897.

Akiyoshi, B. and Gull, K. (2014). Discovery of unconventional kinetochores in kinetoplastids. *Cell* 156 (6), pp.1247-1258.

Albee, A. J. *et al.* (2013). Identification of cilia genes that affect cell cycle progression using whole genome transcriptome analysis in *Chlamydomonas reinhardtti*. *G3 (Bethesda)* 3 (6), pp.979-991.

Alizadehrad, D. *et al.* (2015). Simulating the complex cell design of *Trypanosoma brucei* and its motility. *PLoS Comput Biol* 11 (1), p.e1003967.

Allen, C. L., Goulding, D. and Field, M. C. (2003). Clathrin-mediated endocytosis is essential in *Trypanosoma brucei*. *EMBO J* 22 (19), pp.4991-5002.

Alsford, S. *et al.* (2011). High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res* 21 (6), pp.915-924.

Alvey, P. L. (1986). Do adult centrioles contain cartwheels and lie at right angles to each other? *Cell Biol Int Rep* 10 (8), pp.589-598.

Amack, J. D. (2014). Salient features of the ciliated organ of asymmetry. *Bioarchitecture* 4 (1), pp.6-15.

Amaral, A. *et al.* (2013). Human sperm tail proteome suggests new endogenous metabolic pathways. *Mol Cell Proteomics* 12 (2), pp.330-342.

Andersen, J. S. *et al.* (2003). Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* 426 (6966), pp.570-574.

Anderson, C. T. *et al.* (2008). Primary cilia: cellular sensors for the skeleton. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology* 291 (9), pp.1074-1078.

Anderson, C. T. and Stearns, T. (2009). Centriole age underlies asynchronous primary cilium growth in mammalian cells. *Curr Biol* 19 (17), pp.1498-1502.

Anderson, R. G. (1972). The three-dimensional structure of the basal body from the rhesus monkey oviduct. *J Cell Biol* 54 (2), pp.246-265.

Arabidopsis Genome, I. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408 (6814), pp.796-815.

Archambault, V., Lepine, G. and Kachaner, D. (2015). Understanding the polo kinase machine. *Oncogene*.

Arhin, G. *et al.* (2004). A PCR-based method for gene deletion and protein tagging in *Trypanosoma brucei*. In: Melville, S. (ed.) *Parasite Genomics Protocols*. Vol. 270. Humana Press, pp.277-286.

Armbrust, E. V. *et al.* (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306 (5693), pp.79-86.

Arnaiz, O. *et al.* (2010). Gene expression in a paleopolyploid: a transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genomics* 11, p.547.

Arnaiz, O. *et al.* (2009). Cildb: a knowledgebase for centrosomes and cilia. *Database (Oxford)* 2009, p.bap022.

Aslett, M. *et al.* (2010). TriTrypDB: a functional genomic resource for the *Trypanosomatidae*. *Nucleic Acids Res* 38 (Database issue), pp.D457-462.

Aubusson-Fleury, A. *et al.* (2012). The conserved centrosomal protein FOR20 is required for assembly of the transition zone and basal body docking at the cell surface. *J Cell Sci* 125 (Pt 18), pp.4395-4404.

Aurrecoechea, C. *et al.* (2009). PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res* 37 (Database issue), pp.D539-543.

Aury, J. M. *et al.* (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444 (7116), pp.171-178.

Avasthi, P. and Marshall, W. F. (2011). Stages of ciliogenesis and regulation of ciliary length. *Differentiation*.

Avidor-Reiss, T. *et al.* (2004). Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. *Cell* 117 (4), pp.527-539.

Azimzadeh, J. and Bornens, M. (2005). The centrosome in evolution. In: Nigg, E. A. (ed.) *Centrosomes in development and disease*. Wiley-VCH Verlag GmbH & Co. KGaA, pp.93-122.

Azimzadeh, J. *et al.* (2009). hPOC5 is a centrin-binding protein required for assembly of full-length centrioles. *J Cell Biol* 185 (1), pp.101-114.

Azimzadeh, J. and Marshall, W. F. (2010). Building the centriole. *Curr Biol* 20 (18), pp.R816-825.

Baccetti, B., Dallai, R. and Rosati, F. (1970). The spermatozoon of arthropoda. 8. The 9 + 3 flagellum of spider sperm cells. *J Cell Biol* 44 (3), pp.681-682.

Badano, J. L. *et al.* (2006). The ciliopathies: an emerging class of human genetic disorders. *Annu Rev Genomics Hum Genet* 7, pp.125-148.

Bahe, S. *et al.* (2005). Rootletin forms centriole-associated filaments and functions in centrosome cohesion. *J Cell Biol* 171 (1), pp.27-33.

Bai, D. *et al.* (2009). DNA variants in coding region of EFHC1: SNPs do not associate with juvenile myoclonic epilepsy. *Epilepsia* 50 (5), pp.1184-1190.

Baker, K. and Beales, P. L. (2009). Making sense of cilia in disease: the human ciliopathies. *Am J Med Genet C Semin Med Genet* 151C (4), pp.281-295.

Baker, M. A. *et al.* (2013). Head and flagella subcompartmental proteomic analysis of human spermatozoa. *Proteomics* 13 (1), pp.61-74.

Baldari, C. T. and Rosenbaum, J. (2010). Intraflagellar transport: it's not just for cilia anymore. *Curr Opin Cell Biol* 22 (1), pp.75-80.

Balestra, F. R. *et al.* (2013). Discovering regulators of centriole biogenesis through siRNA-based functional genomics in human cells. *Dev Cell* 25 (6), pp.555-571.

Baron, D. M., Kabututu, Z. P. and Hill, K. L. (2007a). Stuck in reverse: loss of LC1 in *Trypanosoma bruce*i disrupts outer dynein arms and leads to reverse flagellar beat and backward movement. *J Cell Sci* 120 (Pt 9), pp.1513-1520.

Baron, D. M. *et al.* (2007b). Functional genomics in *Trypanosoma brucei* identifies evolutionarily conserved components of motile flagella. *J Cell Sci* 120 (Pt 3), pp.478-491.

Barr, D. J. S. (1981). The phylogenetic and taxonomic implications of flagellar rootlet morphology among zoosporic fungi. *Biosystems* 14 (3–4), pp.359-370.
Barry, S., Towers, K. and Vaughan, S. (2012). Screening ciliopathy genes in the model organism *Trypanosoma brucei*. *Cilia* 1 (Suppl 1), pp.P81-P81.
Bastin, P. *et al.* (1996). A novel epitope tag system to study protein targeting and organelle biogenesis in *Trypanosoma brucei*. *Mol Biochem Parasitol* 77 (2), pp.235-239.
Bastin, P. *et al.* (1999a). Flagellar morphogenesis: protein targeting and assembly in the paraflagellar rod of trypanosomes. *Mol Cell Biol* 19 (12), pp.8191-8200.
Bastin, P. *et al.* (1999b). Protein transport and flagellum assembly dynamics revealed by analysis of the paralysed trypanosome mutant snl-1. *J Cell Sci* 112 (Pt 21), pp.3769-3777.
Bastin, P., Sherwin, T. and Gull, K. (1998). Paraflagellar rod is vital for trypanosome motility. *Nature* 391 (6667), p.548.
Batsios, P. *et al.* (2012). A lamin in lower eukaryotes? *Nucleus* 3 (3), pp.237-243.
Bauer, P. G. (1984). Electron microscopical studies on *Trypanosoma cruzi* and other microorganisms in the reduviid vector. *Mem. Inst. Oswaldo Cruz [online]* 79, pp.25-32.
Baum, P., Furlong, C. and Byers, B. (1986). Yeast gene required for spindle pole body duplication: homology of its product with Ca2+-binding proteins. *Proc Natl Acad Sci U S A* 83 (15), pp.5512-5516.
Beattie, P. and Gull, K. (1997). Cytoskeletal architecture and components involved in the attachment of *Trypanosoma congolense* epimastigotes. *Parasitology* 115 (Pt 1), pp.47-55.
Becker-Heck, A. *et al.* (2011). The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation. *Nat Genet* 43 (1), pp.79-84.
Beisson, J. *et al.* (2010). *Paramecium tetraurelia:* The renaissance of an early unicellular model. *Cold Spring Harbor Protocols* 2010 (1), p.pdb.emo140.
Bernstein, M. *et al.* (1994). A new kinesin-like protein (Klp1) localized to a single microtubule of the *Chlamydomonas* flagellum. *J Cell Biol* 125 (6), pp.1313-1326.
Berriman, M. *et al.* (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309 (5733), pp.416-422.
Berthelot, C. *et al.* (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun* 5, p.3657.
Bettencourt-Dias, M. *et al.* (2011). Centrosomes and cilia in human disease. *Trends Genet* 27 (8), pp.307-315.
Birkett, C. R. *et al.* (1985). Use of monoclonal antibodies to analyse the expression of a multi-tubulin family. *FEBS Lett* 187 (2), pp.211-218.
Blackman, L. M., Harper, J. D. and Overall, R. L. (1999). Localization of a centrin-like protein to higher plant plasmodesmata. *Eur J Cell Biol* 78 (5), pp.297-304.
Blake, D. P. *et al.* (2012). EmaxDB: Availability of a first draft genome sequence for the apicomplexan *Eimeria maxima*. *Mol Biochem Parasitol* 184 (1), pp.48-51.
Bolanos-Garcia, V. M. (2005). Aurora kinases. *Int J Biochem Cell Biol* 37 (8), pp.1572-1577.
Boldt, K. *et al.* (2011). Disruption of intraflagellar protein transport in photoreceptor cilia causes Leber congenital amaurosis in humans and mice. *J Clin Invest* 121 (6), pp.2169-2180.
Bonhivers, M. *et al.* (2008a). A monoclonal antibody marker for the exclusion-zone filaments of *Trypanosoma brucei*. *Parasit Vectors* 1 (1), p.21.

Bonhivers, M. *et al.* (2008b). Biogenesis of the trypanosome endo-exocytotic organelle is cytoskeleton mediated. *PLoS Biol* 6 (5), p.e105.

Boyd, J. S., Mittelmeier, T. M. and Dieckmann, C. L. (2011). New insights into eyespot placement and assembly in *Chlamydomonas*. *Bioarchitecture* 1 (4), pp.196-199.

Bradley, B. A. and Quarmby, L. M. (2005). A NIMA-related kinase, Cnk2p, regulates both flagellar length and cell size in *Chlamydomonas*. *J Cell Sci* 118 (Pt 15), pp.3317-3326.

Branche, C. *et al.* (2006). Conserved and specific functions of axoneme components in trypanosome motility. *J Cell Sci* 119 (Pt 16), pp.3443-3455.

Briggs, L. J. *et al.* (2004a). More than one way to build a flagellum: comparative genomics of parasitic protozoa. *Curr Biol* 14 (15), pp.R611-612.

Briggs, L. J. *et al.* (2004b). The flagella connector of *Trypanosoma brucei*: an unusual mobile transmembrane junction. *J Cell Sci* 117 (Pt 9), pp.1641-1651.

Broadhead, R. *et al.* (2006). Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* 440 (7081), pp.224-227.

Brokaw, C. J. (1991). Microtubule sliding in swimming sperm flagella: direct and indirect measurements on sea urchin and tunicate spermatozoa. *J Cell Biol* 114 (6), pp.1201-1215.

Brokaw, C. J. (2009). Thinking about flagellar oscillation. *Cell Motil Cytoskeleton* 66 (8), pp.425-436.

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* 36 (3), pp.289-292.

Buisson, J. and Bastin, P. (2010). Flagellum structure and function in trypanosomes. In: de Souza, W. (ed.) *Structures and Organelles in Pathogenic Protists*. Springer Berlin / Heidelberg, pp.63-86.

Butler, C. E. and Tyler, K. M. (2012). Membrane traffic and synaptic cross-talk during host cell entry by *Trypanosoma cruzi*. *Cell Microbiol* 14 (9), pp.1345-1353.

Carmena, M. and Earnshaw, W. C. (2003). The cellular geography of aurora kinases. *Nat Rev Mol Cell Biol* 4 (11), pp.842-854.

Carvalho-Santos, Z. *et al.* (2011). Evolution: Tracing the origins of centrioles, cilia, and flagella. *J Cell Biol* 194 (2), pp.165-175.

Carvalho-Santos, Z. *et al.* (2010). Stepwise evolution of the centriole-assembly pathway. *J Cell Sci* 123 (Pt 9), pp.1414-1426.

Castleman, V. H. *et al.* (2009). Mutations in radial spoke head protein genes RSPH9 and RSPH4A cause primary ciliary dyskinesia with central-microtubular-pair abnormalities. *Am J Hum Genet* 84 (2), pp.197-209.

Cavalier-Smith, T. (1974). Basal body and flagellar development during the vegetative cell cycle and the sexual cycle of *Chlamydomonas reinhardii*. *J Cell Sci* 16 (3), pp.529-556.

Cavalier-Smith, T. (1993). Kingdom protozoa and its 18 phyla. *Microbiol Rev* 57 (4), pp.953-994.

Cerqua, C. *et al.* (2010). Trichoplein/mitostatin regulates endoplasmic reticulum-mitochondria juxtaposition. *EMBO Rep* 11 (11), pp.854-860.

Chang, P. and Stearns, T. (2000). Delta-tubulin and epsilon-tubulin: two new human centrosomal tubulins reveal new aspects of centrosome structure and function. *Nat Cell Biol* 2 (1), pp.30-35.

Chen, J. V. and Megraw, T. L. (2013). Cenexin1 and Odf2: splice variants with diverged cilium functions. *Cell Cycle* 12 (6), p.869.

Clarke, M. *et al.* (2013). Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol* 14 (2), p.R11.

Claser, C. *et al.* (2008). Silencing cytokeratin 18 gene inhibits intracellular replication of Trypanosoma cruzi in HeLa cells but not binding and invasion of trypanosomes. *BMC Cell Biol* 9, p.68.

Cleveland, D. W., Kirschner, M. W. and Cowan, N. J. (1978). Isolation of separate mRNAs for alpha- and beta-tubulin and characterization of the corresponding in vitro translation products. *Cell* 15 (3), pp.1021-1031.

Cohen, E., Binet, S. and Meininger, V. (1988). Ciliogenesis and centriole formation in the mouse embryonic nervous system. An ultrastructural analysis. *Biol Cell* 62 (2), pp.165-169.

Conde, C. and Caceres, A. (2009). Microtubule assembly, organization and dynamics in axons and dendrites. *Nat Rev Neurosci* 10 (5), pp.319-332.

Consortium, C. e. S. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282 (5396), pp.2012-2018.

Coss, R. A. (1974). Mitosis in *Chlamydomonas reinhardtii* basal bodies and the mitotic apparatus. *J Cell Biol* 63 (1), pp.325-329.

Cottee, M. A. *et al.* (2011). SAS-6 oligomerization: the key to the centriole? *Nat Chem Biol* 7 (10), pp.650-653.

Cross, G. A. and Manning, J. C. (1973). Cultivation of *Trypanosoma brucei* sspp. in semi-defined and defined media. *Parasitology* 67 (3), pp.315-331.

Czarnecki, P. G. and Shah, J. V. (2012). The ciliary transition zone: from morphology and molecules to medicine. *Trends Cell Biol* 22 (4), p.201=210.

Darde, M. *et al.* (2008). *Toxoplasma gondii*. In: Khan, N. A. (ed.) *Emerging protozoan pathogens*. Taylor & Francis, pp.227-288.

Davidge, J. A. *et al.* (2006). Trypanosome IFT mutants provide insight into the motor location for mobility of the flagella connector and flagellar membrane formation. *J Cell Sci* 119 (Pt 19), pp.3935-3943.

Davids, B. J. *et al.* (2008). *Giardia lamblia* aurora kinase: a regulator of mitosis in a binucleate parasite. *Int J Parasitol* 38 (3-4), pp.353-369.

Dawe, H. R. *et al.* (2007). The hydrocephalus inducing gene product, Hydin, positions axonemal central pair microtubules. *BMC Biol* 5, p.33.

de Graffenried, C. L. *et al.* (2013). Polo-like kinase phosphorylation of bilobe-resident TbCentrin2 facilitates flagellar inheritance in *Trypanosoma brucei*. *Mol Biol Cell* 24 (12), pp.1947-1963.

de Graffenried, C. L., Ho, H. H. and Warren, G. (2008). Polo-like kinase is required for Golgi and bilobe biogenesis in *Trypanosoma brucei*. *J Cell Biol* 181 (3), pp.431-438.

de Nijs, L. *et al.* (2006). EFHC1, a protein mutated in juvenile myoclonic epilepsy, associates with the mitotic spindle through its N-terminus. *Exp Cell Res* 312 (15), pp.2872-2879.

Dean, S. *et al.* (2015). A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. *Open Biol* 5 (1).

Deane, M. P. and Milder, R. (1973). Unusual structures in the flagellar apparatus of a strain of *Trypanosoma cruzi*. *J Protozool* 20 (5), pp.586-589.

Dehal, P. *et al.* (2002). The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298 (5601), pp.2157-2167.

Delaval, B. *et al.* (2011). Centrin depletion causes cyst formation and other ciliopathy-related phenotypes in zebrafish. *Cell Cycle* 10 (22), pp.3964-3972.

Demmel, L. *et al.* (2014). The endocytic activity of the flagellar pocket in *Trypanosoma brucei* is regulated by an adjacent phosphatidylinositol phosphate kinase. *J Cell Sci* 127 (Pt 10), pp.2351-2364.

Denk, W. and Horstmann, H. (2004). Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure. *PLoS Biol* 2 (11), p.e329.

Dentler, W. (2005). Intraflagellar transport (IFT) during assembly and disassembly of *Chlamydomonas* flagella. *J Cell Biol* 170 (4), pp.649-659.

Dippell, R. V. (1967). How ciliary Basal bodies develop. *Science* 158 (3800), p.527.

Dippell, R. V. (1968). The development of basal bodies in *Paramecium*. *Proc Natl Acad Sci U S A* 61 (2), pp.461-468.

Dobbelaere, J. *et al.* (2008). A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in *Drosophila*. *PLoS Biol* 6 (9), p.e224.

Dorus, S. *et al.* (2006). Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nat Genet* 38 (12), pp.1440-1445.

Doudney, K. *et al.* (2001). Comparative physical and transcript maps of approximately 1 Mb around loop-tail, a gene for severe neural tube defects on distal mouse chromosome 1 and human chromosome 1q22-q23. *Genomics* 72 (2), pp.180-192.

Drosophila 12 Genomes, C. *et al.* (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature* 450 (7167), pp.203-218.

DuBois, K. N. *et al.* (2012). NUP-1 Is a large coiled-coil nucleoskeletal protein in trypanosomes with lamin-like functions. *PLoS Biol* 10 (3), p.e1001287.

Dutcher, S. K., Huang, B. and Luck, D. J. (1984). Genetic dissection of the central pair microtubules of the flagella of *Chlamydomonas reinhardtii*. *J Cell Biol* 98 (1), pp.229-236.

Dutcher, S. K. *et al.* (2002). Epsilon-tubulin is an essential component of the centriole. *Mol Biol Cell* 13 (11), pp.3859-3869.

Dutcher, S. K. and Trabuco, E. C. (1998). The UNI3 gene is required for assembly of basal bodies of Chlamydomonas and encodes delta-tubulin, a new member of the tubulin superfamily. *Mol Biol Cell* 9 (6), pp.1293-1308.

Eichinger, L. *et al.* (2005). The genome of the social amoeba *Dictyostelium discoideum*. *Nature* 435 (7038), pp.43-57.

Eisen, J. A. *et al.* (2006). Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol* 4 (9), p.e286.

El-Sayed, N. M. *et al.* (2005). The genome sequence of *Trypanosoma cruzi,* etiologic agent of Chagas disease. *Science* 309 (5733), pp.409-415.

Engstler, M. *et al.* (2007). Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell* 131 (3), pp.505-515.

Erickson, H. P. (1995). FtsZ, a prokaryotic homolog of tubulin? *Cell* 80 (3), pp.367-370.

Erickson, H. P. (1997). FtsZ, a tubulin homologue in prokaryote cell division. *Trends Cell Biol* 7 (9), pp.362-367.

Errabolu, R., Sanders, M. A. and Salisbury, J. L. (1994). Cloning of a cDNA encoding human centrin, an EF-hand protein of centrosomes and mitotic spindle poles. *J Cell Sci* 107 (Pt 1), pp.9-16.

Esson, H. J. *et al.* (2012). Morphology of the trypanosome bilobe, a novel cytoskeletal structure. *Eukaryot Cell* 11 (6), pp.761-772.

Fassan, M. *et al.* (2011). Mitostatin is down-regulated in human prostate cancer and suppresses the invasive phenotype of prostate cancer cells. *PLoS One* 6 (5), p.e19771.

Findeisen, P. *et al.* (2014). Six subgroups and extensive recent duplications characterize the evolution of the eukaryotic tubulin protein family. *Genome Biol Evol* 6 (9), pp.2274-2288.

Finn, R. D. *et al.* (2014). Pfam: the protein families database. *Nucleic Acids Res* 42 (Database issue), pp.D222-230.

Firat-Karalar, E. N. *et al.* (2014). Proteomic analysis of mammalian sperm cells identifies new components of the centrosome. *J Cell Sci*.

Fliegauf, M. *et al.* (2006). Nephrocystin specifically localizes to the transition zone of renal and respiratory cilia and photoreceptor connecting cilia. *J Am Soc Nephrol* 17 (9), pp.2424-2433.

Flores, E. *et al.* (2010). Cryopreservation-induced alterations in boar spermatozoa mitochondrial function are related to changes in the expression and location of midpiece mitofusin-2 and actin network. *Theriogenology* 74 (3), pp.354-363.

Fong, C. S. *et al.* (2014). SAS-6 assembly templated by the lumen of cartwheel-less centrioles precedes centriole duplication. *Dev Cell* 30 (2), pp.238-245.

Francia, M. E. and Striepen, B. (2014). Cell division in apicomplexan parasites. *Nat Rev Micro* 12 (2), pp.125-136.

Franzen, O. *et al.* (2009). Draft genome sequencing of *Giardia intestinalis* assemblage B isolate GS: is human giardiasis caused by two different species? *PLoS Pathog* 5 (8), p.e1000560.

Fritz-Laylin, L. K. and Cande, W. Z. (2010). Ancestral centriole and flagella proteins identified by analysis of Naegleria differentiation. *J Cell Sci* 123 (Pt 23), pp.4024-4031.

Fritz-Laylin, L. K. *et al.* (2010). The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* 140 (5), pp.631-642.

Fry, A. M. (2002). The Nek2 protein kinase: a novel regulator of centrosome structure. *Oncogene* 21 (40), pp.6184-6194.

Fry, A. M. *et al.* (1998). C-Nap1, a novel centrosomal coiled-coil protein and candidate substrate of the cell cycle-regulated protein kinase Nek2. *J Cell Biol* 141 (7), pp.1563-1574.

Fu, J., Hagan, I. M. and Glover, D. M. (2015). The centrosome and its duplication cycle. *Cold Spring Harb Perspect Biol* 7 (2).

Fuge, H. (1969). Electron Microscopic Studies on the Intra-flagellar Structures of Trypanosomes*. *Journal of Eukaryotic Microbiology* 16 (3), pp.460-466.

Furukawa, K. *et al.* (1994). cDNA cloning and functional characterization of a meiosis-specific protein (MNS1) with apparent nuclear association. *Chromosome Res* 2 (2), pp.99-113.

Gache, V. *et al.* (2010). *Xenopus* meiotic microtubule-associated interactome. *PLoS One* 5 (2), p.e9248.

Gadelha, C. *et al.* (2009). Membrane domains and flagellar pocket boundaries are influenced by the cytoskeleton in African trypanosomes. *Proc Natl Acad Sci U S A* 106 (41), pp.17425-17430.

Gadelha, C., Wickstead, B. and Gull, K. (2007). Flagellar and ciliary beating in trypanosome motility. *Cell Motil Cytoskeleton* 64 (8), pp.629-643.

Gadelha, C. *et al.* (2006). Basal body and flagellum mutants reveal a rotational constraint of the central pair microtubules in the axonemes of trypanosomes. *J Cell Sci* 119 (Pt 12), pp.2405-2413.

Gale, M., Jr., Carter, V. and Parsons, M. (1994). Translational control mediates the developmental regulation of the *Trypanosoma brucei* Nrk protein kinase. *J Biol Chem* 269 (50), pp.31659-31665.

Gale, M., Jr. and Parsons, M. (1993). A *Trypanosoma brucei* gene family encoding protein kinases with catalytic domains structurally related to Nek1 and NIMA. *Mol Biochem Parasitol* 59 (1), pp.111-121.

Gallo, J. M. and Anderton, B. H. (1983). A subpopulation of trypanosome microtubules recognized by a monoclonal antibody to tubulin. *EMBO J* 2 (4), pp.479-483.

Gallo, J. M., Precigout, E. and Schrevel, J. (1988). Subcellular sequestration of an antigenically unique beta-tubulin. *Cell Motil Cytoskeleton* 9 (2), pp.175-183.

Gallo, J. M. and Schrevel, J. (1985). Homologies between paraflagellar rod proteins from trypanosomes and euglenoids revealed by a monoclonal antibody. *Eur J Cell Biol* 36 (2), pp.163-168.

Garcia-Gonzalo, F. R. and Reiter, J. F. (2012). Scoring a backstage pass: Mechanisms of ciliogenesis and ciliary access. *J Cell Biol* 197 (6), pp.697-709.

Garcia-Salcedo, J. A. *et al.* (2004). A differential role for actin during the life cycle of *Trypanosoma brucei*. *EMBO J* 23 (4), pp.780-789.

Gardner, M. J. *et al.* (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419 (6906), pp.498-511.

Garreau de Loubresse, N. *et al.* (2001). Role of delta-tubulin and the C-tubule in assembly of *Paramecium* basal bodies. *BMC Cell Biol* 2 (4).

Geimer, S. *et al.* (1998). A novel 95-kD protein is located in a linker between cytoplasmic microtubules and basal bodies in a green flagellate and forms striated filaments in vitro. *J Cell Biol* 140 (5), pp.1149-1158.

Geimer, S. and Melkonian, M. (2004). The ultrastructure of the *Chlamydomonas reinhardtii* basal apparatus: identification of an early marker of radial asymmetry inherent in the basal body. *J Cell Sci* 117 (Pt 13), pp.2663-2674.

Geimer, S. and Melkonian, M. (2005). Centrin scaffold in *Chlamydomonas reinhardtii* revealed by immunoelectron microscopy. *Eukaryot Cell* 4 (7), pp.1253-1263.

Ghossoub, R. *et al.* (2011). The ciliary pocket: a once-forgotten membrane domain at the base of cilia. *Biol Cell* 103 (3), pp.131-144.

Gibbs, R. A. *et al.* (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428 (6982), pp.493-521.

Giessl, A. *et al.* (2006). Centrins, gatekeepers for the light-dependent translocation of transducin through the photoreceptor cell connecting cilium. *Vision Res* 46 (27), pp.4502-4509.

Ginger, M. L. *et al.* (2013). Calmodulin is Required for Paraflagellar Rod Assembly and Flagellum-Cell Body Attachment in Trypanosomes. *Protist* 164 (4), pp.528-540.

Gluenz, E. *et al.* (2010). Beyond 9+0: noncanonical axoneme structures characterize sensory cilia from protists to humans. *FASEB J* 24 (9), pp.3117-3121.

Gluenz, E. *et al.* (2011). The kinetoplast duplication cycle in *Trypanosoma brucei* is orchestrated by cytoskeleton-mediated cell morphogenesis. *Mol Cell Biol* 31 (5), pp.1012-1021.

Goetz, S. C., Liem, K. F., Jr. and Anderson, K. V. (2012). The spinocerebellar ataxia-associated gene Tau tubulin kinase 2 controls the initiation of ciliogenesis. *Cell* 151 (4), pp.847-858.

Goffeau, A. *et al.* (1996). Life with 6000 genes. *Science* 274 (5287), pp.546, 563-547.

Gogendeau, D. *et al.* (2008). Functional diversification of centrins and cell morphological complexity. *J Cell Sci* 121 (Pt 1), pp.65-74.

Gottardo, M., Callaini, G. and Riparbelli, M. G. (2013). The cilium-like region of the Drosophila spermatocyte: an emerging flagellum? *J Cell Sci* 126 (Pt 23), pp.5441-5452.

Gould, R. R. (1975). The basal bodies of *Chlamydomonas reinhardtii*. Formation from probasal bodies, isolation, and partial characterization. *J Cell Biol* 65 (1), pp.65-74.

Gould, S. B. *et al.* (2011). Ciliate pellicular proteome identifies novel protein families with characteristic repeat motifs that are common to alveolates. *Mol Biol Evol* 28 (3), pp.1319-1331.

Gout, J. F. *et al.* (2010). The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet* 6 (5), p.e1000944.

Gouy, M., Guindon, S. and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27 (2), pp.221-224.

Graf, R. (2002). DdNek2, the first non-vertebrate homologue of human Nek2, is involved in the formation of microtubule-organizing centers. *J Cell Sci* 115 (Pt 9), pp.1919-1929.

Griffiths, S. *et al.* (2007). RNA interference mutant induction in vivo demonstrates the essential nature of trypanosome flagellar function during mammalian infection. *Eukaryot Cell* 6 (7), pp.1248-1250.

Groenen, M. A. *et al.* (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491 (7424), pp.393-398.

Groh, K. J. *et al.* (2011). Global proteomics analysis of testis and ovary in adult zebrafish (*Danio rerio*). *Fish Physiol Biochem* 37 (3), pp.619-647.

Gu, W. *et al.* (2005). A new EF-hand containing gene EFHC2 on Xp11.4: tentative evidence for association with juvenile myoclonic epilepsy. *Epilepsy Res* 66 (1-3), pp.91-98.

Guichard, P. *et al.* (2010). Procentriole assembly revealed by cryo-electron tomography. *EMBO J* 29 (9), pp.1565-1572.

Guichard, P. *et al.* (2012). Cartwheel Architecture of *Trichonympha* Basal Body. *Science* 337 (6094), p.553.

Gull, K. *et al.* (1990). The cell cycle and cytoskeletal morphogenesis in *Trypanosoma brucei*. *Biochem Soc Trans* 18 (5), pp.720-722.

Gunay-Aygun, M. (2009). Liver and kidney disease in ciliopathies. *Am J Med Genet C Semin Med Genet* 151C (4), pp.296-306.

Haas, B. J. *et al.* (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461 (7262), pp.393-398.

Hagiwara, H., Ohwada, N. and Takata, K. (2004). Cell biology of normal and abnormal ciliogenesis in the ciliated epithelium. *Int Rev Cytol* 234, pp.101-141.

Hall, N. (2012). Genomic insights into the other malaria. *Nat Genet* 44 (9), pp.962-963.

Hammarton, T. C., Engstler, M. and Mottram, J. C. (2004). The *Trypanosoma brucei* cyclin, CYC2, is required for cell cycle progression through G1 phase and for maintenance of procyclic form cell morphology. *J Biol Chem* 279 (23), pp.24757-24764.

Hammarton, T. C. *et al.* (2007). *Trypanosoma brucei* Polo-like kinase is essential for basal body duplication, kDNA segregation and cytokinesis. *Mol Microbiol* 65 (5), pp.1229-1248.

Harper, J. I. *et al.* (2000). A centrin homologue is localised across the developing cell plate in gymnosperms and angiosperms. *Protoplasma* 211 (3-4), pp.207-216.

Harris, M., Taylor, G. and Taylor, J. (2005). *Maths & Stats for the life and medical sciences.* 1st edition ed. Scion.

Hart, S. R. *et al.* (2009). Analysis of the trypanosome flagellar proteome using a combined electron transfer/collisionally activated dissociation strategy. *J Am Soc Mass Spectrom* 20 (2), pp.167-175.

Hassounah, N. B., Bunch, T. A. and McDermott, K. M. (2012). Molecular pathways: the role of primary cilia in cancer progression and therapeutics with a focus on Hedgehog signaling. *Clin Cancer Res* 18 (9), pp.2429-2435.

Haycraft, C. J. *et al.* (2001). The *C. elegans* homolog of the murine cystic kidney disease gene Tg737 functions in a ciliogenic pathway and is disrupted in osm-5 mutant worms. *Development* 128 (9), pp.1493-1505.

Hayes, J. M. *et al.* (2007). Identification of novel ciliogenesis factors using a new in vivo model for mucociliary epithelial development. *Dev Biol* 312 (1), pp.115-130.

He, C. Y., Pypaert, M. and Warren, G. (2005). Golgi duplication in *Trypanosoma brucei* requires Centrin2. *Science* 310 (5751), pp.1196-1198.

Heddergott, N. *et al.* (2012). Trypanosome motion represents an adaptation to the crowded environment of the vertebrate bloodstream. *PLoS Pathog* 8 (11), p.e1003023.

Hellsten, U. *et al.* (2010). The genome of the Western clawed frog *Xenopus tropicalis*. *Science* 328 (5978), pp.633-636.

Hendriks, E. F. *et al.* (2001). A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form. *EMBO J* 20 (23), pp.6700-6711.

Hildebrandt, F., Benzing, T. and Katsanis, N. (2011). Ciliopathies. *N Engl J Med* 364 (16), pp.1533-1543.

Hill, K. L. (2003). Biology and mechanism of trypanosome cell motility. *Eukaryot Cell* 2 (2), pp.200-208.

Hill, K. L. (2010). Parasites in motion: flagellum-driven cell motility in African trypanosomes. *Curr Opin Microbiol* 13 (4), pp.459-465.

Hinchcliffe, E. H. and Linck, R. W. (1998). Two proteins isolated from sea urchin sperm flagella: structural components common to the stable microtubules of axonemes and centrioles. *J Cell Sci* 111 (Pt 5), pp.585-595.

Hirschner, W. *et al.* (2007). Biosynthesis of Wdr16, a marker protein for kinocilia-bearing cells, starts at the time of kinocilia formation in rat, and wdr16 gene knockdown causes hydrocephalus in zebrafish. *J Neurochem* 101 (1), pp.274-288.

Hodges, M. E. *et al.* (2010). Reconstructing the evolutionary history of the centriole from protein components. *J Cell Sci* 123 (Pt 9), pp.1407-1413.

Hodges, M. E. *et al.* (2011). Conservation of ciliary proteins in plants with no cilia. *BMC Plant Biol* 11, p.185.

Hoh, R. A. *et al.* (2012). Transcriptional program of ciliated epithelial cells reveals new cilium and centrosome components and links to human disease. *PLoS One* 7 (12), p.e52166.

Honeybee Genome Sequencing, C. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443 (7114), pp.931-949.

Hoog, J. L. *et al.* (2012). Cryo-electron tomography and 3-D analysis of the intact flagellum in *Trypanosoma brucei*. *J Struct Biol*.

Howe, K. *et al.* (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496 (7446), pp.498-503.

Hoyer-Fender, S. (2010). Centriole maturation and transformation to basal body. *Semin Cell Dev Biol* 21 (2), pp.142-147.

Hu, C. D., Grinberg, A. V. and Kerppola, T. K. (2006). Visualization of protein interactions in living cells using bimolecular fluorescence complementation (BiFC) analysis. *Curr Protoc Cell Biol* Chapter 21, p.Unit 21 23.

Hughes, L. *et al.* (2014). Serial block face scanning electron microscopy-the future of cell ultrastructure imaging. *Protoplasma* 251 (2), pp.395-401.

Hughes, L. *et al.* (2013). A cell-body groove housing the new flagellum tip suggests an adaptation of cellular morphogenesis for parasitism in the bloodstream form of *Trypanosoma brucei*. *J Cell Sci* 126 (Pt 24), pp.5748-5757.

Hughes, L. C. *et al.* (2012). Three-Dimensional Structure of the Trypanosome Flagellum Suggests that the Paraflagellar Rod Functions as a Biomechanical Spring. *PLoS One* 7 (1), p.e25700.

Hutchings, N. R., Donelson, J. E. and Hill, K. L. (2002). Trypanin is a cytoskeletal linker protein and is required for cell motility in African trypanosomes. *J Cell Biol* 156 (5), pp.867-877.

Hyams, J. S. (1982). The Euglena paraflagellar rod: structure, relationship to other flagellar components and preliminary biochemical characterization. *J Cell Sci* 55, pp.199-210.

Ibi, M. *et al.* (2011). Trichoplein controls microtubule anchoring at the centrosome by binding to Odf2 and ninein. *J Cell Sci* 124 (Pt 6), pp.857-864.

Idei, M. *et al.* (2013). Sperm ultrastructure in the diatoms Melosira and Thalassiosira and the significance of the 9 + 0 configuration. *Protoplasma* 250 (4), pp.833-850.

Ikeda, K. *et al.* (2003). Rib72, a conserved protein associated with the ribbon compartment of flagellar A-microtubules and potentially involved in the linkage between outer doublet microtubules. *J Biol Chem* 278 (9), pp.7725-7734.

Ikeda, K. N. and de Graffenried, C. L. (2012). Polo-like kinase is necessary for flagellum inheritance in *Trypanosoma brucei*. *J Cell Sci*.

Ikeda, T. *et al.* (2005). The mouse ortholog of EFHC1 implicated in juvenile myoclonic epilepsy is an axonemal protein widely conserved among organisms with motile cilia and flagella. *FEBS Lett* 579 (3), pp.819-822.

Inaba, K. (2007). Molecular basis of sperm flagellar axonemes: structural and evolutionary aspects. *Ann N Y Acad Sci* 1101, pp.506-526.

Inoko, A. *et al.* (2012). Trichoplein and Aurora A block aberrant primary cilia assembly in proliferating cells. *J Cell Biol* 197 (3), pp.391-405.

International Chicken Genome Sequencing, C. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432 (7018), pp.695-716.

Ishikawa, H. and Marshall, W. F. (2011). Ciliogenesis: building the cell's antenna. *Nat Rev Mol Cell Biol* 12 (4), pp.222-234.

Ishikawa, H. *et al.* (2012a). Proteomic Analysis of Mammalian Primary Cilia. *Current Biology* 22 (5), pp.414-419.

Ishikawa, H. *et al.* (2012b). Proteomic Analysis of Mammalian Primary Cilia. *Curr Biol*.

Itsuki, Y. *et al.* (2008). Molecular cloning of novel Monad binding protein containing tetratricopeptide repeat domains. *FEBS Lett* 582 (16), pp.2365-2370.

Ivens, A. C. *et al.* (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309 (5733), pp.436-442.

Jakobsen, L. *et al.* (2011). Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. *EMBO J* 30 (8), pp.1520-1535.

Jamieson, B. G. M., Dallai, R. and Afzelius, B. A. (1999). *Insects. Their Spermatozoa and Phylogeny.* Science Publishers inc.

Januschke, J. *et al.* (2011). Drosophila neuroblasts retain the daughter centrosome. *Nat Commun* 2, p.243.

Jenkins, C. *et al.* (2002). Genes for the cytoskeletal protein tubulin in the bacterial genus Prosthecobacter. *Proc Natl Acad Sci U S A* 99 (26), pp.17049-17054.

Johnson, K. A. and Rosenbaum, J. L. (1992). Polarity of flagellar assembly in Chlamydomonas. *J Cell Biol* 119 (6), pp.1605-1611.

Joo, K. *et al.* (2013). CCDC41 is required for ciliary vesicle docking to the mother centriole. *Proc Natl Acad Sci U S A* 110 (15), pp.5987-5992.

Jord, A. A. *et al.* (2014). Centriole amplification by mother and daughter centrioles differs in multiciliated cells. *Nature* 516, pp.104-101-107.

Joukov, V., Walter, J. C. and De Nicolo, A. (2014). The Cep192-organized aurora A-Plk1 cascade is essential for centrosome cycle and bipolar spindle assembly. *Mol Cell* 55 (4), pp.578-591.

Judelson, H. S., Shrivastava, J. and Manson, J. (2012). Decay of genes encoding the oomycete flagellar proteome in the downy mildew *Hyaloperonospora arabidopsidis*. *PLoS One* 7 (10), p.e47624.

Kabututu, Z. P. *et al.* (2010). CMF70 is a subunit of the dynein regulatory complex. *J Cell Sci* 123 (Pt 20), pp.3587-3595.

Kalt, A. and Schliwa, M. (1996). A novel structural component of the Dictyostelium centrosome. *J Cell Sci* 109 (Pt 13), pp.3103-3112.

Kasahara, K. *et al.* (2014). Ubiquitin-proteasome system controls ciliogenesis at the initial step of axoneme extension. *Nat Commun* 5 (5081).

Keller, L. C. *et al.* (2005). Proteomic analysis of isolated chlamydomonas centrioles reveals orthologs of ciliary-disease genes. *Curr Biol* 15 (12), pp.1090-1098.

Kelly, S. *et al.* (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* 154 (1), pp.103-109.

Khona, D. K. *et al.* (2013). Anomalies in the motion dynamics of long-flagella mutants of *Chlamydomonas reinhardtii*. *J Biol Phys* 39 (1), pp.1-14.

Kilburn, C. L. *et al.* (2007). New Tetrahymena basal body protein components identify basal body domain structure. *The Journal of Cell Biology* 178 (6), pp.905-912.

Kilmartin, J. V. (2003). Sfi1p has conserved centrin-binding sites and an essential function in budding yeast spindle pole body duplication. *J Cell Biol* 162 (7), pp.1211-1221.

Kilmartin, J. V., Wright, B. and Milstein, C. (1982). Rat monoclonal antitubulin antibodies derived by using a new nonsecreting rat cell line. *J Cell Biol* 93 (3), pp.576-582.

Kim, Y. R. *et al.* (2010). Mutational analysis of MITOSTATIN, a candidate tumor-suppressor gene, at a mononucleotide repeat in gastric and colorectal carcinoma. *Gut Liver* 4 (1), pp.149-150.

Kissinger, J. C. *et al.* (2003). ToxoDB: accessing the *Toxoplasma gondii* genome. *Nucleic Acids Res* 31 (1), pp.234-236.

Kitagawa, D. *et al.* (2011). Structural basis of the 9-fold symmetry of centrioles. *Cell* 144 (3), pp.364-375.

Kleylein-Sohn, J. *et al.* (2007). Plk4-induced centriole biogenesis in human cells. *Dev Cell* 13 (2), pp.190-202.

Klos Dehring, D. A. *et al.* (2013). Deuterosome-mediated centriole biogenesis. *Dev Cell* 27 (1), pp.103-112.

Kobayashi, T. and Dynlacht, B. D. (2011). Regulating the transition from centriole to basal body. *J Cell Biol* 193 (3), pp.435-444.
Kohl, L. and Bastin, P. (2005). The flagellum of trypanosomes. *Int Rev Cytol* 244, pp.227-285.
Kohl, L., Robinson, D. and Bastin, P. (2003). Novel roles for the flagellum in cell morphogenesis and cytokinesis of trypanosomes. *EMBO J* 22 (20), pp.5336-5346.
Kohl, L., Sherwin, T. and Gull, K. (1999). Assembly of the paraflagellar rod and the flagellum attachment zone complex during the *Trypanosoma brucei* cell cycle. *J Eukaryot Microbiol* 46 (2), pp.105-109.
Kollman, J. M. *et al.* (2015). Ring closure activates yeast gammaTuRC for species-specific microtubule nucleation. *Nat Struct Mol Biol* 22 (2), pp.132-137.
Kollman, J. M. *et al.* (2011). Microtubule nucleation by gamma-tubulin complexes. *Nat Rev Mol Cell Biol* 12 (11), pp.709-721.
Kopczak, S. D. *et al.* (1992). The small genome of *Arabidopsis* contains at least six expressed alpha-tubulin genes. *Plant Cell* 4 (5), pp.539-547.
Kozminski, K. G. *et al.* (1993). A motility in the eukaryotic flagellum unrelated to flagellar beating. *Proc Natl Acad Sci U S A* 90 (12), pp.5519-5523.
Kratzerova, L. *et al.* (2001). Cell cycle-dependent changes in localization of a 210-kDa microtubule-interacting protein in Leishmania. *Exp Cell Res* 266 (2), pp.270-278.
Kremer, H. *et al.* (2006). Usher syndrome: molecular links of pathogenesis, proteins and pathways. *Human Molecular Genetics* 15 (suppl 2), pp.R262-R270.
Kuriyama, R. and Borisy, G. G. (1981). Centriole cycle in Chinese hamster ovary cells as determined by whole-mount electron microscopy. *J Cell Biol* 91 (3 Pt 1), pp.814-821.
La Greca, F. and Magez, S. (2011). Vaccination against trypanosomiasis: can it be done or is the trypanosome truly the ultimate immune destroyer and escape artist? *Hum Vaccin* 7 (11), pp.1225-1233.
Lacomble, S. *et al.* (2011). A Trypanosoma brucei Protein Required for Maintenance of the Flagellum Attachment Zone and Flagellar Pocket ER Domains. *Protist*.
Lacomble, S. *et al.* (2009). Three-dimensional cellular architecture of the flagellar pocket and associated cytoskeleton in trypanosomes revealed by electron microscope tomography. *J Cell Sci* 122 (Pt 8), pp.1081-1090.
Lacomble, S. *et al.* (2010). Basal body movements orchestrate membrane organelle division and cell morphogenesis in *Trypanosoma brucei*. *J Cell Sci* 123 (Pt 17), pp.2884-2891.
Landfear, S. M. and Ignatushchenko, M. (2001). The flagellum and flagellar pocket of trypanosomatids. *Mol Biochem Parasitol* 115 (1), pp.1-17.
Lange, B. M. and Gull, K. (1995). A molecular marker for centriole maturation in the mammalian cell cycle. *The Journal of Cell Biology* 130 (4), pp.919-927.
Lange, B. M. and Gull, K. (1996). A structural study of isolated mammalian centrioles using negative staining electron microscopy. *J Struct Biol* 117 (3), pp.222-226.
Langousis, G. and Hill, K. L. (2014). Motility and more: the flagellum of *Trypanosoma brucei*. *Nat Rev Microbiol* 12 (7), pp.505-518.
Laoukili, J. *et al.* (2000). Differential expression and cellular distribution of centrin isoforms during human ciliated cell differentiation in vitro. *J Cell Sci* 113 (Pt 8), pp.1355-1364.
Lauwaet, T. *et al.* (2011). Mining the *Giardia* genome and proteome for conserved and unique basal body proteins. *Int J Parasitol* 41 (10), pp.1079-1092.

Lechtreck, K. F. *et al.* (2008). Mutations in Hydin impair ciliary motility in mice. *J Cell Biol* 180 (3), pp.633-643.

Lechtreck, K. F., Gould, T. J. and Witman, G. B. (2013). Flagellar central pair assembly in *Chlamydomonas reinhardtii*. *Cilia* 2 (15), pp.1-19.

Lechtreck, K. F. and Witman, G. B. (2007). *Chlamydomonas reinhardtii* hydin is a central pair protein required for flagellar motility. *J Cell Biol* 176 (4), pp.473-482.

Lee, S. and Weatherbee, S. D. (2010). Synergistic interaction between ciliary genes reflects the importance of mutational load in ciliopathies. *Journal of the American Society of Nephrology* 21 (5), pp.724-726.

Lehti, M. S., Kotaja, N. and Sironen, A. (2013). KIF3A is essential for sperm tail formation and manchette function. *Mol Cell Endocrinol* 377 (1-2), pp.44-55.

Leidel, S. *et al.* (2005). SAS-6 defines a protein family required for centrosome duplication in *C. elegans* and in human cells. *Nat Cell Biol* 7 (2), pp.115-125.

Leighton, S. B. (1981). SEM images of block faces, cut by a miniature microtome within the SEM - a technical note. *Scan Electron Microsc* (Pt 2), pp.73-76.

Letcher, P. M. *et al.* (2006). Ultrastructural and molecular phylogenetic delineation of a new order, the *Rhizophydiales (Chytridiomycota)*. *Mycological Research* 110 (8), pp.898-915.

Letcher, P. M. *et al.* (2008). Ultrastructural and molecular analyses of *Rhizophydiales (Chytridiomycota)* isolates from North America and Argentina. *Mycological Research* 112 (7), pp.759-782.

Li, H. *et al.* (2008). *Dictyostelium* Aurora kinase has properties of both Aurora A and Aurora B kinases. *Eukaryotic Cell* 7 (5), pp.894-905.

Li, J. B. *et al.* (2004). Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell* 117 (4), pp.541-552.

Li, L., Stoeckert, C. J., Jr. and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13 (9), pp.2178-2189.

Li, Z., Umeyama, T. and Wang, C. C. (2009). The Aurora kinase in *Trypanosoma brucei* plays distinctive roles in metaphase-anaphase transition and cytokinetic initiation. *PLoS Pathog* 5 (9), p.e1000575.

Li, Z., Umeyama, T. and Wang, C. C. (2010). Polo-like kinase guides cytokinesis in *Trypanosoma brucei* through an indirect means. *Eukaryot Cell* 9 (5), pp.705-716.

Li, Z. and Wang, C. C. (2006). Changing roles of aurora-B kinase in two life cycle stages of *Trypanosoma brucei*. *Eukaryot Cell* 5 (7), pp.1026-1035.

Li, Z., Yao, K. and Cao, Y. (1999). Molecular cloning of a novel tissue-specific gene from human nasopharyngeal epithelium. *Gene* 237 (1), pp.235-240.

Libusova, L. and Draber, P. (2006). Multiple tubulin forms in ciliated protozoan *Tetrahymena* and *Paramecium* species. *Protoplasma* 227 (2-4), pp.65-76.

Lin, J. *et al.* (2011). Building blocks of the nexin-dynein regulatory complex in Chlamydomonas flagella. *J Biol Chem* 286 (33), pp.29175-29191.

Linck, R. *et al.* (2014). Insights into the structure and function of ciliary and flagellar doublet microtubules: tektins, Ca2+-binding proteins, and stable protofilaments. *J Biol Chem* 289 (25), pp.17427-17444.

Linck, R. W., Amos, L. A. and Amos, W. B. (1985). Localization of tektin filaments in microtubules of sea urchin sperm flagella by immunoelectron microscopy. *J Cell Biol* 100 (1), pp.126-135.

Linck, R. W. and Stephens, R. E. (2007). Functional protofilament numbering of ciliary, flagellar, and centriolar microtubules. *Cell Motil Cytoskeleton* 64 (7), pp.489-495.

Lindemann, C. B. and Lesich, K. A. (2010). Flagellar and ciliary beating: the proven and the possible. *J Cell Sci* 123 (Pt 4), pp.519-528.

Links, M. G. *et al.* (2011). De novo sequence assembly of *Albugo candida* reveals a small genome relative to other biotrophic oomycetes. *BMC Genomics* 12, p.503.

Liu, B. *et al.* (1994). gamma-tubulin in *Arabidopsis*: gene sequence, immunoblot, and immunofluorescence studies. *Plant Cell* 6 (2), pp.303-314.

Liu, B. *et al.* (2005). Fellowship of the rings: the replication of kinetoplast DNA. *Trends Parasitol* 21 (8), pp.363-369.

Liu, Q. *et al.* (2007). The proteome of the mouse photoreceptor sensory cilium complex. *Mol Cell Proteomics* 6 (8), pp.1299-1317.

Liu, Z. *et al.* (2012). Proteomic features of potential tumor suppressor NESG1 in nasopharyngeal carcinoma. *Proteomics* 12 (22), pp.3416-3425.

Liu, Z. *et al.* (2011a). Decreased expression of updated NESG1 in nasopharyngeal carcinoma: its potential role and preliminarily functional mechanism. *Int J Cancer* 128 (11), pp.2562-2571.

Liu, Z. *et al.* (2011b). Potential tumor suppressor NESG1 as an unfavorable prognosis factor in nasopharyngeal carcinoma. *PLoS One* 6 (11), p.e27887.

Longcore, J. E., Pessier, A. P. and Nichols, D. K. (1999). *Batrachochytrium Dendrobatidis* gen. et sp. nov., a chytrid pathogenic to amphibians. *Mycologia* 91 (2), pp.219-227.

Lozano-Nunez, A. *et al.* (2013). An Analog-Sensitive Approach Identifies Basal Body Rotation and Flagellum Attachment Zone Elongation as Key Functions of PLK in *Trypanosoma brucei*. *Mol Biol Cell* 24 (9), pp.1321-1333.

Lukasiewicz, K. B. *et al.* (2011). Control of centrin stability by Aurora A. *PLoS One* 6 (6), p.e21291.

Lupas, A. (1996). Coiled coils: new structures and new functions. *Trends Biochem Sci* 21 (10), pp.375-382.

Lutz, W. *et al.* (2001). Phosphorylation of centrin during the cell cycle and its role in centriole separation preceding centrosome duplication. *J Biol Chem* 276 (23), pp.20774-20780.

Mahajan, B. *et al.* (2008). Centrins, cell cycle regulation proteins in human malaria parasite *Plasmodium falciparum*. *J Biol Chem* 283 (46), pp.31871-31883.

Mahjoub, M. R., Qasim Rasi, M. and Quarmby, L. M. (2004). A NIMA-related kinase, Fa2p, localizes to a novel site in the proximal cilia of Chlamydomonas and mouse kidney cells. *Mol Biol Cell* 15 (11), pp.5172-5186.

Mahjoub, M. R., Trapp, M. L. and Quarmby, L. M. (2005). NIMA-related kinases defective in murine models of polycystic kidney diseases localize to primary cilia and centrosomes. *J Am Soc Nephrol* 16 (12), pp.3485-3489.

Marchler-Bauer, A. *et al.* (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* 33 (Database issue), pp.D192-196.

Marchler-Bauer, A. *et al.* (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39 (Database issue), pp.D225-229.

Maric, D., Epting, C. L. and Engman, D. M. (2010). Composition and sensory function of the trypanosome flagellar membrane. *Curr Opin Microbiol* 13 (4), pp.466-472.

Marshall, W. F. (2008). Basal bodies platforms for building cilia. *Curr Top Dev Biol* 85, pp.1-22.

Marshall, W. F. and Rosenbaum, J. L. (2001). Intraflagellar transport balances continuous turnover of outer doublet microtubules: implications for flagellar length control. *J Cell Biol* 155 (3), pp.405-414.

Martinez-Sanz, J. *et al.* (2010). Structure, dynamics and thermodynamics of the human centrin 2/hSfi1 complex. *J Mol Biol* 395 (1), pp.191-204.
Martinez-Sanz, J. *et al.* (2006). Binding of human centrin 2 to the centrosomal protein hSfi1. *FEBS J* 273 (19), pp.4504-4515.
Matsui, T. and Bessho, Y. (2012). Left-right asymmetry in zebrafish. *Cell Mol Life Sci* 69 (18), pp.3069-3077.
Matthews, K. R., Sherwin, T. and Gull, K. (1995). Mitochondrial genome repositioning during the differentiation of the African trypanosome between life cycle forms is microtubule mediated. *J Cell Sci* 108 (Pt 6), pp.2231-2239.
Mayer, U. *et al.* (2009). The proteome of rat olfactory sensory cilia. *Proteomics* 9 (2), pp.322-334.
Mayor, T. *et al.* (2000). The centrosomal protein C-Nap1 is required for cell cycle-regulated centrosome cohesion. *J Cell Biol* 151 (4), pp.837-846.
McClintock, T. S. *et al.* (2008). Tissue expression patterns identify mouse cilia genes. *Physiol Genomics* 32 (2), pp.198-206.
McFadden, G. I. *et al.* (1987). Basal body reorientation mediated by a Ca2+-modulated contractile protein. *J Cell Biol* 105 (2), pp.903-912.
McKean, P. G. *et al.* (2003). Gamma-tubulin functions in the nucleation of a discrete subset of microtubules in the eukaryotic flagellum. *Curr Biol* 13 (7), pp.598-602.
McLaughlin, D. J. *et al.* (2009). The search for the fungal tree of life. *Trends Microbiol* 17 (11), pp.488-497.
Medina, M. T. *et al.* (2008). Novel mutations in Myoclonin1/EFHC1 in sporadic and familial juvenile myoclonic epilepsy. *Neurology* 70 (22 Pt 2), pp.2137-2144.
Meijering, E. *et al.* (2004). Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry A* 58 (2), pp.167-176.
Merchant, S. S. *et al.* (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318 (5848), pp.245-250.
Mita, K. *et al.* (2004). The genome sequence of silkworm, *Bombyx mori*. *DNA Res* 11 (1), pp.27-35.
Mitchell, D. R. (2004). Speculations on the evolution of 9+2 organelles and the role of central pair microtubules. *Biol Cell* 96 (9), pp.691-696.
Mitchell, D. R. (2009). Chapter 8 - The Flagellar Central Pair Apparatus. In: Witman, E. H. H. B. S. B. (ed.) *The Chlamydomonas Sourcebook (Second Edition)*. London: Academic Press, pp.235-252.
Mitchell, D. R. and Nakatsugawa, M. (2004). Bend propagation drives central pair rotation in *Chlamydomonas reinhardtii* flagella. *J Cell Biol* 166 (5), pp.709-715.
Mitchell, D. R. and Sale, W. S. (1999). Characterization of a *Chlamydomonas* insertional mutant that disrupts flagellar central pair microtubule-associated structures. *J Cell Biol* 144 (2), pp.293-304.
Mitchell, D. R. and Smith, B. (2009). Chapter 13 - Analysis of the Central Pair Microtubule Complex in *Chlamydomonas reinhardtii*. In: Stephen, M. K. and Gregory, J. P. (eds.) *Methods in Cell Biology*. Vol. Volume 92. Academic Press, pp.197-213.
Mitchison, T. and Kirschner, M. (1984). Dynamic instability of microtubule growth. *Nature* 312 (5991), pp.237-242.
Mockel, A. *et al.* (2011). Retinal dystrophy in Bardet-Biedl syndrome and related syndromic ciliopathies. *Prog Retin Eye Res* 30 (4), pp.258-274.
Molla-Herman, A. *et al.* (2010). The ciliary pocket: an endocytic membrane domain at the base of primary and motile cilia. *J Cell Sci* 123 (Pt 10), pp.1785-1795.

Moreira-Leite, F. F. *et al.* (2001). A trypanosome structure involved in transmitting cytoplasmic information during cell division. *Science* 294 (5542), pp.610-612.
Morgan, G. W. *et al.* (2005). An evolutionarily conserved coiled-coil protein implicated in polycystic kidney disease is involved in basal body duplication and flagellar biogenesis in *Trypanosoma brucei*. *Molecular and Cellular Biology* 25 (9), pp.3774-3783.
Morita, Y. and Shingyoji, C. (2004). Effects of imposed bending on microtubule sliding in sperm flagella. *Curr Biol* 14 (23), pp.2113-2118.
Morrissette, N. S. and Sibley, L. D. (2002). Cytoskeleton of apicomplexan parasites. *Microbiol Mol Biol Rev* 66 (1), pp.21-38.
Morriswood, B. *et al.* (2013). Novel bilobe components in *Trypanosoma brucei* identified using proximity-dependent biotinylation. *Eukaryot Cell* 12 (2), pp.356-367.
Morriswood, B. *et al.* (2009). The bilobe structure of *Trypanosoma brucei* contains a MORN-repeat protein. *Mol Biochem Parasitol* 167 (2), pp.95-103.
Mouse Genome Sequencing, C. *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420 (6915), pp.520-562.
Muller, H. *et al.* (2010). Proteomic and functional analysis of the mitotic *Drosophila* centrosome. *EMBO J* 29 (19), pp.3344-3357.
Nakachi, M. *et al.* (2011). Proteomic profiling reveals compartment-specific, novel functions of ascidian sperm proteins. *Mol Reprod Dev* 78 (7), pp.529-549.
Nakazawa, Y. *et al.* (2007). SAS-6 is a cartwheel protein that establishes the 9-fold symmetry of the centriole. *Curr Biol* 17 (24), pp.2169-2174.
Narasimhan, V. *et al.* (2014). Mutations in CCDC11, which encodes a coiled-coil containing ciliary crotein, causes *situs inversus* due to dysmotility of monocilia in the left-right organizer. *Hum Mutat* 36 (3), pp.307-318.
Narita, K. *et al.* (2012). Proteomic analysis of multiple primary cilia reveals a novel mode of ciliary development in mammals. *Biol Open* 1 (8), pp.815-825.
Nett, I. R. *et al.* (2009). Identification and specific localization of tyrosine-phosphorylated proteins in *Trypanosoma brucei*. *Eukaryot Cell* 8 (4), pp.617-626.
Nielsen, M. G. *et al.* (2001). Axoneme-specific beta-tubulin specialization: a conserved C-terminal motif specifies the central pair. *Curr Biol* 11 (7), pp.529-533.
Nigg, E. A. (2007). Centrosome duplication: of rules and licenses. *Trends Cell Biol* 17 (5), pp.215-221.
Nigg, E. A. and Stearns, T. (2011). The centrosome cycle: Centriole biogenesis, duplication and inherent asymmetries. *Nat Cell Biol* 13 (10), pp.1154-1160.
Nishizawa, M. *et al.* (2005). Identification of trichoplein, a novel keratin filament-binding protein. *J Cell Sci* 118 (Pt 5), pp.1081-1090.
Nojima, D., Linck, R. W. and Egelman, E. H. (1995). At least one of the protofilaments in flagellar microtubules is not composed of tubulin. *Current Biology* 5 (2), pp.158-167.
Norrander, J. M. *et al.* (2000). The Rib43a protein is associated with forming the specialized protofilament ribbons of flagellar microtubules in Chlamydomonas. *Mol Biol Cell* 11 (1), pp.201-215.
Novak, Z. A. *et al.* (2014). Asterless licenses daughter centrioles to duplicate for the first time in Drosophila embryos. *Curr Biol* 24 (11), pp.1276-1282.
Nynca, J. *et al.* (2014a). Proteomic identification of rainbow trout sperm proteins. *Proteomics* 14 (12), pp.1569-1573.
Nynca, J. *et al.* (2014b). Proteomic identification of rainbow trout seminal plasma proteins. *Proteomics* 14 (1), pp.133-140.

O'Connell, M. J., Krien, M. J. and Hunter, T. (2003). Never say never. The NIMA-related protein kinases in mitotic control. *Trends Cell Biol* 13 (5), pp.221-228.

O'Toole, E. T. *et al.* (2003). Three-dimensional organization of basal bodies from wild-type and delta-tubulin deletion strains of *Chlamydomonas reinhardtii*. *Mol Biol Cell* 14 (7), pp.2999-3012.

Oakley, C. E. and Oakley, B. R. (1989). Identification of gamma-tubulin, a new member of the tubulin superfamily encoded by mipA gene of *Aspergillus nidulans*. *Nature* 338 (6217), pp.662-664.

Oberholzer, M. *et al.* (2011). Independent analysis of the flagellum surface and matrix proteomes provides insight into flagellum signaling in mammalian-infectious *Trypanosoma brucei*. *Mol Cell Proteomics* 10 (10), p.M111 010538.

Oberholzer, M. *et al.* (2006). A vector series for rapid PCR-mediated C-terminal in situ tagging of *Trypanosoma brucei* genes. *Mol Biochem Parasitol* 145 (1), pp.117-120.

Ogbadoyi, E. *et al.* (2000). Architecture of the *Trypanosoma brucei* nucleus during interphase and mitosis. *Chromosoma* 108 (8), pp.501-513.

Ogbadoyi, E. O., Robinson, D. R. and Gull, K. (2003). A high-order trans-membrane structural linkage is responsible for mitochondrial genome positioning and segregation by flagellar basal bodies in trypanosomes. *Mol Biol Cell* 14 (5), pp.1769-1779.

Ohkura, H., Hagan, I. M. and Glover, D. M. (1995). The conserved *Schizosaccharomyces pombe* kinase plo1, required to form a bipolar spindle, the actin ring, and septum, can drive septum formation in G1 and G2 cells. *Genes Dev* 9 (9), pp.1059-1073.

Olbrich, H. *et al.* (2012). Recessive HYDIN mutations cause primary ciliary dyskinesia without randomization of left-right body asymmetry. *Am J Hum Genet* 91 (4), pp.672-684.

Omoto, C. K. *et al.* (1999). Rotation of the central pair microtubules in eukaryotic flagella. *Mol Biol Cell* 10 (1), pp.1-4.

Omoto, C. K. and Kung, C. (1980). Rotation and twist of the central-pair microtubules in the cilia of Paramecium. *J Cell Biol* 87 (1), pp.33-46.

Osmani, S. A., Pu, R. T. and Morris, N. R. (1988). Mitotic induction and maintenance by overexpression of a G2-specific gene that encodes a potential protein kinase. *Cell* 53 (2), pp.237-244.

Ostrowski, L. E. *et al.* (2002). A proteomic analysis of human cilia: identification of novel components. *Mol Cell Proteomics* 1 (6), pp.451-465.

Ostrowski, L. E., Dutcher, S. K. and Lo, C. W. (2011). Cilia and models for studying structure and function. *Proc Am Thorac Soc* 8 (5), pp.423-429.

Ou, Y. Y. *et al.* (2002). CEP110 and ninein are located in a specific domain of the centrosome associated with centrosome maturation. *J Cell Sci* 115 (Pt 9), pp.1825-1835.

Page, A. M. and Lagnado, J. R. (1998). Novel filamentous bundles in the cytoplasm of a unicellular eukaryote *Crithidia fasciculata*. *Protoplasma* 201 (1), pp.64-70.

Page, A. M. and Lagnado, J. R. (2000). Ultrastructural analysis of Phytomonas species from *Euphorbia pinea* reveals trans-cytoplasmic filaments 10 nm in diameter. *Parasitol Res* 86 (10), pp.843-850.

Pain, A. *et al.* (2005). Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. *Science* 309 (5731), pp.131-133.

Paintrand, M. *et al.* (1992). Centrosome organization and centriole architecture: their sensitivity to divalent cations. *J Struct Biol* 108 (2), pp.107-128.

Pampaloni, F. and Florin, E. L. (2008). Microtubule architecture: inspiration for novel carbon nanotube-based biomimetic materials. *Trends Biotechnol* 26 (6), pp.302-310.

Papon, J. F. *et al.* (2010). A 20-year experience of electron microscopy in the diagnosis of primary ciliary dyskinesia. *Eur Respir J* 35 (5), pp.1057-1063.

Paridaen, J. T., Wilsch-Brauninger, M. and Huttner, W. B. (2013). Asymmetric inheritance of centrosome-associated primary cilium membrane directs ciliogenesis after cell division. *Cell* 155 (2), pp.333-344.

Parsons, M. *et al.* (2005). Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major, Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics* 6, p.127.

Pazour, G. J. *et al.* (2005). Proteomic analysis of a eukaryotic cilium. *J Cell Biol* 170 (1), pp.103-113.

Pazour, G. J. *et al.* (2000). *Chlamydomonas* IFT88 and its mouse homologue, polycystic kidney disease gene tg737, are required for assembly of cilia and flagella. *J Cell Biol* 151 (3), pp.709-718.

Peddie, C. J. and Collinson, L. M. (2014). Exploring the third dimension: volume electron microscopy comes of age. *Micron* 61, pp.9-19.

Pelletier, L. *et al.* (2006). Centriole assembly in *Caenorhabditis elegans*. *Nature* 444 (7119), pp.619-623.

Pennekamp, P. *et al.* (2015). *Situs inversus* and ciliary abnormalities: 20 years later, what is the connection? *Cilia* 4 (1), p.1.

Perles, Z. *et al.* (2012). A human laterality disorder associated with recessive CCDC11 mutation. *J Med Genet* 49 (6), pp.386-390.

Pigino, G. *et al.* (2009). Electron-tomographic analysis of intraflagellar transport particle trains in situ. *J Cell Biol* 187 (1), pp.135-148.

Pigino, G. *et al.* (2012). Comparative structural analysis of eukaryotic flagella and cilia from *Chlamydomonas, Tetrahymena*, and sea urchins. *J Struct Biol* 178 (2), pp.199-206.

Pilhofer, M. *et al.* (2011). Microtubules in bacteria: Ancient tubulins build a five-protofilament homolog of the eukaryotic cytoskeleton. *PLoS Biol* 9 (12), p.e1001213.

Piperno, G. *et al.* (1994). Mutations in the "dynein regulatory complex" alter the ATP-insensitive binding sites for inner arm dyneins in *Chlamydomonas* axonemes. *J Cell Biol* 125 (5), pp.1109-1117.

Plattner, H. (2010). How to design a highly organized cell: an unexpectedly high number of widely diversified SNARE proteins positioned at strategic sites in the ciliate, *Paramecium tetraurelia*. *Protist* 161 (4), pp.497-516.

Plotnikova, O. V., Pugacheva, E. N. and Golemis, E. A. (2009). Primary cilia and the cell cycle *Methods Cell Biol*. Vol. 94. pp.137-160.

Ploubidou, A. *et al.* (1999). Evidence for novel cell cycle checkpoints in trypanosomes: kinetoplast segregation and cytokinesis in the absence of mitosis. *J Cell Sci* 112 (Pt 24), pp.4641-4650.

Poon, S. K. *et al.* (2012). A modular and optimized single marker system for generating *Trypanosoma brucei* cell lines expressing T7 RNA polymerase and the tetracycline repressor. *Open Biology* 2 (2).

Pradel, L. C. *et al.* (2006). NIMA-related kinase TbNRKC is involved in basal body separation in *Trypanosoma brucei*. *J Cell Sci* 119 (Pt 9), pp.1852-1863.

Price, H. P. *et al.* (2012). The orthologue of Sjogren's syndrome nuclear autoantigen 1 (SSNA1) in *Trypanosoma brucei* is an immunogenic self-assembling molecule. *PLoS One* 7 (2), p.e31842.

Prigent, C., Glover, D. M. and Giet, R. (2005). Drosophila Nek2 protein kinase knockdown leads to centrosome maturation defects while overexpression causes centrosome fragmentation and cytokinesis failure. *Exp Cell Res* 303 (1), pp.1-13.

Prochnik, S. E. *et al.* (2010). Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* 329 (5988), pp.223-226.

Punta, M. *et al.* (2012). The Pfam protein families database. *Nucleic Acids Res* 40 (Database issue), pp.D290-301.

Quarmby, L. M. and Mahjoub, M. R. (2005). Caught Nek-ing: cilia and centrioles. *J Cell Sci* 118 (Pt 22), pp.5161-5169.

Ralston, K. S. and Hill, K. L. (2006). Trypanin, a component of the flagellar dynein regulatory complex, is essential in bloodstream form African trypanosomes. *PLoS Pathog* 2 (9), p.e101.

Ralston, K. S. *et al.* (2006). Flagellar motility contributes to cytokinesis in *Trypanosoma brucei* and is modulated by an evolutionarily conserved dynein regulatory system. *Eukaryot Cell* 5 (4), pp.696-711.

Reboutier, D. *et al.* (2013). Aurora A is involved in central spindle assembly through phosphorylation of Ser 19 in P150Glued. *The Journal of Cell Biology* 201 (1), pp.65-79.

Redmond, S., Vadivelu, J. and Field, M. C. (2003). RNAit: an automated web-based tool for the selection of RNAi targets in *Trypanosoma brucei*. *Mol Biochem Parasitol* 128 (1), pp.115-118.

Reid, A. J. *et al.* (2012). Comparative genomics of the apicomplexan parasites *Toxoplasma gondii* and *Neospora caninum*: Coccidia differing in host range and transmission strategy. *PLoS Pathog* 8 (3), p.e1002567.

Reininger, L. *et al.* (2011). An essential Aurora-related kinase transiently associates with spindle pole bodies during *Plasmodium falciparum* erythrocytic schizogony. *Molecular Microbiology* 79 (1), pp.205-221.

Renaud, F. L. and Swift, H. (1964). The development of basal bodies and flagella in *Allomyces Arbusculus*. *J Cell Biol* 23, pp.339-354.

Rhesus Macaque Genome, S. *et al.* (2007). Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316 (5822), pp.222-234.

Rice, P., Longden, I. and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16 (6), pp.276-277.

Rider, S. D. and Zhu, G. (2008). *Cryptosporidium* spp. In: Khan, N. A. (ed.) *Emerging protozoan pathogens*. Taylor & Francis, pp.193-226.

Ringo, D. L. (1967). Flagellar motion and fine structure of the flagellar apparatus in *Chlamydomonas*. *The Journal of Cell Biology* 33 (3), pp.543-571.

Riparbelli, M. G. *et al.* (2013). Unique properties of Drosophila spermatocyte primary cilia. *Biol Open* 2 (11), pp.1137-1147.

Robinson, D. R. and Gull, K. (1991). Basal body movements as a mechanism for mitochondrial genome segregation in the trypanosome cell cycle. *Nature* 352 (6337), pp.731-733.

Robinson, D. R. *et al.* (1995). Microtubule polarity and dynamics in the control of organelle positioning, segregation, and cytokinesis in the trypanosome cell cycle. *J Cell Biol* 128 (6), pp.1163-1172.

Roditi, I. *et al.* (1998). Unravelling the procyclin coat of *Trypanosoma brucei*. *Mol Biochem Parasitol* 91 (1), pp.117-130.

Rodrigues-Martins, A. *et al.* (2007). DSAS-6 organizes a tube-like centriole precursor, and its absence suggests modularity in centriole assembly. *Curr Biol* 17 (17), pp.1465-1472.

Rosenblum, E. B. *et al.* (2013). Complex history of the amphibian-killing chytrid fungus revealed with genome resequencing data. *Proc Natl Acad Sci U S A* 110 (23), pp.9385-9390.

Ross, A. J. *et al.* (2007). Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. *Am J Respir Cell Mol Biol* 37 (2), pp.169-185.

Ross, I. *et al.* (2013). epsilon-tubulin is essential in *Tetrahymena thermophila* for the assembly and stability of basal bodies. *J Cell Sci* 126 (15), pp.3441-3451.

Rossetto, M. G. *et al.* (2011). Defhc1.1, a homologue of the juvenile myoclonic gene EFHC1, modulates architecture and basal activity of the neuromuscular junction in Drosophila. *Hum Mol Genet* 20 (21), pp.4248-4257.

Rotureau, B. *et al.* (2014). Flagellar adhesion in *Trypanosoma brucei* relies on interactions between different skeletal structures in the flagellum and cell body. *J Cell Sci* 127 (Pt 1), pp.204-215.

Roux, K. J. *et al.* (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of Cell Biology* 196 (6), pp.801-810.

Roy, A., Kucukural, A. and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5 (4), pp.725-738.

Rual, J. F. *et al.* (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437 (7062), pp.1173-1178.

Ruiz, F. *et al.* (2000). The SM19 gene, required for duplication of basal bodies in Paramecium, encodes a novel tubulin, eta-tubulin. *Curr Biol* 10 (22), pp.1451-1454.

Russell, D. G. and Burns, R. G. (1984). The polar ring of coccidian sporozoites: a unique microtubule-organizing centre. *J Cell Sci* 65, pp.193-207.

Saeki, M. *et al.* (2006). Monad, a WD40 repeat protein, promotes apoptosis induced by TNF-alpha. *Biochem Biophys Res Commun* 342 (2), pp.568-572.

Salisbury, J. L. (1995). Centrin, centrosomes, and mitotic spindle poles. *Curr Opin Cell Biol* 7 (1), pp.39-45.

Salisbury, J. L. (2007). A mechanistic view on the evolutionary origin for centrin-based control of centriole duplication. *J Cell Physiol* 213 (2), pp.420-428.

Salisbury, J. L. *et al.* (1984). Striated flagellar roots: isolation and partial characterization of a calcium-modulated contractile organelle. *The Journal of Cell Biology* 99 (3), pp.962-970.

Salisbury, J. L., Baron, A. T. and Sanders, M. A. (1988). The centrin-based cytoskeleton of *Chlamydomonas reinhardtii*: distribution in interphase and mitotic cells. *J Cell Biol* 107 (2), pp.635-641.

Salisbury, J. L. *et al.* (2002). Centrin-2 is required for centriole duplication in mammalian cells. *Curr Biol* 12 (15), pp.1287-1292.

Sanders, M. A. and Salisbury, J. L. (1994). Centrin plays an essential role in microtubule severing during flagellar excision in *Chlamydomonas reinhardtii*. *J Cell Biol* 124 (5), pp.795-805.

Sanggaard, K. W. *et al.* (2014). Spider genomes provide insight into composition and evolution of venom and silk. *Nat Commun* 5, p.3765.

Sarvetnick, N. *et al.* (1989). A mouse chromosome 17 gene encodes a testes-specific transcript with unusual properties. *Immunogenetics* 30 (1), pp.34-41.

Sauer, G. *et al.* (2005). Proteome analysis of the human mitotic spindle. *Mol Cell Proteomics* 4 (1), pp.35-43.

Schlieper, D. *et al.* (2005). Structure of bacterial tubulin BtubA/B: evidence for horizontal gene transfer. *Proc Natl Acad Sci U S A* 102 (26), pp.9170-9175.

Schmidt, K. N. *et al.* (2012). Cep164 mediates vesicular docking to the mother centriole during early steps of ciliogenesis. *J Cell Biol* 199 (7), pp.1083-1101.

Schmidt, T. *et al.* (2000). A novel protein (Fbf-1) that binds to CD95/APO-1/FAS and shows sequence similarity to trichohyalin and plectin. *Biochim Biophys Acta* 1493 (1-2), pp.249-254.

Schnarwiler, F. *et al.* (2014). Trypanosomal TAC40 constitutes a novel subclass of mitochondrial beta-barrel proteins specialized in mitochondrial genome inheritance. *Proc Natl Acad Sci U S A* 111 (21), pp.7624-7629.

Schneider, A. *et al.* (1987). Subpellicular and flagellar microtubules of *Trypanosoma brucei brucei* contain the same alpha-tubulin isoforms. *J Cell Biol* 104 (3), pp.431-438.

Scholey, J. M. (2008). Intraflagellar transport motors in cilia: moving along the cell's antenna. *J Cell Biol* 180 (1), pp.23-29.

Scott, V., Sherwin, T. and Gull, K. (1997). gamma-tubulin in trypanosomes: molecular characterisation and localisation to multiple and diverse microtubule organising centres. *J Cell Sci* 110 (Pt 2), pp.157-168.

Seebeck, T. *et al.* (1988). The cytoskeleton of *Trypanosoma brucei*-the beauty of simplicity. *Protoplasma* 145 (2-3), pp.188-194.

Selvapandiyan, A. *et al.* (2004). Centrin gene disruption impairs stage-specific basal body duplication and cell cycle progression in *Leishmania*. *J Biol Chem* 279 (24), pp.25703-25710.

Selvapandiyan, A. *et al.* (2001). Expression of a mutant form of *Leishmania donovani* centrin reduces the growth of the parasite. *J Biol Chem* 276 (46), pp.43253-43261.

Selvapandiyan, A. *et al.* (2007). Centrin1 is required for organelle segregation and cytokinesis in *Trypanosoma brucei*. *Mol Biol Cell* 18 (9), pp.3290-3301.

Selvapandiyan, A. *et al.* (2012). Role of centrins 2 and 3 in organelle segregation and cytokinesis in *Trypanosoma brucei*. *PLoS One* 7 (9), p.e45288.

Serluca, F. C. *et al.* (2009). Mutations in zebrafish leucine-rich repeat-containing six-like affect cilia motility and result in pronephric cysts, but have variable effects on left-right patterning. *Development* 136 (10), pp.1621-1631.

Shen, S. *et al.* (2001). *In vivo* epitope tagging of *Trypanosoma brucei* genes using a one step PCR-based strategy. *Mol Biochem Parasitol* 113 (1), pp.171-173.

Sherwin, T. and Gull, K. (1989). The cell division cycle of *Trypanosoma brucei brucei*: timing of event markers and cytoskeletal modulations. *Philos Trans R Soc Lond B Biol Sci* 323 (1218), pp.573-588.

Shi, D. *et al.* (2011). Analysis of ciliary beat frequency and ovum transport ability in the mouse oviduct. *Genes Cells* 16 (3), pp.282-290.

Shi, J. *et al.* (2008). Centrin4 coordinates cell and nuclear division in *T. brucei*. *J Cell Sci* 121 (Pt 18), pp.3062-3070.

Siman-Tov, M. M., Ivens, A. C. and Jaffe, C. L. (2001). Identification and cloning of Lmairk, a member of the Aurora/Ipl1p protein kinase family, from the human protozoan parasite Leishmania. *Biochim Biophys Acta* 1519 (3), pp.241-245.

Simon, J. R. and Salmon, E. D. (1990). The structure of microtubule ends during the elongation and shortening phases of dynamic instability examined by negative-stain electron microscopy. *J Cell Sci* 96 (Pt 4), pp.571-582.

Sinden, R. E. *et al.* (1978). Gametocyte and gamete development in *Plasmodium falciparum*. *Proc R Soc Lond B Biol Sci* 201 (1145), pp.375-399.

Skerget, S. *et al.* (2013). The Rhesus macaque (*Macaca mulatta*) sperm proteome. *Mol Cell Proteomics* 12 (11), pp.3052-3067.

Sluder, G. (2005). Centrosome duplication and its regulation in the higher animal cell *Centrosomes in Development and Disease*. Wiley-VCH Verlag GmbH & Co. KGaA, pp.167-189.

Smith, D. and Price, H. P. (2013). *T. brucei TREU927 total flagellar proteome from bloostream form.* Online release: TriTrypDB 5.0.

Smith, E. F. (2002). Regulation of flagellar dynein by the axonemal central apparatus. *Cell Motil Cytoskeleton* 52 (1), pp.33-42.

Smith, E. F. and Lefebvre, P. A. (1996). PF16 encodes a protein with armadillo repeats and localizes to a single microtubule of the central apparatus in *Chlamydomonas* flagella. *J Cell Biol* 132 (3), pp.359-370.

Smith, E. F. and Lefebvre, P. A. (1997a). PF20 gene product contains WD repeats and localizes to the intermicrotubule bridges in *Chlamydomonas* flagella. *Mol Biol Cell* 8 (3), pp.455-467.

Smith, E. F. and Lefebvre, P. A. (1997b). The role of central apparatus components in flagellar motility and microtubule assembly. *Cell Motil Cytoskeleton* 38 (1), pp.1-8.

Smith, J. C. *et al.* (2005). Robust method for proteome analysis by MS/MS using an entire translated genome: demonstration on the ciliome of *Tetrahymena thermophila*. *J Proteome Res* 4 (3), pp.909-919.

Snustad, D. P. *et al.* (1992). The small genome of *Arabidopsis* contains at least nine expressed beta-tubulin genes. *Plant Cell* 4 (5), pp.549-556.

Spang, A. *et al.* (1993). The calcium-binding protein cell division cycle 31 of *Saccharomyces cerevisiae* is a component of the half bridge of the spindle pole body. *The Journal of Cell Biology* 123 (2), pp.405-416.

Springer, A. L. *et al.* (2011). Silencing of a putative inner arm dynein heavy chain results in flagellar immotility in *Trypanosoma brucei*. *Mol Biochem Parasitol* 175 (1), pp.68-75.

Stearns, T. (2005). The tubulin superfamily*Centrosomes in Development and Disease*. Wiley-VCH Verlag GmbH & Co. KGaA, pp.17-25.

Stechmann, A. and Cavalier-Smith, T. (2003). The root of the eukaryote tree pinpointed. *Curr Biol* 13 (17), pp.R665-666.

Steffen, W. and Linck, R. W. (1988). Evidence for tektins in centrioles and axonemal microtubules. *Proc Natl Acad Sci U S A* 85 (8), pp.2643-2647.

Stemm-Wolf, A. J., Meehl, J. B. and Winey, M. (2013). Sfr13 is a member of a large family of asymmetrically localized Sfi1-repeat proteins and is important for basal body separation and stability in *Tetrahymena thermophila*. *J Cell Sci*.

Stephan, A. *et al.* (2007). An essential quality control mechanism at the eukaryotic basal body prior to intraflagellar transport. *Traffic* 8 (10), pp.1323-1330.

Steverding, D. (2008). The history of African trypanosomiasis. *Parasit Vectors* 1 (3), pp.1-8.

Stolc, V. *et al.* (2005). Genome-wide transcriptional analysis of flagellar regeneration in *Chlamydomonas reinhardtii* identifies orthologs of ciliary disease genes. *Proc Natl Acad Sci U S A* 102 (10), pp.3703-3707.

Straschil, U. *et al.* (2010). The Armadillo repeat protein PF16 is essential for flagellar structure and function in *Plasmodium* male gametes. *PLoS One* 5 (9), p.e12901.

Subota, I. *et al.* (2011). ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation. *Mol Biol Cell* 22 (22), pp.4205-4219.

Sugimoto, M. *et al.* (2008). The keratin-binding protein Albatross regulates polarization of epithelial cells. *J Cell Biol* 183 (1), pp.19-28.

Sui, H. and Downing, K. H. (2010). Structural basis of interprotofilament interaction and lateral deformation of microtubules. *Structure* 18 (8), pp.1022-1031.

Summers, K. E. and Gibbons, I. R. (1971). Adenosine triphosphate-induced sliding of tubules in trypsin-treated flagella of sea-urchin sperm. *Proc Natl Acad Sci U S A* 68 (12), pp.3092-3096.

Sun, L. and Wang, C. C. (2011). The structural basis of localizing polo-like kinase to the flagellum attachment zone in *Trypanosoma brucei*. *PLoS One* 6 (11), p.e27303.

Sunkel, C. E. and Glover, D. M. (1988). Ppolo, a mitotic mutant of *Drosophila* displaying abnormal spindle poles. *J Cell Sci* 89 (Pt 1), pp.25-38.

Sunter, J. D. *et al.* (2015). A dynamic coordination of flagellum and cytoplasmic cytoskeleton assembly specifies cell morphogenesis in trypanosomes. *J Cell Sci* 128 (8), pp.1580-1594.

Surpili, M. J., Delben, T. M. and Kobarg, J. (2003). Identification of proteins that interact with the central coiled-coil region of the human protein kinase NEK1. *Biochemistry* 42 (51), pp.15369-15376.

Suzuki, T. *et al.* (2004). Mutations in EFHC1 cause juvenile myoclonic epilepsy. *Nat Genet* 36 (8), pp.842-849.

Suzuki, T. *et al.* (2002). Identification and mutational analysis of candidate genes for juvenile myoclonic epilepsy on 6p11-p12: LRRC1, GCLC, KIAA0057 and CLIC5. *Epilepsy Res* 50 (3), pp.265-275.

Taylor, J. E. and Rudenko, G. (2006). Switching trypanosome coats: what's in the wardrobe? *Trends Genet* 22 (11), pp.614-620.

Tetley, L. and Vickerman, K. (1985). Differentiation in *Trypanosoma brucei*: host-parasite cell junctions and their persistence during acquisition of the variable antigen coat. *J Cell Sci* 74, pp.1-19.

Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* Chapter 2, p.Unit 2 3.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22 (22), pp.4673-4680.

Tilney, L. G. *et al.* (1973). Microtubules: evidence for 13 protofilaments. *J Cell Biol* 59 (2 Pt 1), pp.267-275.

Tobin, J. L. and Beales, P. L. (2009). The nonmotile ciliopathies. *Genet Med* 11 (6), pp.386-402.

Towers, K. (2010). *Functional investigation of Trypanosoma brucei microtubule associated proteins and their role in cellular morphogenesis.*

Trevor, K. T., McGuire, J. G. and Leonova, E. V. (1995). Association of vimentin intermediate filaments with the centrosome. *J Cell Sci* 108 (Pt 1), pp.343-356.

Tribolium Genome Sequencing, C. *et al.* (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452 (7190), pp.949-955.

Trojan, P. *et al.* (2008). Centrins in retinal photoreceptor cells: regulators in the connecting cilium. *Prog Retin Eye Res* 27 (3), pp.237-259.

Tu, X. *et al.* (2006). An aurora kinase homologue is involved in regulating both mitosis and cytokinesis in *Trypanosoma brucei*. *J Biol Chem* 281 (14), pp.9677-9687.

Turk, E. *et al.* (2015). Zeta-Tubulin is a member of a conserved tubulin module and is a component of the centriolar basal foot in multiciliated cells. *Current Biology* ePrint, pp.790-793.

Umeyama, T. and Wang, C. C. (2008). Polo-like kinase is expressed in S/G2/M phase and associated with the flagellum attachment zone in both procyclic and bloodstream forms of *Trypanosoma brucei*. *Eukaryot Cell* 7 (9), pp.1582-1590.

Upadhya, P. *et al.* (2000). Mutations in a NIMA-related kinase gene, Nek1, cause pleiotropic effects including a progressive polycystic kidney disease in mice. *Proc Natl Acad Sci U S A* 97 (1), pp.217-221.

Uppaluri, S. *et al.* (2011). Impact of microscopic motility on the swimming behavior of parasites: straighter trypanosomes are more directional. *PLoS Comput Biol* 7 (6), p.e1002058.

Vadnais, M. L., Lin, A. M. and Gerton, G. L. (2014). Mitochondrial fusion protein MFN2 interacts with the mitostatin-related protein MNS1 required for mouse sperm flagellar structure and function. *Cilia* 3 (5), pp.1-15.

van Reeuwijk, J., Arts, H. H. and Roepman, R. (2011). Scrutinizing ciliopathies by unraveling ciliary interaction networks. *Hum Mol Genet* 20 (R2), pp.R149-157.

van Rooijen, E. *et al.* (2008). LRRC50, a conserved ciliary protein implicated in polycystic kidney disease. *J Am Soc Nephrol* 19 (6), pp.1128-1138.

van Wijk, E. *et al.* (2009). Usher syndrome and Leber congenital amaurosis are molecularly linked via a novel isoform of the centrosomal ninein-like protein. *Human Molecular Genetics* 18 (1), pp.51-64.

Vaughan, S. (2010). Assembly of the flagellum and its role in cell morphogenesis in *Trypanosoma brucei*. *Curr Opin Microbiol* 13 (4), pp.453-458.

Vaughan, S. *et al.* (2000). New tubulins in protozoal parasites. *Curr Biol* 10 (7), pp.R258-259.

Vaughan, S. *et al.* (2008). A repetitive protein essential for the flagellum attachment zone filament structure and function in *Trypanosoma brucei*. *Protist* 159 (1), pp.127-136.

Vaughan, S., Shaw, M. and Gull, K. (2006). A post-assembly structural modification to the lumen of flagellar microtubule doublets. *Curr Biol* 16 (12), pp.R449-450.

Vecchione, A. *et al.* (2009). MITOSTATIN, a putative tumor suppressor on chromosome 12q24.1, is downregulated in human bladder and breast cancer. *Oncogene* 28 (2), pp.257-269.

Venter, J. C. *et al.* (2001). The sequence of the human genome. *Science* 291 (5507), pp.1304-1351.

Vickerman, K. (1969). On the surface coat and flagellar adhesion in trypanosomes. *J Cell Sci* 5 (1), pp.163-193.

Vickerman, K. *et al.* (1988). Biology of African trypanosomes in the tsetse fly. *Biol Cell* 64 (2), pp.109-119.

Vilela, R. C. and Benchimol, M. (2011). Interaction of *Trichomonas vaginalis* and *Tritrichomonas foetus* with keratin: an important role in parasite infection. *Mem Inst Oswaldo Cruz* 106 (6), pp.701-704.

Vincensini, L., Blisnick, T. and Bastin, P. (2011). 1001 model organisms to study cilia and flagella. *Biol Cell* 103 (3), pp.109-130.

Vorobjev, I. A. and Chentsov Yu, S. (1982). Centrioles in the cell cycle. I. Epithelial cells. *J Cell Biol* 93 (3), pp.938-949.

Wakid, M. H. and Bates, P. A. (2004). Flagellar attachment of Leishmania promastigotes to plastic film *in vitro*. *Exp Parasitol* 106 (3-4), pp.173-178.

Wang, L., Ciganda, M. and Williams, N. (2013). Association of a novel preribosomal complex in *Trypanosoma brucei* determined by fluorescence resonance energy transfer. *Eukaryot Cell* 12 (2), pp.322-329.

Wang, M., Gheiratmand, L. and He, C. Y. (2012). An interplay between Centrin2 and Centrin4 on the bi-lobed structure in *Trypanosoma brucei*. *Mol Microbiol* 83 (6), pp.1153-1161.

Wargo, M. J., Dymek, E. E. and Smith, E. F. (2005). Calmodulin and PF6 are components of a complex that localizes to the C1 microtubule of the flagellar central apparatus. *J Cell Sci* 118 (Pt 20), pp.4655-4665.

Wargo, M. J. and Smith, E. F. (2003). Asymmetry of the central apparatus defines the location of active microtubule sliding in *Chlamydomonas* flagella. *Proc Natl Acad Sci U S A* 100 (1), pp.137-142.

Wasbrough, E. R. *et al.* (2010). The *Drosophila melanogaster* sperm proteome-II (DmSP-II). *J Proteomics* 73 (11), pp.2171-2185.

Waters, A. M. and Beales, P. L. (2011). Ciliopathies: an expanding disease spectrum. *Pediatr Nephrol* 26 (7), pp.1039-1056.

Weber, C. *et al.* (1994). High level expression in *Escherichia coli* and characterization of the EF-hand calcium-binding protein caltractin. *J Biol Chem* 269 (22), pp.15795-15802.

Wei, Y. *et al.* (2014). Centrin3 in trypanosomes maintains the stability of a flagellar inner-arm dynein for cell motility. *Nat Commun* 5 (4060), pp.1-11.

Weisse, S. *et al.* (2012). A quantitative 3D motility analysis of *Trypanosoma brucei* by use of digital in-line holographic microscopy. *PLoS One* 7 (5), p.e37296.

Wheeler, R. J. (2012). *Generation, regulation and function of morphology in Leishmania and Trypanosoma.*

Wheeler, R. J., Gluenz, E. and Gull, K. (2013a). The limits on trypanosomatid morphological diversity. *PLoS One* 8 (11), p.e79581.

Wheeler, R. J. *et al.* (2013b). Cytokinesis in *Trypanosoma brucei* differs between bloodstream and tsetse trypomastigote forms: implications for microtubule-based morphogenesis and mutant analysis. *Mol Microbiol* 90 (6), pp.1339-1355.

Wheeler, T. J., Clements, J. and Finn, R. D. (2014). Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* 15 (7), pp.1-9.

Wheway, G., Parry, D. A. and Johnson, C. A. (2014). The role of primary cilia in the development and disease of the retina. *Organogenesis* 10 (1), pp.69-85.

WHO, W. H. O. (2012). *Human African Trypanosomiasis (sleeping sickness).Fact sheets.*

Wiche, G. (1998). Role of plectin in cytoskeleton organization and dynamics. *J Cell Sci* 111 (Pt 17), pp.2477-2486.

Wickstead, B., Ersfeld, K. and Gull, K. (2002). Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei*. *Mol Biochem Parasitol* 125 (1-2), pp.211-216.

Wickstead, B. and Gull, K. (2007). Dyneins across eukaryotes: a comparative genomic analysis. *Traffic* 8 (12), pp.1708-1721.

Wickstead, B. and Gull, K. (2011). The evolution of the cytoskeleton. *J Cell Biol* 194 (4), pp.513-525.

Williams, C. L. *et al.* (2011). MKS and NPHP modules cooperate to establish basal body/transition zone membrane associations and ciliary gate function during ciliogenesis. *J Cell Biol* 192 (6), pp.1023-1041.

Wilson, N. F. *et al.* (2008). Regulation of flagellar length in *Chlamydomonas*. *Semin Cell Dev Biol* 19 (6), pp.494-501.

Winkles, J. A. and Alberts, G. F. (2005). Differential regulation of polo-like kinase 1, 2, 3, and 4 gene expression in mammalian cells and tissues. *Oncogene* 24 (2), pp.260-266.

Wood, V. *et al.* (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415 (6874), pp.871-880.

Woods, A., Baines, A. J. and Gull, K. (1989a). Evidence for a Mr 88,000 glycoprotein with a transmembrane association to a unique flagellum attachment region in *Trypanosoma brucei*. *J Cell Sci* 93 (Pt 3), pp.501-508.

Woods, A. *et al.* (1989b). Definition of individual components within the cytoskeleton of *Trypanosoma brucei* by a library of monoclonal antibodies. *J Cell Sci* 93 (Pt 3), pp.491-500.

Woodward, R., Carden, M. J. and Gull, K. (1995). Immunological characterization of cytoskeletal proteins associated with the basal body, axoneme and flagellum attachment zone of *Trypanosoma brucei*. *Parasitology* 111 (Pt 1), pp.77-85.

Woodward, R. and Gull, K. (1990). Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of *Trypanosoma brucei*. *J Cell Sci* 95 (Pt 1), pp.49-57.

Yanagisawa, H. A. and Kamiya, R. (2004). A tektin homologue is decreased in chlamydomonas mutants lacking an axonemal inner-arm dynein. *Mol Biol Cell* 15 (5), pp.2105-2115.

Yanagisawa, J. *et al.* (1997). The molecular interaction of Fas and FAP-1. A tripeptide blocker of human Fas interaction with FAP-1 promotes Fas-induced apoptosis. *J Biol Chem* 272 (13), pp.8539-8545.

Yang, J., Adamian, M. and Li, T. (2006). Rootletin interacts with C-Nap1 and may function as a physical linker between the pair of centrioles/basal bodies in cells. *Mol Biol Cell* 17 (2), pp.1033-1040.

Yang, J. *et al.* (2002). Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet. *J Cell Biol* 159 (3), pp.431-440.

Yano, J. *et al.* (2013). Proteomic analysis of the cilia membrane of *Paramecium tetraurelia*. *J Proteomics* 78 (0), pp.113-122.

Ying, G. *et al.* (2014). Centrin 2 is required for mouse olfactory ciliary trafficking and development of ependymal cilia planar polarity. *J Neurosci* 34 (18), pp.6377-6388.

Yu, Z., Liu, Y. and Li, Z. (2012). Structure-function relationship of the Polo-like kinase in *Trypanosoma brucei*. *J Cell Sci* 125 (Pt 6), pp.1519-1530.

Zaghloul, N. A. and Katsanis, N. (2010). Functional modules, mutational load and human genetic disease. *Trends Genet* 26 (4), pp.168-176.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9 (40), pp.1-8.

Zhao, Z. *et al.* (2008). p166, a link between the trypanosome mitochondrial DNA and flagellum, mediates genome segregation. *EMBO J* 27 (1), pp.143-154.

Zhou, J. *et al.* (2012). MNS1 is essential for spermiogenesis and motile ciliary functions in mice. *PLoS Genet* 8 (3), p.e1002516.

Zhou, Q. *et al.* (2010). A comparative proteomic analysis reveals a new bi-lobe protein required for bi-lobe duplication and cell division in *Trypanosoma brucei*. *PLoS One* 5 (3), p.e9660.

Zhou, Q. *et al.* (2011). A coiled-coil- and C2-domain-containing protein is required for FAZ assembly and cell morphology in *Trypanosoma brucei*. *J Cell Sci* 124 (Pt 22), pp.3848-3858.

Zhu, X. and Kaverina, I. (2013). Golgi as an MTOC: making microtubules for its own good. *Histochem Cell Biol* 140 (3), pp.361-367.

Appendix

Appendix 1: Motility assay macro

Motility assay macro (Wheeler, 2012). See CD insert for macro file

Appendix 2: EFHC1 and EFHC2 orthologs from Chordata

Alignment (

Figure 105) and phylogenetic tree (Figure 106) of EFHC1 and EFHC2 orthologs from Chordata.

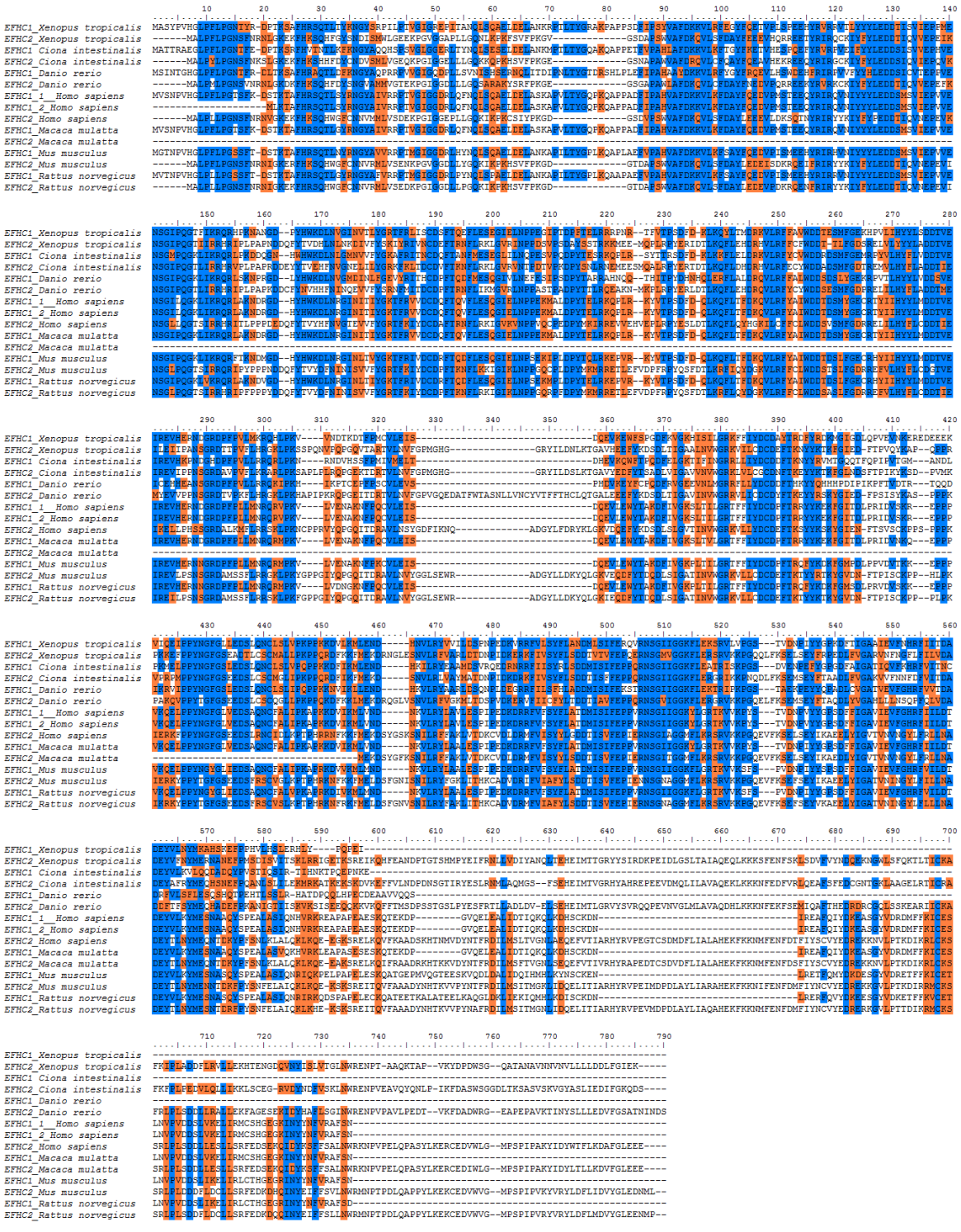


Figure 105: Alignment of EFHC1 and EFHC2 orthologs

Alignment of the orthologous protein sequences from chordata. (Separate alignments of each group are in the appendix; Figure 107 and Figure 108). Identical amino acids are blue and similar amino acids are orange.

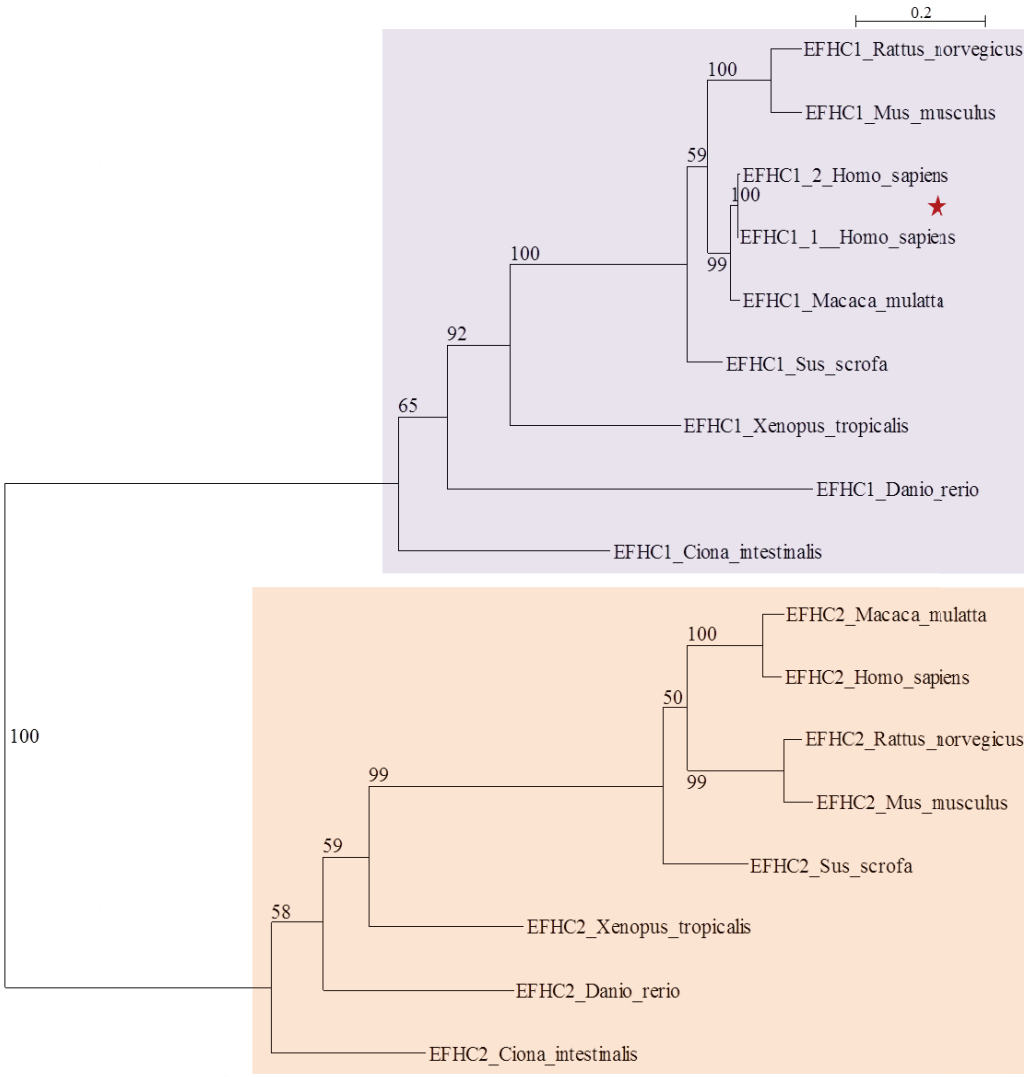


Figure 106: Relationship between EFHC1 and EFHC2 orthologs in chordata

The protein sequences of EFHC1 (purple) and EFHC2 (orange) orthologs were aligned and assembled into a phylogenetic tree. Branch positions were bootstrapped with 100 replicates, bootstrap values are displayed at the branch nodes. The red star highlights the two isoforms of *H. sapiens* EFHC1. Scale bar = 0.2 amino acid substitutions.

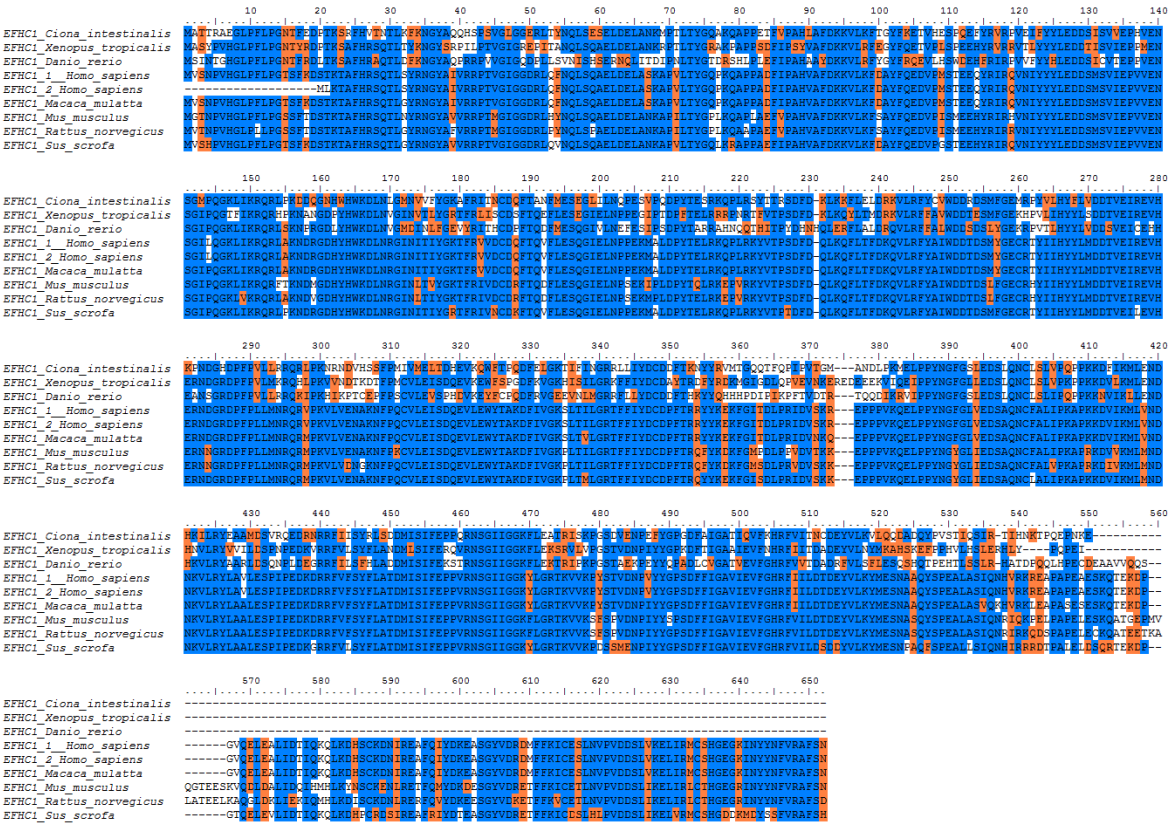


Figure 107: Conservation of EFHC1 orthologs

Alignment of the orthologous protein sequences from chordata. Identical amino acids are blue and similar amino acids are orange

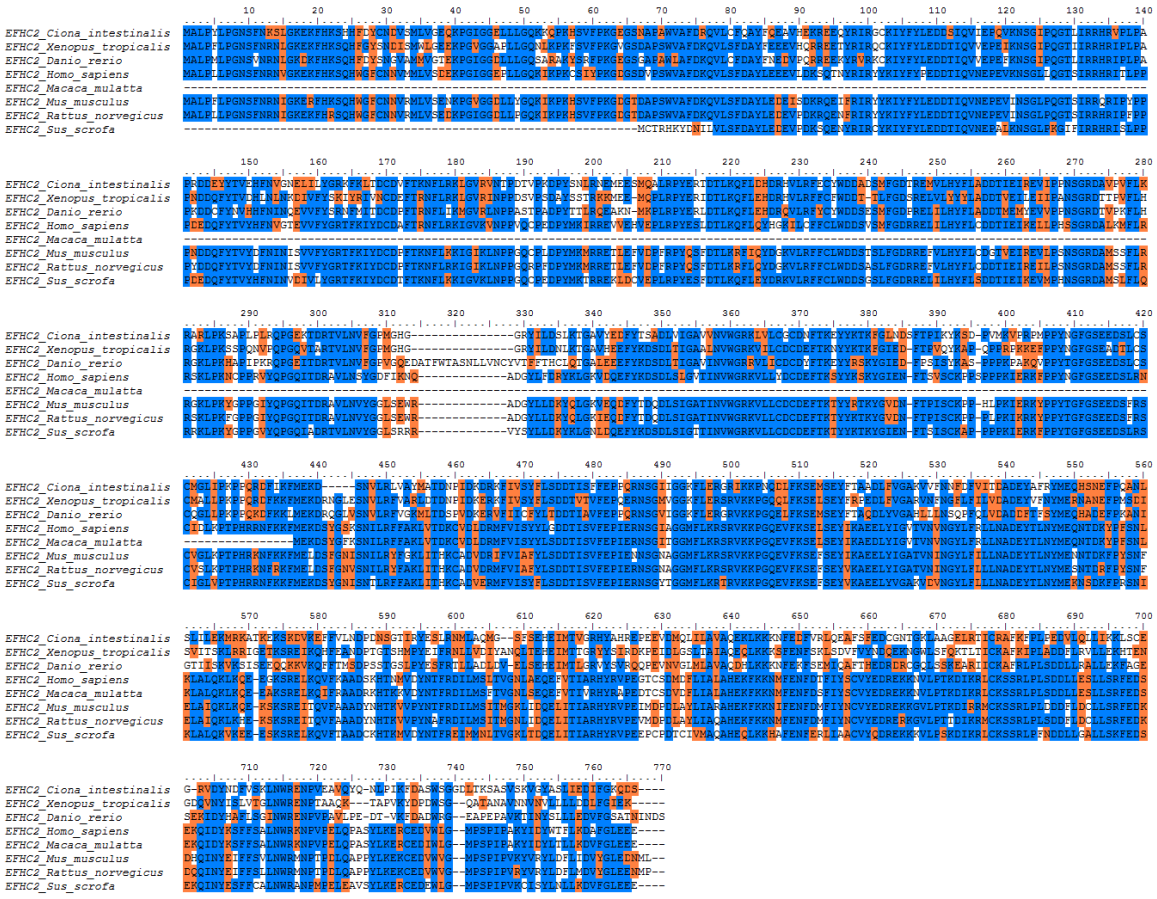


Figure 108: Conservation of EFHC2 orthologs

Alignment of the orthologous protein sequences from chordata. Identical amino acids are blue and similar amino acids are orange

Appendix 3: CCDC19 isoforms

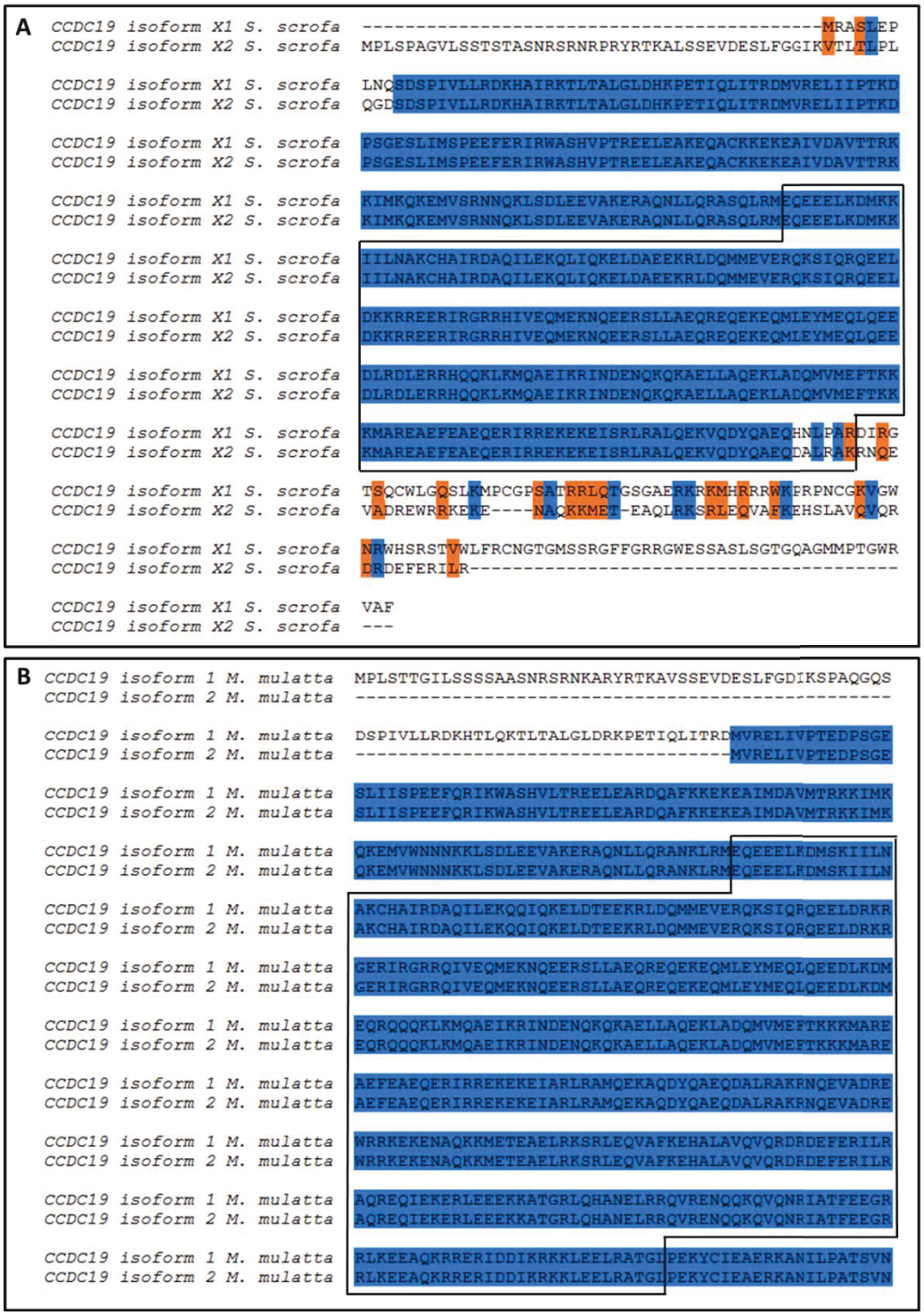


Figure 109: Alignment of CCDC19 isoforms

A) Alignment of *Sus scrofa* CCDC19 isoforms X1 and X2 sequences (XP_005663268/XP_005663267). B) Alignment of *M. mulatta* CCDC19 isoforms X1 and X2 sequences (XP_001115183/XP_002801869). The black box indicates the position of the TPH domains. Identical amino acids are coloured blue and similar amino acids are coloured orange.

Appendix 4: Identity matrix of CCDC19 and CCDC19-like proteins

Table 22: Sequence identity matrix between *Drosophila* spp. CCDC19 and CCDC19-like proteins

			CCDC19												CCDC19-like												
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
CCDC19	1	GM22444_sechellia	100																							
	2	GD17655_simulans	94.59	100																						
	3	CG11449_melanogaster	96.84	90.99	100																					
	4	GG16255_erecta	94.73	87.5	94.35	100																				
	5	GE22616_yakuba	95.14	89.81	94.58	95.86	100																			
	6	GF10977_ananassae	84.76	80.18	84.39	84.24	85.79	100																		
	7	GL23132_persimilis	78.57	73.64	77.81	76.42	77.63	80	100																	
	8	GA11010_pseudoobscura	80.07	73.64	79.33	78.91	79.59	81.48	98.48	100																
	9	GK16758_willistoni	77.05	71.56	76.49	75.66	76.55	75.65	78.83	80.26	100															
	10	GJ11634_virilis	71.54	61.11	70.97	71.64	71.19	71.06	73.72	75.61	72.95	100														
	11	GI11378_mojavensis	72.42	65.09	71.86	71.78	71.7	71	73.4	74.63	73.08	85.74	100													
	12	GH16691_grimshaw	71.35	63.89	70.79	72.35	72.13	73.05	75.96	77.05	74.95	85.58	82.19	100												
CCDC19-like	13	GM12655_sechellia	23.77	17.48	23.58	23.81	23.72	23.83	24.87	23.73	24.81	22.74	22.89	23.73	100											
	14	GD24597_simulans	23.77	17.48	23.58	23.81	23.72	23.83	24.87	23.73	24.81	22.74	22.89	23.73	99.31	100										
	15	CG5062_melanogaster	23.4	16.5	23.21	23.43	23.34	23.6	24.55	23.68	24.44	22.66	22.43	23.64	97.58	97.58	100									
	16	GE16824_yakuba	23.96	18.45	23.77	24	23.91	23.97	25.06	23.87	24.62	22.7	22.47	23.68	95.85	96.02	95.69	100								
	17	GG18507_erecta	23.96	17.48	23.77	24	23.91	23.78	24.55	23.5	24.81	22.51	22.66	23.68	93.77	93.94	93.28	95.52	100							
	18	GF20956_ananassae	23.82	17.65	23.63	24.05	23.19	24.02	25.91	24.11	24.48	21.99	22.33	23.35	83.83	84.17	83.68	82.81	81.42	100						
	19	GL14344_persimilis	22.67	14.42	22.87	23.2	23.2	22.31	20.86	22.39	22.67	20.85	20.81	22.44	61.43	61.25	61.46	61.21	60.32	62.19	100					
	20	GA18630_pseudoobscura	23.06	14.42	23.25	23.19	23.57	22.7	21.45	22.6	23.06	21.28	21.24	22.83	61.95	61.95	61.98	61.74	60.87	63.05	97.54	100				
	21	GK10338_willistoni	24.63	12.73	25	25.05	25.14	24.21	23.1	22.86	24.81	23.52	23.38	24.86	59.27	59.45	59.14	58.72	58.2	60.35	56.06	57.56	100			
	22	GJ16580_virilis	22.47	10.28	22.28	23.06	22.6	22.01	23.39	22.47	22.7	21.76	21.43	21.43	48.26	48.08	48.52	48.17	46.78	48.16	46.79	47.82	49.66	100		
	23	GI15308_mojavensis	20.9	11.01	20.9	21.28	21.01	20.52	21.88	21.19	22.01	20.93	20.59	20.96	46.43	46.26	45.85	45.75	44.71	46.95	46.63	47.66	48.5	79.26	100	
	24	GH23457_grimshaw	19.49		18.97	20	20.51	16.92	19.61	18.97	20	18.46	20.51	19.49	45.73	45.23	45.23	44.72	43.72	45.73	45.45	45.96	47.98	74.87	65.33	100

Appendix 5: Identity matrix of MNS1-like A and B proteins

Table 23: Sequence identity matrix between *Drosophila* spp. MNS1-like A and B proteins

		GeneID_subspecies	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
MNS1-like A	1	GJ24529_virilis	100																							
	2	GI24886_mojavensis	65.72	100																						
	3	GH14366_grimshaw	63.87	55.87	100																					
	4	GM23730_sechellia	48.07	43.97	43.98	100																				
	5	GD18540_simulans	48.07	43.53	43.98	98.27	100																			
	6	CG7352_melanogaster	48.59	42.7	43.38	91.99	92.86	100																		
	7	GG10368_erecta	46.27	42.04	42.95	88.31	89.18	86.7	100																	
	8	GE25874_yakuba	46.53	42.57	42.39	88.72	89.15	86.88	92.69	100																
	9	GF18348_ananassae	48.34	40.92	45.49	64.72	65.15	65.02	64.81	65.38	100															
	10	GL24191_persimilis	43.04	41.7	39.61	51.22	51.44	50.33	50.11	50.44	49.13	100														
	11	GA20287_pseudoobscura	41.79	43.07	39.81	52.49	52.26	51.31	50.83	50.95	49.65	98.33	100													
	12	GK20748_willistoni	41.48	39.87	39.78	41.77	41.34	40.3	40.95	40.82	40.86	40.09	41.47	100												
MNS1-like B	13	GK13186_willistoni	46.43	45.17	45.61	47.14	47.14	45.71	46.43	46.78	48.69	42.93	42.72	50.47	100											
	14	GM24187_sechellia	18.49	17.44	17	19.38	19.38	19.65	19.43	18.81	17.1	16.04	17.8	15.49	14.36	100										
	15	GD18979_simulans	18.44	17.4	16.96	19.33	19.33	19.6	19.38	18.76	17.06	16	17.76	15.23	14.56	98.98	100									
	16	CG3610_melanogaster	17.71	17.22	16.34	19.15	19.15	19.43	18.98	18.58	16.23	15.59	17.32	15.04	14.84	96.72	96.11	100								
	17	GG16878_erecta	17.71	15.89	16.56	18.26	18.26	18.54	17.88	17.48	16.02	16.04	17.56	15.49	13.87	86.89	86.07	87.5	100							
	18	GE24259_yakuba	18.23	16.56	17.44	18.93	18.93	19.43	18.54	18.58	17.1	16.04	17.8	15.93	14.6	88.11	87.09	88.93	93.24	100						
	19	GF19498_ananassae	20.05	18.18	17.29	19.91	19.91	19.51	19.29	19.56	17.83	17.67	19.51	15.33	16.79	55.97	55.97	55.76	55.97	56.58	100					
	20	GL26723_persimilis	18.77	17.39	15.91	18.58	18.35	18.64	18.41	19.13	16.7	17.66	19.55	15.49	14.25	35.46	35.38	34.39	34.82	35.46	38.32	100				
	21	GA23029_pseudoobscura	18.3	17.04	15.6	17.96	17.74	17.8	17.36	18.06	15.73	16.19	17.87	15.2	13.49	35.8	35.73	35.19	35.8	36.21	38.16	88.75	100			
	22	GJ14772_virilis	16.71	17.66	17.28	19.26	19.26	19.52	19.09	19.35	17.66	15.9	17.14	17.39	16.75	29.98	29.92	29.77	29.98	30.18	32.32	27.77	28.95	100		
	23	GI15335_mojavensis	18.04	16.56	16.23	17.32	17.54	18.04	16.96	16.78	16.2	14.41	15.75	15.25	14.87	29.36	29.3	28.95	28.95	28.95	29.47	26.72	27.53	72.02	100	
	24	GH24395_grimshaw	16.15	17.14	17.42	17.43	16.99	17.28	17.06	17.75	14.83	15.4	16.59	16.88	16.43	27.66	27.81	27.66	27.46	27.05	30.02	25.83	27.07	71.68	61.71	100

Appendix 6: HMM for the TPH domain

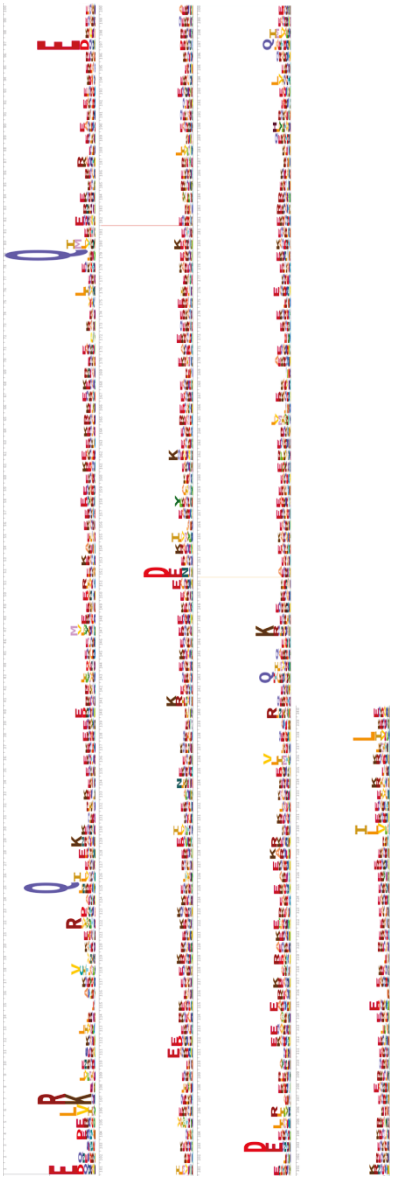


Figure 110: HMM for the TPH domain

The TPH domain was excised from each protein (n=113) *in silico* using positions annotated in Appendix 7.

The isolated domains were used to generate a hidden markov model (HMM) in skyalign.

Appendix 7: Architecture of TPH domain-containing protein sequences

Table 24: Annotation of TPH domain-containing protein sequences

The structure of eukaryotic TPH domain-containing protein sequences, including the position of the TPH domain according to CDD and Pfam.

Protein group	NCBI Accession	Organism	AA	TPHD				N terminus				C terminus				Annotation	
				Start	End	Length	% of whole protein	Start	End	Length	% of whole protein	Start	End	Length	% of whole protein	CDD	Pfam
TCHP	NP_001035433	*Danio rerio*	499	146	484	338	67.74	1	145	145	29.058	485	499	15	3.006	146-484	146-488
TCHP	CDQ66337	*Oncorhynchus mykiss*	537	150	476	326	60.71	1	149	149	27.747	477	537	61	11.359	150-476	149-492
TCHP	NP_115676	*Homo sapiens*	498	146	476	330	66.27	1	145	145	29.116	477	498	22	4.4177	146-476	145-488
TCHP	EHH21167	*Macaca mulatta*	498	146	476	330	66.27	1	145	145	29.116	477	498	22	4.4177	146-476	145-488
TCHP	XP_003483491	*Sus scrofa*	500	188	478	290	58.00	1	187	187	37.4	479	500	22	4.4	188-478	147-490
TCHP	NP_084268	*Mus musculus*	497	145	444	299	60.16	1	144	144	28.974	445	497	53	10.664	145-444	144-490
TCHP	NP_001178595	*Rattus norvegicus*	497	145	481	336	67.61	1	144	144	28.974	482	497	16	3.2193	145-481	144-493
TCHP	XP_422992	*Gallus gallus*	471	128	428	300	63.69	1	127	127	26.964	429	471	43	9.1295	128-428	128-466
TCHP	XP_004910671	*Xenopus tropicalis*	500	152	478	326	65.20	1	151	151	30.2	479	500	22	4.4	152-478	147-484
TCHP	XP_969483	*Tribolium castaneum*	504	175	485	310	61.51	1	174	174	34.524	486	504	19	3.7698	175-485	
TCHP	XP_004928902	*Bombyx mori*	495														
TCHP	KFM76999	*Stegodyphus mimosarum*	362														
TCHP	XP_003435775	*Anopheles gambiae*	510														
MNS1	XP_008199885	*Tribolium castaneum*	460	100	449	349	75.87	1	99	99	21.522	450	460	11	2.3913	100-449	100-449
MNS1	NP_001096482	*Xenopus tropicalis*	498	115	464	349	70.08	1	114	114	22.892	465	498	34	6.8273	115-464	115-464
MNS1	NP_060835	*Homo sapiens*	495	116	465	349	70.51	1	115	115	23.232	466	495	30	6.0606	116-465	116-465
MNS1	XP_005168964	*Danio rerio*	503	117	466	349	69.38	1	116	116	23.062	467	503	37	7.3559	117-466	117-466
MNS1	CDQ61895	*Oncorhynchus mykiss*	527	142	491	349	66.22	1	141	141	26.755	492	527	36	6.8311	142-491	142-491
MNS1	XP_845569	*Trypanosoma brucei*	436	73	420	347	79.59	1	72	72	16.514	421	436	16	3.6697	73-420	72-421
MNS1	XP_002947788	*Volvox carteri*	857	127	475	348	40.61	1	126	126	14.702	476	857	382	44.574	127-475	127-475
MNS1	XP_002896208	*Phytopthora infestans*	463	98	447	349	75.38	1	97	97	20.95	448	463	16	3.4557	98-447	98-447
MNS1	XP_001349839	*Plasmodium falciparum*	495	130	480	350	70.71	1	129	129	26.061	481	495	15	3.0303	130-480	130-480
MNS1	XP_002683398	*Naegleria gruberi*	426	46	394	348	81.69	1	45	45	10.563	395	426	32	7.5117	46-394	45-394
MNS1	XP_001699616	*Chlamydomonas reinhardtii*	426	48	396	348	81.69	1	47	47	11.033	397	426	30	7.0423	48-396	48-396
MNS1	XP_001017577	*Tetrahymena thermophila*	447	67	416	349	78.08	1	66	66	14.765	417	447	31	6.9351	67-416	67-416
MNS1	XP_002286669	*Thalassiosira pseudonana*	411	58	406	348	84.67	1	57	57	13.869	407	411	5	1.2165	58-406	58-407
MNS1	XP_002371809	*Toxoplasma gondii*	335	107	335	228	68.06	1	106	106	31.642	336	335	0	0	107-335	107-335
MNS1	XP_009861318	*Ciona intestinalis*	497	117	466	349	70.22	1	116	116	23.34	467	497	31	6.2374	117-466	117-466
MNS1	CCI46780	*Albugo candida*	487	119	468	349	71.66	1	118	118	24.23	469	487	19	3.9014	119-468	119-468
MNS1	XP_004922932	*Bombyx mori*	303	101	300	199	65.68	1	100	100	33.003	301	303	3	0.9901	101-300	101-302
MNS1	KFM81249 (partial	*Stegodyphus mimosarum*	476	100	449	349	73.32	1	99	99	20.798	450	476	27	5.6723	100-449	100-449
MNS1	XP_313110	*Anopheles gambiae*	444	103	425	322	72.52	1	102	102	22.973	426	444	19	4.2793	103-425	103-428
MNS1	XP_006563783	*Apis mellifera*	428	139	391	252	58.88	1	138	138	32.243	392	428	37	8.6449	139-391	133-407
MNS1	XP_001684925	*Leishmania major*	411	38	385	347	84.43	1	37	37	9.0024	386	411	26	6.326	38-385	37-386
MNS1	NP_001253790	*Macaca mulatta*	495	116	465	349	70.51	1	115	115	23.232	466	495	30	6.0606	116-465	116-465
MNS1	**ENSSSCP000000	*Sus scrofa*	497	115	464	349	70.22	1	114	114	22.938	465	497	33	6.6398	115-464	115-464
MNS1	NP_032639	*Mus musculus*	491	116	465	349	71.08	1	115	115	23.422	466	491	26	5.2953	116-465	116-465
MNS1	NP_001007753	*Rattus norvegicus*	498	116	465	349	70.08	1	115	115	23.092	466	498	33	6.6265	116-465	116-465
MNS1	XP_423957	*Gallus gallus*	464	83	432	349	75.22	1	82	82	17.672	433	464	32	6.8966	83-432	83-432
MNS1	XP_003884529	*Neospora caninum*	485	108	457	349	71.96	1	107	107	22.062	458	485	28	5.7732	108-457	108-457
MNS1	CDJ59336	*Eimeria maxima*	508	112	461	349	68.70	1	111	111	21.85	462	508	47	9.252	112-461	112-461
MNS1-B	NP_650396	*Drosophila melanogaster*	488	125	473	348	71.31	1	124	124	25.41	474	488	15	3.0738	125-473	124-473
MNS1-A	NP_649795 (CG735	*Drosophila melanogaster*	466	97	442	345	74.03	1	96	96	20.601	443	466	24	5.1502	97-442	96-442
MNS1	XP_001434111	*Paramecium tetraurelia*	421	44	394	350	83.14	1	43	43	10.214	395	421	27	6.4133	44-394	43-394
CCDC11	XP_001691996	*Chlamydomonas reinhardtii*	482	134	464	330	68.46	1	133	133	27.593	465	482	18	3.7344	134-464	134-467

CCDC11	XP_001691996	*Chlamydomonas reinhardtii*	482	134	464	330	68.46	1	133	133	27.593	465	482	18	3.7344	134-464	134-467
CCDC11	XP_004226085	*Ciona intestinalis*	516	159	488	329	63.76	1	158	158	30.62	489	516	28	5.4264	159-488	159-493
CCDC11	CCI44136	*Albugo candida*	491	140	468	328	66.80	1	139	139	28.31	469	491	23	4.6843	140-468	140-474
CCDC11	XP_001683505	*Leishmania major*	490	138	463	325	66.33	1	137	137	27.959	464	490	27	5.5102	138-463	134-477
CCDC11	XP_003883663	*Neospora caninum*	528	152	486	334	63.26	1	151	151	28.598	487	528	42	7.9545	152-486	152-498
CCDC11	CDJ59796	*Eimeria maxima*	361	162	355	193	53.46	1	161	161	44.598	356	361	6	1.662	162-355	
CCDC11	XP_004933599	*Bombyx mori*	524	182	485	303	57.82	1	181	181	34.542	486	524	39	7.4427	182-485	
CCDC11	XP_557804	*Anopheles gambiae*	424														
CCDC11	XP_006562850	*Apis mellifera*	434	86	392	306	70.51	1	85	85	19.585	393	434	42	9.6774	86-392	
CCDC11	CDQ60305	*Oncorhynchus mykiss*	521	161	495	334	64.11	1	160	160	30.71	496	521	26	4.9904	161-495	161-496
CCDC11	XP_001091617	*Macaca mulatta*	514	159	460	301	58.56	1	158	158	30.739	461	514	54	10.506	159-460	159-493
CCDC11	XP_005654480	*Sus scrofa*	356	161	334	173	48.60	1	160	160	44.944	335	356	22	6.1798	161-334	159-335
CCDC11	NP_083224	*Mus musculus*	514	159	490	331	64.40	1	158	158	30.739	491	514	24	4.6693	159-490	159-510
CCDC11	XP_001053914	*Rattus norvegicus*	542	187	518	331	61.07	1	186	186	34.317	519	542	24	4.428	187-518	187-527
CCDC11	NP_001038595	*Danio rerio*	519	159	493	334	64.35	1	158	158	30.443	494	519	26	5.0096	159-493	159-500
CCDC11	NP_659457	*Homo sapiens*	514	159	460	301	58.56	1	158	158	30.739	461	514	54	10.506	159-460	159-491
CCDC11	XP_001440964	*Paramecium tetraurelia*	484	136	470	334	69.01	1	135	135	27.893	471	484	14	2.8926	136-470	136-482
CCDC11	XP_002898153	*Phytophthora infestans*	495	164	474	310	62.63	1	163	163	32.929	475	495	21	4.2424	164-474	144-479
CCDC11	XP_001021909	*Tetrahymena thermophila*	489	142	475	333	68.10	1	141	141	28.834	476	489	14	2.863	142-475	141-487
CCDC11	XP_002365948	*Toxoplasma gondii*	528	152	486	334	63.26	1	151	151	28.598	487	528	42	7.9545	152-486	152-503
CCDC11	XP_001808852	*Tribolium castaneum*	535	197	512	315	58.88	1	196	196	36.636	513	535	23	4.2991	197-512	174-513
CCDC11	XP_844797	*Trypanosoma brucei*	483	131	473	342	70.81	1	130	130	26.915	474	483	10	2.0704	131-473	
CCDC11	XP_002946740	*Volvox carteri*	483	134	463	329	68.12	1	133	133	27.536	464	483	20	4.1408	134-463	134-467
CCDC11	NP_001120281	*Xenopus tropicalis*	516	160	491	331	64.15	1	159	159	30.814	492	516	25	4.845	160-491	160-491
CCDC19	XP_001701792	*Chlamydomonas reinhardtii*	501	142	484	342	68.26	1	141	141	28.144	485	501	17	3.3932	142-484	142-489
CCDC19	XP_002127322	*Ciona intestinalis*	544	180	522	342	62.87	1	179	179	32.904	523	544	22	4.0441	180-522	180-527
CCDC19	NP_001076465	*Danio rerio*	543	183	514	331	60.96	1	182	182	33.517	515	543	29	5.3407	183-514	183-530
CCDC19	NP_649390	*Drosophila melanogaster*	538	171	515	344	63.94	1	170	170	31.599	516	538	23	4.2751	210-552	210-557
CCDC19-B	NP_572207 (CG506	*Drosophila melanogaster*	581	210	552	342	58.86	1	209	209	35.972	553	581	29	4.9914	210-552	210-557
CCDC19	NP_036469	*Homo sapiens*	551	186	530	344	62.43	1	185	185	33.575	531	551	21	3.8113	186-530	186-533
CCDC19	XP_002677033	*Naegleria gruberi*	532	151	493	342	64.29	1	150	150	28.195	494	532	39	7.3308	151-493	151-498
CCDC19	XP_001437318	*Paramecium tetraurelia*	492	142	484	342	69.51	1	141	141	28.659	485	492	8	1.626	131-468	131-478
CCDC19	XP_002900117	*Phytophthora infestans*	473	111	453	342	72.30	1	110	110	23.256	454	473	20	4.2283	111-453	111-458
CCDC19	EWS71661	*Tetrahymena thermophila*	494	135	476	341	69.03	1	134	134	27.126	477	494	18	3.6437	135-476	135-482
CCDC19	EPT25970	*Toxoplasma gondii*	507	137	469	332	65.48	1	136	136	26.824	470	507	38	7.4951	137-469	137-481
CCDC19	XP_002288680	*Thalassiosira pseudonana*	511	150	488	338	66.14	1	149	149	29.159	489	511	23	4.501	150-488	148-495
CCDC19	XP_008195146	*Tribolium castaneum*	511	148	492	344	67.32	1	147	147	28.767	493	511	19	3.7182	148-492	148-495
CCDC19	XP_847289	*Trypanosoma brucei*	483	120	461	341	70.60	1	119	119	24.638	462	483	22	4.5549	120-461	120-467
CCDC19	XP_001681449	*Leishmania major*	480	123	464	341	71.04	1	122	122	25.417	465	480	16	3.3333	123-464	123-470
CCDC19	EET00855	*Giardia lamblia*	464	106	444	338	72.84	1	105	105	22.629	445	464	20	4.3103	106-444	106-451
CCDC19	XP_001238576	*Anopheles gambiae*	472	112	454	342	72.46	1	111	111	23.517	455	472	18	3.8136	112-454	112-459
CCDC19	XP_006570023	*Apis mellifera*	455	95	437	342	75.16	1	94	94	20.659	438	455	18	3.956	95-437	95-442
CCDC19	XP_004926031	*Bombyx mori*	528	166	508	342	64.77	1	165	165	31.25	509	528	20	3.7879	166-508	166-513
CCDC19	KFM61683	*Stegodyphus mimosarum*	510	150	491	341	66.86	1	149	149	29.216	492	510	19	3.7255	150-491	150-497
CCDC19	CDQ63523	*Oncorhynchus mykiss*	576	212	556	344	59.72	1	211	211	36.632	557	576	20	3.4722	212-556	212-559
CCDC19	XP_001115183	*Macaca mulatta*	550	186	529	343	62.36	1	185	185	33.636	530	550	21	3.8182	186-529	186-532
CCDC19	XP_005663267	*Sus scrofa*	460	147	353	206	44.78	1	146	146	31.739	354	460	107	23.261	147-353	147-355
CCDC19	NP_082248	*Mus musculus*	551	186	530	344	62.43	1	185	185	33.575	531	551	21	3.8113	186-530	186-533
CCDC19	XP_006250358	*Rattus norvegicus*	551	186	530	344	62.43	1	185	185	33.575	531	551	21	3.8113	186-530	186-533
CCDC19	XP_424097	*Gallus gallus*	478	102	444	342	71.55	1	101	101	21.13	445	478	34	7.113	102-444	102-449
CCDC19	XP_006677515	*Batrachochytrium dendrobatidis*	502	133	474	341	67.93	1	132	132	26.295	475	502	28	5.5777	133-474	133-480
CCDC19	XP_002946373	*Volvox carteri*	502	143	485	342	68.13	1	142	142	28.287	486	502	17	3.3865	143-485	143-490
CCDC19	NP_989263	*Xenopus tropicalis*	447	83	427	344	76.96	1	82	82	18.345	428	447	20	4.4743	83-427	83-430
CCDC173	XP_001701659	*Chlamydomonas reinhardtii*	456	93	420	327	71.71	1	92	92	20.175	421	456	36	7.8947	93-420	
CCDC173	XP_002128435	*Ciona intestinalis*	569	154	464	310	54.48	1	153	153	26.889	465	569	105	18.453	154-464	
CCDC173	XP_002663472	*Danio rerio*	551	139	465	326	59.17	1	138	138	25.045	466	551	86	15.608	139-465	
CCDC173	NP_001078916	*Homo sapiens*	552	148	456	308	55.80	1	147	147	26.63	457	552	96	17.391	148-456	147-482
CCDC173	XP_001441738	*Paramecium tetraurelia*	473	113	435	322	68.08	1	112	112	23.679	436	473	38	8.0338	113-435	113-441
CCDC173	CDQ57194	*Oncorhynchus mykiss*	548	139	454	315	57.48	1	138	138	25.182	455	548	94	17.153	139-454	
CCDC173	XP_001098455	*Macaca mulatta*	567	217	485	268	47.27	1	216	216	38.095	486	567	82	14.462	217-485	216-390
CCDC173	XP_003133517	*Sus scrofa*	552	147	468	321	58.15	1	146	146	26.449	469	552	84	15.217	147-468	146-481
CCDC173	NP_001071152	*Mus musculus*	547	142	453	311	56.86	1	141	141	25.777	454	547	94	17.185	142-453	141-475
CCDC173	NP_001120958	*Rattus norvegicus*	546	142	463	321	58.79	1	141	141	25.824	464	546	83	15.201	142-463	
CCDC173	XP_004942810	*Gallus gallus*	674	277	590	313	46.44	1	276	276	40.95	591	674	84	12.463	277-590	276-611
CCDC173	CCI42766	*Albugo candida*	468	99	430	331	70.73	1	98	98	20.94	431	468	38	8.1197	99-430	
CCDC173	XP_002899662	*Phytophthora infestans*	468	99	443	344	73.50	1	98	98	20.94	444	468	25	5.3419	99-443	
CCDC173	XP_001015372	*Tetrahymena thermophila*	583	223	561	338	57.98	1	222	222	38.079	562	583	22	3.7736	223-561	
CCDC173	XP_002947389	*Volvox carteri*	465	102	429	327	70.32	1	101	101	21.72	430	465	36	7.7419	102-429	102-449
CCDC173	XP_002934659	*Xenopus tropicalis*	553	142	341	199	35.99	1	141	141	25.497	342	553	212	38.336	142-341	141-367
FAP98	XP_001692676	*Chlamydomonas reinhardtii*	522	191	491	300	57.47	1	190	190	36.398	492	522	31	5.9387		185-496
FAP98	XP_002957631	*Volvox carteri*	521	219	490	271	52.02	1	218	218	41.843	491	521	31	5.9501		
FAP241	XP_001701817	*Chlamydomonas reinhardtii*	357	54	316	262	73.39	1	53	53	14.846	317	357	41	11.485	54-316	42-332
FAP241	XP_002948551	*Volvox carteri*	365	163	364	201	55.07	1	162	162	44.384	365	365	1	0.274	163-364	
Unknown	XP_966290	*Plasmodium falciparum*	396	99	375	276	69.70	1	98	98	24.747	376	396	21	5.303	99-375	85-396

Appendix 8: Synonyms

Across the literature of eukaryotic organisms the TPH domain-containing proteins have been ascribed several names. Table 25 summarise the difference names.

Table 25: TPH domain-containing protein nomenclature

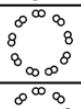
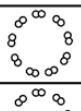
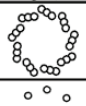
Name used in this thesis	Synonym 1	Synonym 2	Synonym 3
CCDC19	NESG1	FAP45	CFAP45
CCDC11	HTX6	FAP53	CFAP53
MSN1	SPATA40	FAP127	
CCDC173	C2orf77 (Hs)	FAP210	
Trichoplein	TCHP	Mitostatin	TpMs
FAP98			
FAP241			

Appendix 9: Centriole and cilia structure

Table 26: Centriole and cilia structures

For the reference organisms used in this study the organisation of centriole and cilia microtubules is illustrated below. Absence of a centriole or cilium in indicated by a black line (-). If a centriole, motile cilia or immotile cilia has been documented in an organisms this is indicated by a cartoon of the structure. If the presence/absence of a structure is unknown this is indicated with '?'. Structures were collated from previous bioinformatics studies (Carvalho-Santos *et al.*, 2011; Carvalho-Santos *et al.*, 2010; Hodges *et al.*, 2010; Hodges *et al.*, 2011; Wickstead and Gull, 2007; Wickstead and Gull, 2011).

Organism	Centriole/ Basal body	Motile cilia	Immotile cilia
Acanthamoeba castellani	—	—	—
Albugo candida			—
Anopheles gambiae			
Apis mellifera			
Arabidopsis thaliana	—	—	—
Batrachochytrium dendrobatidis			—
Bombyx mori			
Caenorhabditis elegans		—	
Chlamydomonas reinhardtii			—
Ciona intestinalis			
Cryptosporidium parvum		—	—
Danio rerio			
Dictyostelium discoideum	—	—	—
Drosophila melanogaster			
Eimeria maxima			—

Gallus gallus			
Giardia lamblia			—
Homo sapiens			
Leishmania major			
Macaca mulatta			
Mus musculus			
Naegleria gruberi			—
Neospora caninum			—
Oncorhynchus mykiss			
Paramecium tetraurelia			—
Phytophthora infestans			—
Plasmodium falciparum			—
Rattus norvegicus			
Saccharomyces cerevisiae	—	—	—
Schizosaccharomyces pombe	—	—	—
Stegodyphus mimosarum			
Sus scrofa			
Tetrahymena thermophila			—
Thalassiosira pseudonana	?		—
Toxoplasma gondii			—
Tribolium castaneum			
Trypanosoma brucei			—
Volvox carteri			—
Xenopus tropicalis			

Appendix 10: TritrypDB searches

Searching within TriTrypDB using the search parameter 'Trichohyalin' brings back a list of results from across *Trypanosoma spp.*

Gene ID	Organism	Genomic Location	Product Description
Tb427tmp.01.3320	*T. brucei Lister strain 427*	Tb427_11_01_v4: 3,018,715 - 3,020,691 (+)	Trichohyalin, putative
Tb927.11.11480	*T. brucei TREU927*	Tb927_11_v5: 3,073,596 - 3,077,460 (+)	Trichohyalin, putative
Tbg972.11.12890	*T. brucei gambiense DAL972*	Tbg972_11: 3,048,579 - 3,050,555 (+)	Trichohyalin, putative
TcCLB.509595.70	*T. cruzi CL Brener Esmeraldo-like*	TcChr30-S: 175,271 - 177,124 (-)	Trichohyalin, putative
TcCLB.511805.20	*T. cruzi CL Brener Esmeraldo-like*	TcChr30-S: 536,027 - 538,093 (+)	Trichohyalin, putative
TcCLB.508817.160	*T. cruzi CL Brener Non-Esmeraldo-like*	TcChr30-P: 174,782 - 177,061 (-)	Trichohyalin, putative
TcCLB.511015.30	*T. cruzi CL Brener Non-Esmeraldo-like*	TcChr30-P: 532,510 - 533,880 (+)	Trichohyalin, putative
TCSYLVIO_002528	*T. cruzi Sylvio X10/1*	ADWP02012599: 10,259 - 12,535 (-)	Trichohyalin, putative
TCSYLVIO_005618	*T. cruzi Sylvio X10/1*	ADWP02015175: 3,538 - 4,728 (-)	Trichohyalin, putative
Tc_MARK_1262	*T. cruzi marinkellei strain B7*	TcMARK_CONTIG_1432: 13,790 - 15,877 (-)	Trichohyalin, putative
Tc_MARK_1412	*T. cruzi marinkellei strain B7*	TcMARK_CONTIG_1441: 24,425 - 25,201 (+)	Trichohyalin, putative
TvY486_1112510	*T. vivax Y486*	TvY486_11: 3,264,083 - 3,265,510 (+)	Trichohyalin, putative

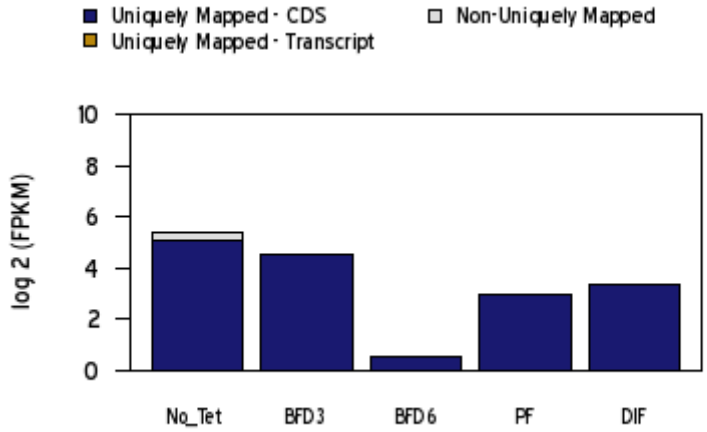
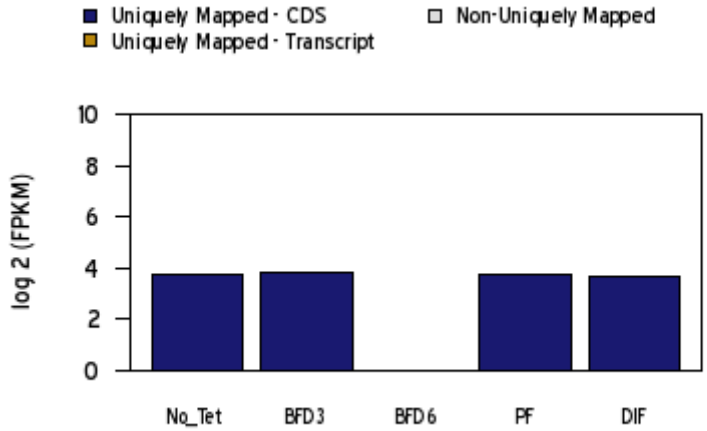
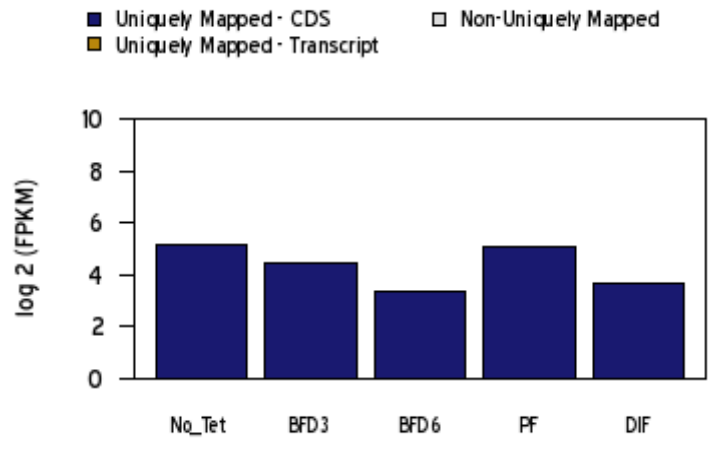
Searching within TriTrypDB using the search parameter 'Plectin' brings back a list of results from within *T. cruzi*, which do not have synteny to *T. brucei*

Gene ID	Organism	Product Description
TcCLB.503897.110	*T. cruzi CL Brener Esmeraldo-like*	plectin-like protein, putative
TcCLB.509561.80	*T. cruzi CL Brener Non-Esmeraldo-like*	plectin-like protein, putative
TCSYLVIO_009961	*T. cruzi Sylvio X10/1*	plectin-like protein, putative
Tc_MARK_9763	*T. cruzi marinkellei strain B7*	plectin-like protein, putative
Tc_MARK_9764	*T. cruzi marinkellei strain B7*	plectin-like protein, putative

Appendix 11: RNAi screen result

Table 27: High-throughput RNAi target sequence detection

The results of sequencing a non-clonal population of *T. brucei* before (No_Tet) and after treatment with tetracycline. BFD3; Bloodstreamform day 3, BFD6; Bloodstreamform day 6, PF; Procyclic form, DIF; differentiation. Data published (Alsford *et al.*, 2011) and graphs taken from TriTrypDB.

Gene	RNAi Phenotype
Tb927.8.4580 aka TbCCDC19	
Tb927.6.4520 aka TbMNS1	
Tb927.5.1230 aka TbCCDC11	

Appendix 12: Patient mutation 1

>gi|94721321|ref|NP_659457.2| coiled-coil domain-containing protein 11 [Homo sapiens]
MYSQRFGTVQREVKGPTPKVVIVRSKPPKGQGAEHHLERIRRSHQKHNAILASIKSSERDRLKAEWDQHNDCK
ILDSLVRARIKDAVQGFIINIEERRNKLRELLALEENEYFTEMQLKKETIEEKKDRMREKTKLLKEKNEKERQ
DFVAEKLDQQFRERCEELRVELLSIHQKKVCEERKAQIAFNEELSRQKLVEEQMFSKLWEEDRLAKEKREAQE
ARRQKELMENTRLGLNAQITSIKAQRQATQLLKEEEARLVESNNAQIKHENEQDMLKKQKAKQETRTILQKAL
QERIEHIQQEYRDEQDLNMKLVQRALQDLQEEADKKKQKREDMIREQKIYHKYLAQRREEEKAQEKEFDRILE
EDKAKKLAEKDKELRLEKEARRQLVDEVMCTRKLQVQEKLQREAKEQEERAMEQKHINESLKELNCEEKENFA
RRQRLAQEYRKQLQMQIAYQQQSQEAEKEEKRREFEAGVAANKMCLDKVQEVLSTHQVLPQNIHPMRKACPSK
LPP

>Truncated CCDC11 of patient OP-1069-II1
MYSQRFGTVQREVKGPTPKVVIVRSKPPKGQGAEHHLERI*

The nonsense mutation of patient OP-1069-II1 (Narasimhan *et al.*, 2014) caused ARG41*, therefore terminating the protein. ARG41 is highlighted in red; the nonsense mutation (*) is highlighted in green.

Appendix 13: Patient mutation 2

>gi|94721321|ref|NP_659457.2| coiled-coil domain-containing protein 11 [Homo sapiens]
MYSQRFGTVQREVKGPTPKVVIVRSKPPKGQGAEHHLERIRRSHQKHNAILASIKSSERDRLKAEWDQHNDCK
ILDSLVRARIKDAVQGFIINIEERRNKLRELLALEENEYFTEMQLKKETIEEKKDRMREKTKLLKEKNEKERQ
DFVAEKLDQQFRERCEELRVELLSIHQKKVCEERKAQIAFNEELSRQKLVEEQMFSKLWEEDRLAKEKREAQE
ARRQKELMENTRLGLNAQITSIKAQRQATQLLKEEEARLVESNNAQIKHENEQDMLKKQKAKQETRTILQKAL
QERIEHIQQEYRDEQDLNMKLVQRALQDLQEEADKKKQKREDMIREQKIYHKYLAQRREEEKAQEKEFDRILE
EDKAKKLAEKDKELRLEKEARRQLVDEVMCTRKLQVQEKLQREAKEQEERAMEQKHINESLKELNCEEKENFA
RRQRLAQEYRKQLQMQIAYQQQSQEAEKEEKRREFEAGVAANKMCLDKVQEVLSTHQVLPQNIHPMRKACPSK
LPP

TPH domain (start of)Exon 5 (start of)Exon 7

Sequence annotation taken from the clinical report (Perles *et al.*, 2012). The patient had a mutation in the coding sequence for CCDC11. This lead to Exon 6 being absent, causing a frame shift and therefore Exon 7 (red) was out of frame and it created a premature stop codon. This mutation is within the TPH domain (blue).

Appendix 14: Bioinformatics of centrin family in *T. brucei*

In the trypanosome field there are two nomenclature systems for centrins. In this thesis the one published in the genome and TrytriDB (Aslett *et al.*, 2010; Berriman *et al.*, 2005) is used (Table 28). This system is the most widely accepted (Shi *et al.*, 2008; Wang *et al.*, 2012). This is a different system to the one used by (Selvapandiyan *et al.*, 2004; Selvapandiyan *et al.*, 2001; Selvapandiyan *et al.*, 2007; Selvapandiyan *et al.*, 2012).

Table 28: Accession numbers of centrin

T. brucei centrin protein accession numbers as assigned on TriTrypDB (Aslett *et al.*, 2010) and in the genome (Berriman *et al.*, 2005).

TbCen1	Tb927.4.2260
TbCen2	Tb927.8.1080
TbCen3	Tb927.10.8710
TbCen4	Tb927.7.3410
TbCen5	Tb927.11.13900

Centrins are highly conserved among eukaryotes (Azimzadeh and Bornens, 2005) and there are 5 paralogs of centrin conserved among kinetoplastids (Berriman *et al.*, 2005). The exception to this is *T. cruzi,* which has two duplicates of centrin 2 and four duplicates on centrin 4 (Berriman *et al.*, 2005), these extra copies were not seen in other kinetoplastid species. An alignment of the 5 centrin proteins from 8 kinetoplastid species (*T. brucei, T. vivax, T. cruzi, T.grayi, T. congolense, C. fasiculata, L. major* and *L. Mexicana)* shows a high level of conservation at the amino acid level (Figure 111). When the aligned sequences were processed into a phylogenetic tree, the 5 centrin groups are identifiable (Figure 112). The work in this chapter will focus on centrin 2 (red group; Figure 112) from *T. brucei*, known as TbCen2.

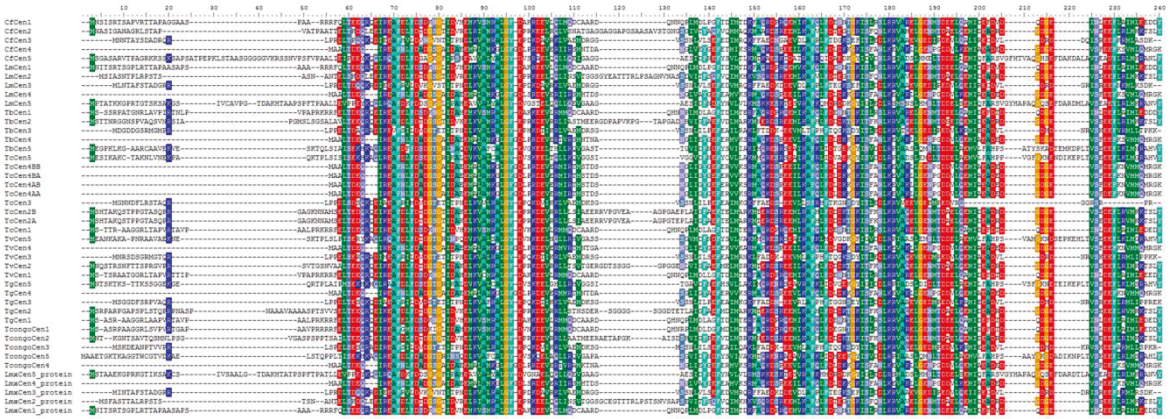


Figure 111: Alignment of kinetoplastid centrin proteins

Centrin protein sequences from 8 kinetoplastid species were used for the alignment; T. brucei, T. vivax, T. cruzi, T.grayi, T. congolense, C. fasiculata, L. major and L. Mexicana

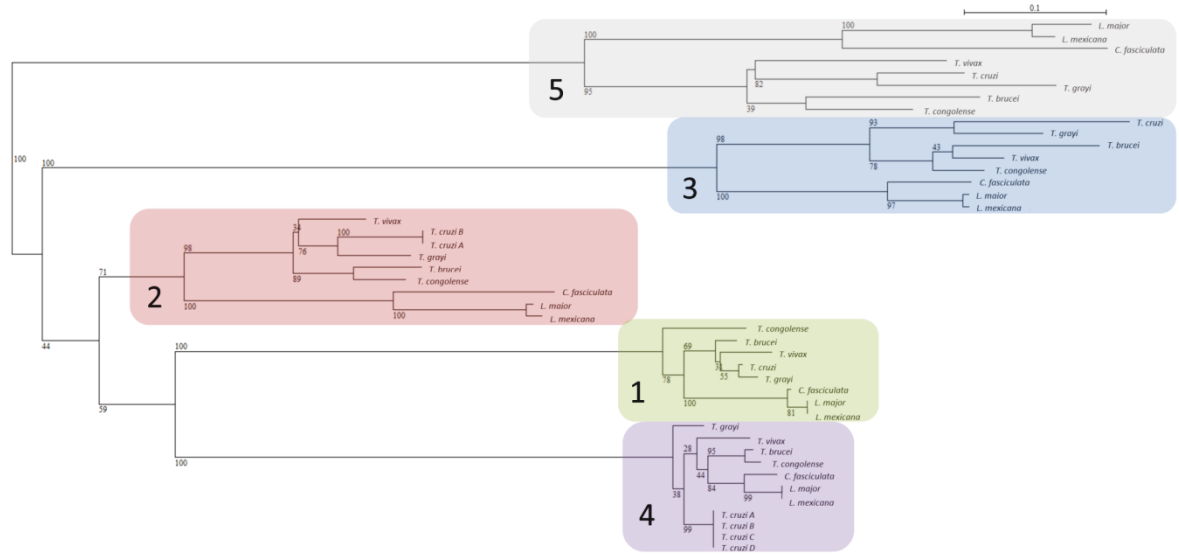


Figure 112: Phylogenetic analysis of kinetoplastid centrins.

A maximum likelihood phylogenetic tree of centrin proteins in selected kinetoplastid species (*T. brucei, T. vivax, T. cruzi, T.grayi, T. congolense, C. fasiculata, L. major* and *L. Mexicana*). Protein sequences were sourced from TriTrypDB (Aslett et al., 2010). Aligned sequences were bootstrapped with 100 replicates, bootstrap values are diaplyed at the branch nodes. Scale bar = 0.1 aa.

Appendix 15: published RNAi of TbCen2

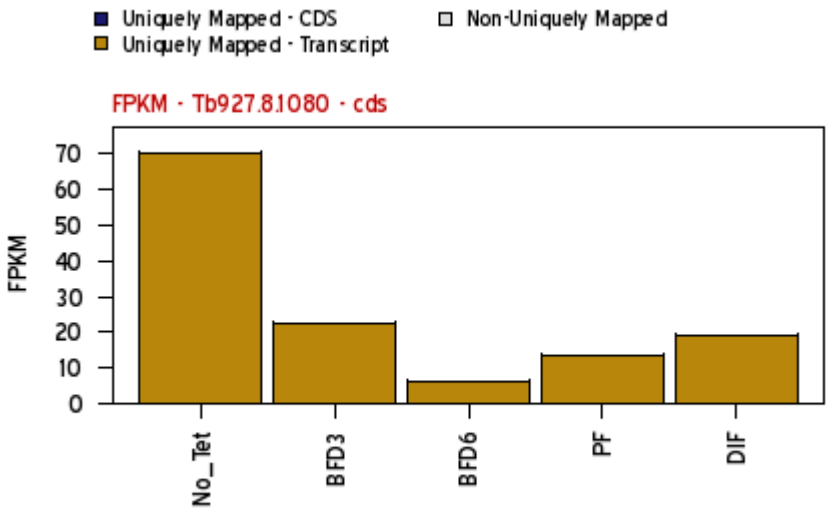


Figure 113: RNAi target sequence detection of TbCen2

The results of sequencing a non-clonal population of *T. brucei* before (No_Tet) and after treatment with tetracycline. BFD3; Bloodstreamform day 3, BFD6; Bloodstreamform day 6, PF; Procyclic form, DIF; differentiation. Data published (Alsford *et al.*, 2011) and graphs taken from TriTrypDB.

Appendix 16: SBFSEM 1K1N cell movie

Video 'Cell B.mpg' of rendered model from 1K1N procyclic cell. It is the same cell as in Figure 93 (i). The original SBFSEM data that the cell was segmented from is also shown. The cell has one kinetoplast and one nucleus (DNA containing organelles, blue) and therefore is also referred to as a 1K1N cell. There is a single flagellum (purple) with a single pair of basal bodies (green). The cell has a single, elaborate mitochondrion (dark orange), which occupies space along the length of the cell body. Most of the organelles within *T. brucei* are single copy; exceptions acidocalcisomes (white).

Appendix 17: SBFSEM DIVK1N cell movie

Video of rendered data from a procyclic cell with a dividing kinetoplast. File is called 'dividing kinetoplast.mpg'. Organelles share the same colour scheme as Appendix 1 with the addition of a new flagellum (red). It is the same cell as shown in Figure 93 (iii).

Appendix 18: TbHydin localisation

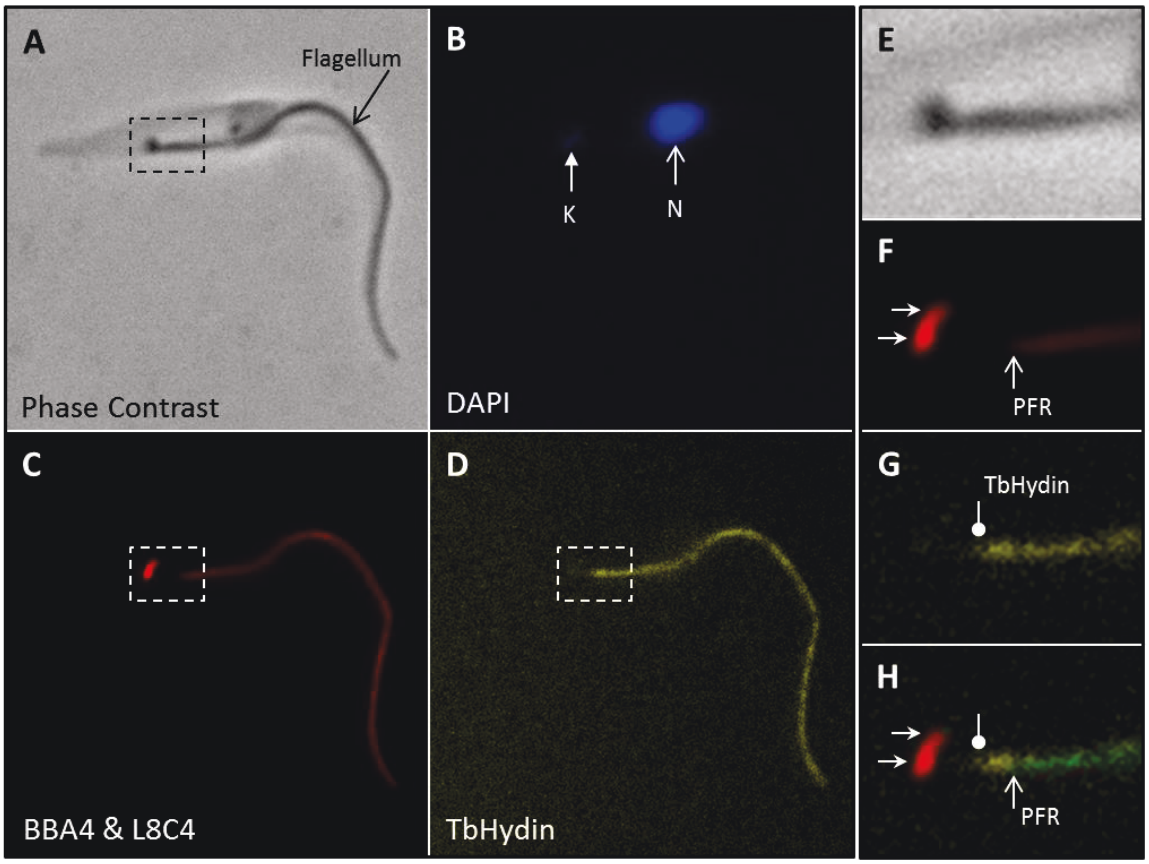


Figure 114: Localisation of YFP::TbHydin

TbHydin was tagged at the N terminus with YFP (YFP::TbHydin; D; yellow) and expressed from the endogenous locus in procyclic *T. brucei*. Cells were detergent extracted, fixed, labelled with DAPI and immunolabelled with BBA4 and L8C4 to stain the basal bodies and PFR, respectively. (A) An example cell is shown with a single flagellum. (B) 1 kinetoplast (K) and 1 nucleus (N). (C) both BBA4 and L8C4 immunolabelling is shown. (D) The localisation pattern of TbHydin.

Insets (E-H) show the proximal end of the flagellum from the 1K1N cell. (E) Phase contrast image of the basal bodies and the flagellum. (F) Immunolabelling of the basal bodies (solid arrowheads) with BBA4 (Woods *et al.*, 1989b) and the PFR (open arrow head) with L8C4 (Kohl *et al.*, 1999). (G) YFP::TbHydin signal localised to the flagellum (round arrow head). (H) Z-projection of the standard deviation between the BB/PFR (F) and TbHydin (G). Shared pixels are shown in green (H). Data courtesy of Dr Sam Dean.

List of publications and abstracts arising from thesis

1. Screening ciliopathy genes in the model *Trypanosoma brucei*

Poster

First International Cilia in Development and Disease Scientific Conference, May 2012

2. Functional analysis of genes implicated in ciliopathies

Oral presentation

Oxford Brookes University postgraduate symposium, January 2013

3. The global view: characterisation of organelle segregation and distribution throughout the *T. brucei* bloodstream form cell cycle

Oral presentation (By Dr Sue Vaughan)

5th Kinetoplastid Molecular Cell Biology Meeting (KMCBM), April 2013

4. TPHD proteins are localised to the flagellum of *Trypanosoma brucei*

Poster

8th French Meeting on Cilia, Flagella and Centrioles, October 2013

5. Functional analysis of genes implicated in ciliopathies

Poster

Oxford Brookes University postgraduate symposium, January 2014

6. New insights into the role of TbCentrin2 in basal body and flagellum biogenesis

Poster

British society of Parasitology, April 2014

7. Investigation of organelle distribution in the insect form of the protozoan parasite *Trypanosoma brucei*

Poster

Microscience Microscopy congress, July 2014

Copies of published material

REMOVED FROM ELECTRONIC VERSION

1. Screening ciliopathy genes in the model *Trypanosoma brucei*

Barry, Samantha (Oxford Brookes University); Hughes, Louise (Oxford Brookes University); Towers, Katie (Oxford Brookes University); Vaughan, Sue (Oxford Brookes University).

We have carried out a bioinformatics study to search for novel proteins of the flagellum/cilium. A total of 26 candidate hypothetical genes were established from previously published studies including a flagellar proteome (Broadhead *et al.*, 2006) and an RNAi study of motility mutants of *Trypanosoma brucei* (Baron *et al.*, 2007b). All candidates are predicted to have orthologs in the human genome, the dysregulation of which is associated with or predicted to be involved in at least one ciliopathy.

The protozoan parasite *T. brucei* is a well-established experimental model to study defects in flagellum assembly and function. In this project we will confirm that the proteins from the candidate genes are localised to the flagellum or basal body by GFP-tagging. The function of each will be assessed using inducible RNAi methods and in future work will involve taking the work forward into human cell lines for some candidate genes to see if we can reproduce the same phenotype in the cilia of human cells.

Abstract published: (Barry *et al.*, 2012).

2. Functional analysis of genes implicated in ciliopathies

Barry, Samantha (Oxford Brookes University); Vaughan, Sue (Oxford Brookes University)

We study the structure and function of cilia/flagella in order to understand human diseases resulting from cilia/flagella dysfunction. This class of diseases are referred to as ciliopathies. We use the well-established model organism *Trypanosoma brucei*, a protozoan parasite responsible for causing Human African Trypanosomiasis (HAT).

We carried out a bioinformatics search for uncharacterised proteins within the *T. brucei* flagellar proteome which are conserved within humans and predicted to play a role within ciliopathic diseases, 26 candidate proteins were identified using these criteria. To study the function of these proteins inducible overexpression and RNAi constructs have been produced, proteins will also be localised by creating endogenous YFP fusions.

One of our target genes is the trypanosomal ortholog of CCDC19 (aka NESG1) implicated in nasopharyngeal carcinoma. Overexpression of this protein leads to morphological defects in *T. brucei* and RNAi knock down leads to reduced growth of the population in culture and organelle inheritance defects.

Future work will involve motility assays and ultrastructural studies within trypanosomes. We also aim to conduct work in mammalian cell lines to enable a comparison between phenotypes observed in *T. brucei* flagella and mammalian cilia.

3. The global view: characterisation of organelle segregation and distribution throughout the *T. brucei* bloodstream form cell cycle

Vaughan, Sue (Oxford Brookes University); Hughes, Louise (Oxford Brookes University);Towers, Katie (Oxford Brookes University); Barry, Samantha (Oxford Brookes University);Starborg,Tobias (University of Manchester); Gull, Keith (University of Oxford).

Serial block face scanning electron microscopy (SBFSEM) is a new technique that enables visualization of large volumes containing numerous cells at different cell cycle stages. This technique provides a whole cell view whilst at the same time revealing internal ultrastructural detail of organelle morphology and distribution. The T. brucei cell has a long tubular shape with a single flagellum. Some organelles, present as single copy or in low copy number are asymmetrically localised to certain regions of the cell such as the flagellum, flagellar pocket, basal body, kinetoplast, nucleus and Golgi. The mitochondrion and endoplasmic reticulum are distributed throughout the cell body, as are multi-copy organelles such as glycosomes and acidocalciosomes. Whole cell volumes at different stages of the cell division cycle were extracted and all major organelles and cellular structures were modelled and quantified. This allowed a complete view of the temporal order and spatial positioning in which each organelle duplicated, segregated or assembled. New flagellum length, kinetoplast and nuclear division, acknowledged cell cycle stage markers, could now be placed in context with other cell organelles. Our data included the discovery of novel architecture at the tip of the growing new flagellum, further extension of our understanding of mitochondrial division and new features of nuclear division. Our 3D global modelling provides cartographic snapshots combining spatial information with quantitative volumetric analyses of all major organelles across the cell division cycle.

4. TPH domain-containing proteins are localised to the flagellum of *Trypanosoma brucei*

Barry, Samantha (Oxford Brookes University); Vaughan, Sue (Oxford Brookes University)

The protozoan parasite *T. brucei* is a well-established experimental model to study defects in flagellum assembly and function. The published *Trypanosoma brucei* flagellar proteome was interrogated to select proteins conserved between *T. brucei* and *Homo sapiens* and therefore potentially involved in ciliopathy diseases. Using this approach we have identified a novel family of highly conserved proteins with a shared 'trichohyalin-plectin homology' (TPH) domain. The proteomes of *Homo sapiens* and the green algae, *Chlamydomonas reinhardtii* each possess six members of this family whilst *Trypanosoma brucei* has 3 proteins belonging to this TPHD family. Analysis of published proteomes reveals that TPH domain-containing proteins are absent from eukaryotes with immotile cilia, for example, *Caenorhabditis elegans* and are only present in organisms with motile cilia/flagella which suggests a link between the TPH domain and motility. Using a YFP-tagging method we have localised the 3 TPH domain-containing proteins to the flagellum of *T. brucei* for the first time. We found that these proteins localised along the whole length of the flagellum and therefore suggest that they are associated with the axoneme. As part of our future investigations, the function of each TPH domain-containing protein will be assessed using inducible RNAi methods.

5. Functional analysis of genes implicated in ciliopathies

Barry, Samantha (Oxford Brookes University); Vaughan, Sue (Oxford Brookes University).

We study the structure and function of cilia/flagella in order to understand human diseases resulting from cilia/flagella dysfunction. This class of diseases are referred to as ciliopathies. We use the well-established model organism *Trypanosoma brucei*, a protozoan parasite responsible for causing Human African Trypanosomiasis (HAT).

The published *Trypanosoma brucei* flagellar proteome was interrogated to select proteins conserved between *T. brucei* and *Homo sapiens* and therefore potentially involved in ciliopathy diseases. Using this approach we have identified a novel family of highly conserved proteins with a shared 'trichohyalin-plectin homology domain' (TPHD), the role of this domain is as of yet uncharacterised. The proteomes of *Homo sapiens* and the green algae, *Chlamydomonas reinhardtii* each possess six members of this family whilst *Trypanosoma brucei* has 3 proteins belonging to this TPHD family. Analysis of published proteomes reveals that TPHD proteins are absent from eukaryotes with immotile cilia, for example, *Caenorhabditis elegans* and are only present in organisms with motile cilia/flagella. This suggests a link between TPHD proteins and motility. Using a YFP-tagging method we have localised the 3 TPHD proteins of *T. brucei* to the flagellum for the first time. We found that these proteins localised along the whole length of the flagellum and therefore suggest that they are associated with the axoneme. As part of our future investigations, the function of each TPHD protein will be assessed using inducible RNAi methods.

6. New insights into the role of TbCentrin2 in basal body and flagellum biogenesis

Barry, Samantha (Oxford Brookes University); Towers, Katie (Oxford Brookes University); Vaughan, Sue (Oxford Brookes University).

The single motile flagellum of *Trypanosoma brucei* is essential for pathogenicity and is assembled from a microtubule-based cylindrical structure called a basal body. During the cell division cycle a new flagellum is assembled alongside the old flagellum and assembly requires duplication and segregation of basal bodies. Basal bodies exist as a pair with a defined 'age'. The mature basal body can extend a flagellum, but the pro-basal body which is positioned alongside cannot extend a flagellum until the next cell cycle after it was assembled. During the cell division cycle the pro-basal body matures and can assemble the new flagellum and two new pro-basal bodies form alongside each mature basal body. Centrin is a calcium-binding protein which has been localised to a number of cytoskeletal structures including the basal body and flagellum in eukaryotic organisms. There are 5 centrin proteins in *Trypanosoma brucei*. Centrin 2 (TbCen2) has been localised to basal bodies, bilobe and flagellum and was found to be required for duplication of the bilobe.

Localisation studies of 20H5 which recognises Centrin 1 and 2 in *T. brucei* revealed a distinct pattern of labelling corresponding to the 'age' of each basal body. This pattern of labelling was also demonstrated by localisation studies of endogenous YFP::TbCen2 fusion in addition to localisation at the bilobe and flagellum. We also discovered defects in assembly of central pair microtubules of the axoneme following knockdown of TbCen2 by RNAi. These results give further insights into the multi-functional role of this family of proteins in eukaryotic cells.

7. Investigation of organelle distribution in the insect form of the protozoan parasite *Trypanosoma brucei*

Barry, Samantha (Oxford Brookes University); Hughes, Louise (Oxford Brookes University); Starborg, Tobias (University of Manchester); Vaughan, Sue (Oxford Brookes University).

We have used SBFSEM (Denk and Horstmann, 2004) to investigate the relationship between organelles during the cell cycle of Trypanosoma brucei, an infective protozoan parasite that survives in the bloodstream of the mammalian host. Here we present data from the insect stage form of the parasite which is transmitted by an insect vector between hosts. The environment in the insect gut is very different to that of the mammalian bloodstream, our data shows what organelle based adaptations the parasite undergoes to survive and proliferate in both of these difference environments. We have already shown using SBFSEM that the bloodstream form of the parasite has a unique membrane invagination to protect the distal tip of the growing flagellum (Hughes *et al.*, 2013) and here we investigate the lifecycle stage specific cell biology of the insect form. Insect form 29:13 T. brucei cells were fixed and processed as previously described for SBFSEM (Hughes *et al.*, 2013). Data were collected on a Quanta FEG 250 SEM (FEI) fitted with a Gatan 3View system. Briefly, the cells were fixed with 2.5% glutaraldehyde, 2% paraformaldehyde and 0.1% tannic acid in 0.1 M phosphate buffer (pH 7.0) for 2 hours at room temperature. Cells were post-fixed in 1% osmium tetroxide in the same buffer for 1 hour before being stained en bloc for 40 minutes in 2% uranyl acetate, dehydrated in an ascending acetone series and embedded in Agar 100 resin. Block face images were recorded using digital micrograph (Gatan) and assembled into stacks using ImageJ (NIH). Data segmentation was performed using Amira (FEI). Cell cycle markers by light microscopy are well known but distinct ultrastructural changes across a whole cell volume in three dimensions has not been characterised in T. brucei. Duplication and segregation of individual organelles such as the flagellum, nucleus and kinetoplast (trypanosomatid specific mitochondrial genome) are relatively well known but the timing of other discrete organelles including the Golgi apparatus and mitochondrion lack characterisation. SBFSEM allows the opportunity to tie in discrete events in organelle duplication with the length of the growing new flagellum. This provides a detailed series of events culminating in cytokinesis. We have used serial block face scanning electron microscopy (SBFSEM) to examine the organelle morphology and distribution within the cell during the course of

the insect stage cell cycle. We have modelled major organelles and structures including the lifecycle stage specific structure, the flagella connector (Briggs *et al.*, 2004b). Authors gratefully acknowledge funding from the BBSRC under grant number BB/1000402/1.