Initiation of Nuclear DNA Replication in *Trypanosoma brucei* and *Leishmania*

Catarina de Almeida Marques

BSc, MSc

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

Wellcome Trust Centre for Molecular Parasitology
Institute of Infection, Immunity and Inflammation
College of Medical, Veterinary and Life Sciences

University of Glasgow

April 2015

Abstract

Replication of the genome is a central process in cellular life, which must be tightly regulated at the risk of genomic instability. DNA replication has been extensively studied in bacteria and eukaryotes, and recently, in archaea. Universally, DNA replication is started at specific genomic sites termed origins of replication, which are recognised by an initiator factor. While in both bacteria and archaea the initiator factor is a single protein, it is assumed that this role is performed in eukaryotes by a highly conserved six-subunit origin recognition complex (ORC). Recent phylogenetic studies, however, suggest that the presence of a six-subunit ORC might not be as conserved as initially believed.

Trypanosoma brucei is a protozoan parasite in which little is known about nuclear DNA replication. To date, initiation of *T. brucei* DNA replication has been associated with a single factor, TbORC1/CDC6, though highly diverged interacting partners have been identified. To elucidate whether *T. brucei* possesses a diverged ORC-like complex, TbORC1/CDC6 and its known interacting partners, TbORC1B, TbORC4, Tb7980, Tb3120, and a novel factor, Tb1120, were analysed. First, the protein sequences of these factors were re-analysed, revealing varying degrees of conservation and divergence with other eukaryotes' ORC proteins. Second, expression silencing by inducible-RNA interference (RNAi) of TbORC1/CDC6, TbORC1B, TbORC4, and Tb3120, in procyclic form (PCF) and bloodstream form (BSF) cells, confirmed their involvement in DNA replication. Third, subcellular localisation and dynamics of TbORC1/CDC6 and its interacting partners during the cell cycle of PCF and BSF cells was investigated by immunofluorescence, revealing TbORC1B to be the sole factor to display an apparent cell cycle-dependent localisation pattern, perhaps suggesting that it might be a DNA replication regulatory factor. Finally, immunoprecipitation and gel filtration assays support the existence of an ORC-like complex, apparently large enough to be composed of TbORC1/CDC6 and known interacting partners, and potentially, additional factors.

TbORC1/CDC6-binding sites and origins of replication in *T. brucei* have been mapped in PCF cells. Like most eukaryotes, no specific sequence elements were found to define TbORC1/CDC6-binding sites or origins of replication, and the TbORC1/CDC6-binding sites outnumbered the mapped origins, which appeared to

be activated at different times during S phase. It has been reported in other eukaryotes that different cell types activate different origins or the activation timing differs. Here, origins of replication were mapped in *T. brucei* PCF and BSF cells, revealing a pronounced inflexibility in origin usage in these two life cycle stages. Only one, notable genome-wide difference was found: in BSF cells, the single active variant surface glycoprotein (VSG) expression site was found to be early replicating, whereas all other silent VSG expression sites were late replicating; in PCF cells however, where all VSG expression sites are silenced, these were all late replicating. These data reveal a locus-specific link between DNA replication and transcription in *T. brucei*, which may relate to immune evasion.

The genomes of *T. brucei* and related kinetoplastids are highly syntenic. Since most eukaryotic origins of replication are not defined by consensus DNA sequences, but appear instead to be defined by, among other features, chromatin context and status, origins were here mapped in both *L. major* and *L. mexicana* promastigotes in order to ask if common features could be found relative to *T. brucei*. Surprisingly, only a single origin could be found per *Leishmania* chromosome, in contrast with all eukaryotes examined to date, where each chromosome is replicated from multiple detectable origins. Origin-active loci in *Leishmania* were found to be distinguishable from related non-origin loci in terms of size, a characteristic not observed in *T. brucei*, although around 40% of the mapped origins are conserved in location relative to *T. brucei*. These data reveal pronounced differences in replication dynamics between the two genera, despite the considerable overlap in genome organisation.

Table of Contents

List of Tables

List of Equations

List of Figures

List of Accompanying Material

Supplemental CD-ROM containing:

Appendices. Sections, Figures and Tables are numbered as if the section was Chapter 7.

“To infinity… and beyond!”

Buzz Lightyear, from Toy Story, 1995, Disney, all rights reserved.

Acknowledgements

First of all, to Richard McCulloch, for being such a great supervisor. For accepting me as a PhD student out of the blue; for giving me this AMAZING project; for supporting and challenging me during these 4 years; for giving me independence to think and plan by myself; and all the fun at tea time.

To Fundação para a Ciência e Tecnologia (FCT), Portugal, for funding my PhD abroad, even after the country sunk into deep economical crisis.

To Calvin Tiengwe, who started this project during his PhD, and for being involved when he was already a Post-Doc somewhere else. Without him, this project wouldn't have gone this far. To Lucio Marcello for being my chaperone and mentor in the lab during my first year and introduce me to molecular biology, which was not really my area at the time. To Nick Dickens, obviously for all the help with bioinformatics, not just the origins, but for everything, and for teaching me very patiently how to programme. And Kathryn Crouch, for all the help with TriTryp.

To my assessors, Dave Barry, Liam Morrison, Tansy Hammarton, and Daniel Paape, for their assessments, suggestions, and constructive criticism over these years. A special thanks to Liam, who was my assessor throughout the 4 years.

To Alex MacKay, as well as Mulu Gedle and Margaret Agnew without whom I would honestly be lost in bureaucracy.

To everyone that helped me in terms of technical support and are mentioned throughout this thesis: to Prof. Markus Meissner, Dr Gurman Pall, and Carmen for the help with the Elyra microscope; to Amy Goundry for the *Leishmania* cells, as well as for teaching me how to use the Delta Vision microscope, together with Ben Cull and Sam Duncan; to Dr Richard Burchmore, Christina Naula, and Stefan Weidt at Glasgow Polyomics for the mass spectrometry analysis; to Alan Scott, Jaspreet and Jeziel for the gel filtration run and analysis; to Marko Prorocic, our molecular biology guru; and obviously, to Jim Scott for being the king technician in level 6, and Elaine Brown. And finally to Craig Lapsley, the coolest technician in the lab.

To all the great friends I made over these 4 years at the WTCMP (including everyone abovementioned). I'm sorry not to name you individually, but when I tried, it filled a whole page (and this thesis is already massive enough), but you know who you are, so thank you! A special thanks to Amy Goundry and Elizabeth Thomas, who were the first people I met and took me in not just as a colleague, but also as a friend. For all the adventures in the lab, office, and outside, these were 4 amazing years; they wouldn't have been the same without you. To Becky, for being such a good lab mate and friend, as well as thesis writing buddy across the Atlantic. To everyone that was in the level 6 end office during these 4 years, who were great office mates and friends, with whom I spent so many hours and had so much fun both in and outside of the office/lab. Also a special thanks to all the Portuguese and Brazilian friends I made over the years, even though they were here for short periods of time, it was very important to be able to speak my own language at work, and for making me feel like I never left Portugal at all!

And to my family, who always supported me and saw me leave to somewhere far away. Ao meu pai e à minha mãe, por serem tão espectaculares, e embarcarem nesta aventura comigo. À minha maninha Carol, que teve de aprender a ser filha única, quando nunca o foi, e de quem tenho sempre tantas saudades. E aos meus quatro avós, que me viram partir para longe já em tão avançada idade, e mesmo sem perceberem o que faço aqui, me apoiam sempre. E à família do Zé, pelos mesmos motivos.

And finally, to Zé, to whom I thank everyday for being always by my side.

Author's Declaration

I here declare that this thesis and the results herein presented are a result of my own work, except where otherwise stated and acknowledged. None of the results herein presented have been used previously to obtain a degree at any university.

Catarina de Almeida Marques

List of Abbreviations

aa	amino acid
ACS	autonomous consensus sequence
ARS	autonomous replication sequence
ADP	adenosine diphosphate
ATP	adenosine triphosphate
BAH	bromo-adjacent homology domain
BES	bloodstream VSG expression site
BIR	break induced replication
BLAST	Basic Local Alignment Search Tool
BLE	bleomycin resistance gene
bp	base pair
BrdU	bromo-5'-deoxyuridine
BSA	bovine serine albumin
BSD	blasticidin resistance gene
BSF	bloodstream form
CDK	cyclin-dependent kinase
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
Cl	clones
CL	cutaneous leishmaniasis
CMG	Cdc45, MCM2-7 and GINS, complex
CNS	central nervous system
co-IP	co-immunoprecipitation
CTD	C-terminal domain
DAPI	4',6-diamidino-2-phenylindole
DDK	Dbf4-dependent kinase
DGC	directional gene cluster
DIC	differential interference contrast
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
DTT	1,4-dithiothreitol
DUE	DNA-unwinding element
EdU	5-ethynyl-2'-deoxyuridine
EDTA	ethylenediaminetetraacetic acid
Ef1α	transcription elongation factor 1 α
ESAG	expression site associated gene
FACS	Fluorescence Activated Cell Sorting
FB	fractionation buffer
FITC	fluorescein isothiocyanate
gDNA	genomic DNA
GOI	gene of interest
h	hour(s)
H2I	helix-2 insert motif
HA	human influenza virus hemagglutinin epitope
HAT	Human African Trypanosomiasis
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	horseradish peroxidase
HTH	helix-turn-helix motif

HU	hydroxyurea
HYG	hygromycin phosphotransferase gene
IFA	immunofluorescence assay
IP	immunoprecipitation
ISM	initiator-specific motif
K	kinetoplast
kbp	Kilobase-pairs
kDa	Kilo-dalton
kDNA	kinetoplast DNA
LB	lysogeny broth
Mbp	Megabase-pairs
MCM	minichromosome maintenance
MES	metacyclic VSG expression site
MFA	Marker Frequency Analysis
MFA-seq	Marker Frequency Analysis by deep sequencing
MFA-qPCR	Marker Frequency Analysis by qPCR
min	minute(s)
mRNA	messenger RNA
MVSG	metacyclic variant surface glycoprotein
myc	c-myc epitope
N	nucleus
NEB	New England Biolabs
NEO	neomycin phosphotransferase gene
NLS	nuclear localisation signal
nLS- ES MS/MS	nanoflow liquid chromatography-electrospray tandem mass spectrometry
NTD	N-terminal domain
OPB	oligopeptidase B
ORC	origin recognition complex
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline buffer solution
PCF	procyclic form
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PDB	Protein Data Bank
PI	propidium iodide
pre-RC	pre-replication complex
PURO	puromycin-N-acetyltransferase gene
pre-IC	pre-initiation complex
qPCR	quantitative real-time PCR
RNA	ribonucleic acid
RNA Pol I	RNA polymerase I
RNA Pol II	RNA polymerase II
RNAi	RNA interference
ROI	region of interest
RT	reverse transcriptase
RT-qPCR	reverse transcription quantitative real-time PCR
SD	standard deviation
SDS	sodium dodecyl sulphate
sec	second(s)
SEM	standard error of the mean

S.O.C	super optimal broth supplemented with glucose
SR-SIM	super resolution structure illumination microscopy
ssDNA	single-stranded DNA
SSR	strand switch region
T7RNAP	T7 RNA polymerase
TAE	tris-acetate-EDTA
Tet	tetracycline
TetO	tetracycline operator
TetR	tetracycline repressor
TREU	Trypanosomiasis Research Edinburg University
TRITC	tetramethylrhodamine
TriTryps	*T. brucei*, *T. cruzi*, and *L. major*
UTR	untranslated region
UV	ultraviolet
VL	visceral leishmaniasis
VSG	variant surface glycoprotein
WCE	whole cell extract buffer
WHD	winged-helix domain
WHO	World Health Organisation
WTCMP	Wellcome Trust Centre for Molecular Parasitology

1 General Introduction

1.1 *Trypanosoma brucei* and related kinetoplastids

1.1.1 The kinetoplastids

Recent taxonomic studies divide the *Eukarya* domain into five supergroups (Adl *et al.*, 2012; Lukes *et al.*, 2014), as shown in Figure 1.1. In these classifications, kinetoplastid protists are included into the Excavata supergroup, group Discoba, subgroup Discicristata, phylum Euglenozoa, and class Kinetoplastea (Lukes *et al.*, 2014; Moreira *et al.*, 2004; Adl *et al.*, 2012) (Figure 1.1). As the name suggests, this class owes its name to the presence of a specific cellular structure, the kinetoplast (Moreira *et al.*, 2004; Vickerman, 1976), which is the cell's single mitochondrion DNA, organised as a mass of concatenated DNA circles (reviewed in Jensen and Englund, 2012; Liu *et al.*, 2005). This early-branching group (Dacks *et al.*, 2008) appears to be evolutionarily more ancestral than other major groups of protozoan parasites (Lukes *et al.*, 2014), such as the apicomplexans (which incorporates clinically important genera such as *Plasmodium* and *Toxoplasma*), and includes a wide range of ubiquitous organisms, from free-living species to obligatory parasites, and pathogens of invertebrates, vertebrates and plants (reviewed in Simpson *et al.*, 2006; Lukes *et al.*, 2014). Generally, kinetoplastids are seen as highly diverged eukaryotes, as they show unusual aspects of core biological processes, such as mitochondrial RNA editing, genomic organisation, mRNA *trans*-splicing, and compartmentalisation of glycolysis (reviewed in Simpson *et al.*, 2006), when compared to the canonical or model eukaryotes (e.g. yeasts, domestic fly, mice, and human), which belong to the Opisthokonta supergroup. The class Kinetoplastea has been divided into two subclasses (Moreira *et al.*, 2004), the early-branching Prokinetoplastina, and the Metakinetoplastina, which incorporates the order Trypanosomatida, containing exclusively parasitic and uniflagellate organisms (Adl *et al.*, 2012; Moreira *et al.*, 2004). It is within the order Trypanosomatida that disease-causing parasites from the dixeneous (have a vertebrate/plant host and an invertebrate host) genera *Trypanosoma* and *Leishmania* are placed (Adl *et al.*, 2012; Lukes *et al.*, 2014; Moreira *et al.*, 2004). These include pathogenic species of humans and other mammals that are of clinical and veterinary importance, and impose a great economic burden on the affected countries: *Trypanosoma brucei* (sleeping sickness disease in sub-Saharan Africa), *Trypanosoma cruzi* (Chagas disease in

South America) and the various *Leishmania* spp. (cutaneous and visceral leishmaniasis, affecting countries worldwide).

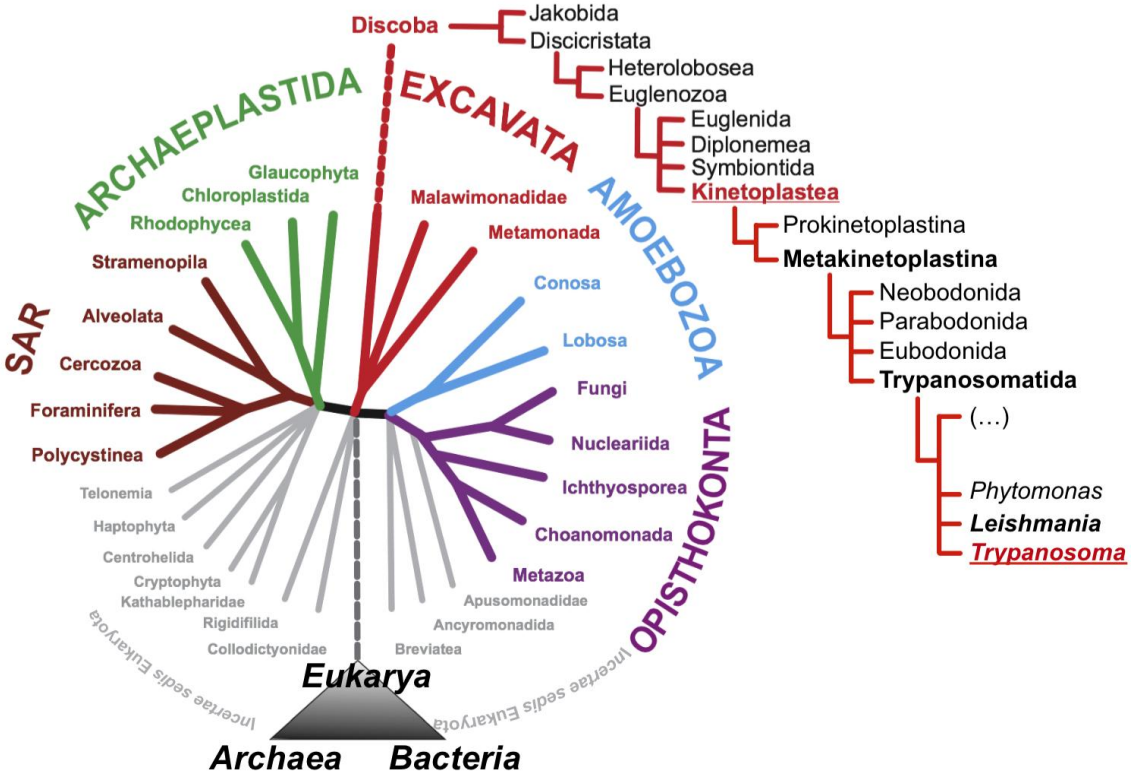


Figure 1.1. The phylogeny of the *Eukarya* domain.
Diagram showing a general overview of the eukaryote phylogeny based on the recent classification of the *Eukarya* domain by (Adl *et al.*, 2012). In this case, the taxonomic positions of the class Kinetoplastea, order Trypanosomatida, and the *Trypanosoma* and *Leishmania* genera are highlighted. Distances are not to scale. (…) represents the other genera within the order Trypanosomatida, but that are not relevant for the present work. SAR refers to the group containing the Stramenopiles, Alveolates, and Rhizaria groups, previously classified differently (Adl *et al.*, 2005). The groups in grey refer to organism presently classified as *incertae sedis*, as their inclusion into the other groups is not clear (Adl *et al.*, 2012). Figure adapted from (Lukes *et al.*, 2014), reproduced with permission (license number 3586570624161).

1.1.2 Human African Trypanosomiasis: past, present and future

Sleeping sickness disease, or human African trypanosomiasis (HAT), is caused by two subspecies of *Trypanosoma brucei*: *T. b. gambiense* and *T. b. rhodesiense*. Although generally called HAT, the two subspecies of *T. brucei* are considered to cause different diseases (Franco *et al.*, 2014), as the clinical pattern (see below) and epidemiology of the infections are distinct. *T. b. gambiense* causes a chronic form of HAT (gambiense HAT) in western and central Africa (Figure 1.2, A), accounting for 98% of the HAT cases reported in the last 10 years (Franco *et al.*, 2014). In contrast, *T. b. rhodesiense* is responsible for an acute form of the disease (rhodesiense HAT) in eastern and southern Africa (Figure 1.2, B), where

it occasionally infects humans; it is considered a zoonotic disease, infecting livestock and wildlife animals. Both subspecies of the parasite are transmitted to the mammalian host (see life cycle below) by infected tsetse flies, of the genus *Glossina* (reviewed in Fevre *et al.*, 2006), when taking a blood meal. The distribution of HAT reflects the regions of the African continent that are inhabited by the tsetse fly, which is delimited to most of mid-continental Africa, between the Sahara and Kalahari deserts (tsetse belt) (Figure 1.2) (Franco *et al.*, 2014; Simarro *et al.*, 2010), where it is estimated that around 70 million people are at risk (Simarro *et al.*, 2012; Franco *et al.*, 2014). Generally, working-age adults and young adults of rural areas are at higher risk of HAT, as many work in environments colonised by tsetse flies such as forests, savannah, and plantations (Franco *et al.*, 2014). Together with the highly morbidity due to the incapacitating and chronic nature of the disease, HAT has a devastating impact on the income-generating capacity of families, social acceptance and, overall, is responsible for a significant socio-economic burden on the affected countries (Franco *et al.*, 2014; Simarro *et al.*, 2010; Steverding, 2008; Fevre *et al.*, 2006). A third *T. brucei* subspecies, *T. b. brucei*, as well as the distinct species *T. congolense* and *T. vivax*, are not infective to humans, but infective to both wild and domestic animals, being responsible for veterinary trypanosomiasis, also known as nagana. While in wild animals nagana displays relatively mild symptoms, in domestic animals it results in severe and debilitating disease that is often fatal, rendering stock farming impractical in countries within the tsetse belt, thus adding substantially to the economic impact already caused by HAT (reviewed in Brun *et al.*, 2010; Steverding, 2008; Franco *et al.*, 2014).

It is thought that trypanosomes have been infecting humans since hominids first appeared in Africa, having most likely contributed to early hominid evolution (Steverding, 2008). Indeed, there are several reports throughout human history documenting HAT and nagana (Steverding, 2008). Although HAT epidemics during the early 20th century were devastating, implementation of successful control measures and adequate surveillance programs led to a highly efficient reduction in the cases of HAT by the 1960's where the landmark of <5,000 cases was reached (reviewed in Steverding, 2008; Franco *et al.*, 2014). However, subsequent socio-political instability in a large number of the affected countries led to the abandonment of these measures, and the consequent resurgent of

HAT in the 1980's, with an alarming increase in the number of estimated cases (>300,000) by the late 1990's (Steverding, 2008; Franco *et al.*, 2014; Brun *et al.*, 2010). Since 2000, joint efforts led and coordinated by the World Health Organisation (WHO) have been successful in reversing this trend: epidemiological data shows that in 2013 there were 6,228 new declared cases of gambiense HAT (although around 20,000 new cases were estimated) (Figure 1.2, A), and 98 reported cases of rhodesiense HAT (Figure 1.2, B), a reduction of 75% and 88%, respectively, since 2000 (Franco *et al.*, 2014). These results are encouraging, and gambiense HAT has been targeted for elimination by 2020 (WHO, 2012).

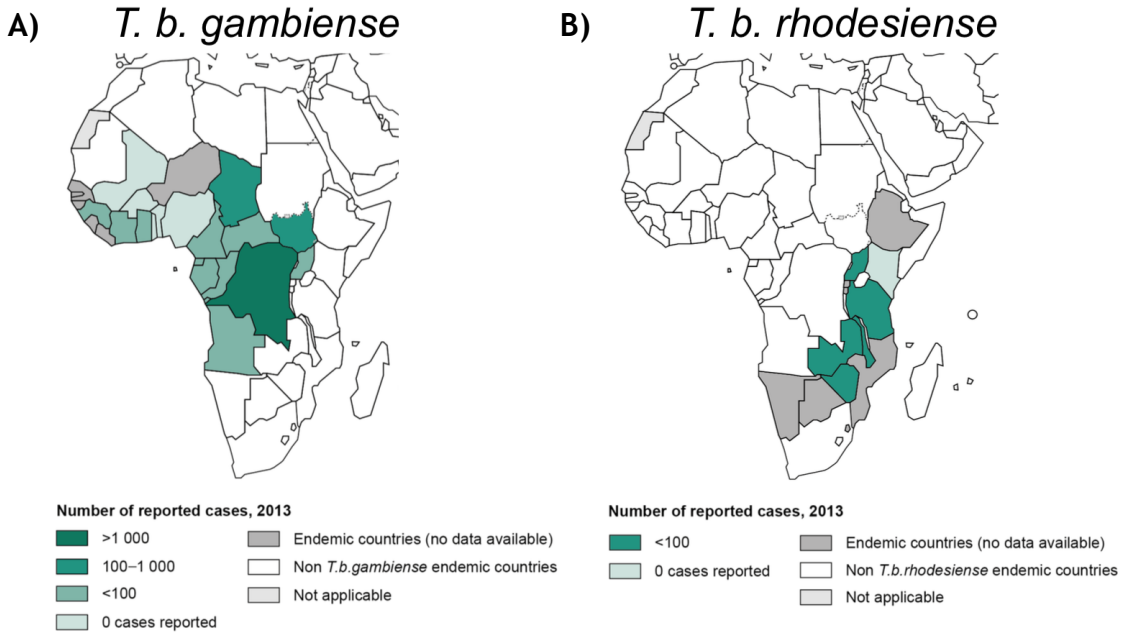


Figure 1.2. Epidemiological distribution of HAT as of 2013.
A) Distribution of gambiense HAT (caused by *T. b. gambiense*) in 2013. B) Distribution of rhodesiense HAT (caused by *T. b. rhodesiense*) in 2013. Adapted from the worldwide distributions of human African trypanosomiasis maps available from WHO on **http://gamapserver.who.int/mapLibrary/Files/Maps/HAT_ga_2013.png** (*T. b. gambiense*) and **http://gamapserver.who.int/mapLibrary/Files/Maps/HAT_rh_2013.png** (*T. b. rhodesiense*), accessed in March 2015.

HAT consists of two clinical stages. The first, early, or hemo-lymphatic stage, takes place when the parasite resides and multiplies in the bloodstream and lymphatic system, causing general symptoms such as intermittent fever, headache, weakness, anaemia and musculoskeletal pains, commonly misdiagnosed as other fever-causing diseases (reviewed in Franco *et al.*, 2014; Brun *et al.*, 2010). If not treated, HAT develops to the second, late, or meningo-encephalitic stage, in which the parasites cross the blood-brain barrier, invading the central nervous system and progressively causing serious neurological damage. It is during this stage that common sleeping sickness-associated

symptoms such as sleep disturbances and neuropsychiatric symptoms are noticeable and, if not treated, eventually result in coma and subsequent death (reviewed in Franco *et al.*, 2014; Brun *et al.*, 2010). During both stages, the parasite is exclusively extracellular, with the levels of the parasite in the blood (parasitemia) fluctuating throughout the infection, partly as a result of the infected-host immune response and the parasite's evasion strategy, termed antigenic variation (see below). In rhodesiense HAT, the first stage of the disease is much shorter than in gambiense HAT, resulting in death within 6-8 months of infection. On the other hand, individuals suffering from gambiense HAT remain mildly symptomatic for years, although the chronic meningo-encephalic stage is seen, if untreated, within an average of 3 years (reviewed in Franco *et al.*, 2014; Fevre *et al.*, 2006; Brun *et al.*, 2010). Diagnosis is not easy, and new improved methods are necessary, as the present ones based on microscopic analysis of blood and cerebrospinal fluid, or serologic and polymerase chain reaction (PCR)-based assays, appear to be insufficient (Brun *et al.*, 2010).

Treatment of HAT is performed according to the stage of the disease using a very small range of drugs: first stage gambiense HAT is treated with pentamidine; second stage gambiense HAT can be treated with either eflornithine or melarsoprol; first stage rhodesiense HAT is treated with suramin; and second stage rhodesiense HAT is treated with melarsoprol (Brun *et al.*, 2010; Steverding, 2010). Melarsoprol and eflornithine have been also used in combination with nifurtimox (oral drug used to treat Chagas disease), for the treatment of second stage gambiense HAT, with better results than monotherapy (reviewed in Brun *et al.*, 2010). In all cases, the treatment regimens require either intramuscular (pentamidine) or intravenous (eflornithine, melarsoprol, and suramin) injections, given in various doses over, sometimes, long periods of time, making the treatments cumbersome and not always feasible in the most affected areas (Brun *et al.*, 2010; Fevre *et al.*, 2006). In addition, these drugs are highly toxic, and result in mild to severe, and sometimes life threatening (in the case of melarsoprol), side effects (Brun *et al.*, 2010; Steverding, 2010). For these reasons, and considering that most of these drugs were developed in the first half of the 20th century, it is imperative that new, safer drugs are developed in the near future.

1.1.3 The life cycle of *Trypanosoma brucei*

T. brucei parasites are transmitted to their mammalian hosts through the bite of an infected tsetse fly (*Glossina* spp.) (reviewed in Fevre *et al.*, 2006). While probing its blood meal, the fly inoculates the host with the non-replicating, cell cycle arrested, mammal infective life cycle stage of the parasite, the metacyclic trypomastigote form, which is found in the fly's salivary glands (see below), and injected into the mammalian host with the fly's saliva (reviewed in Franco *et al.*, 2014). Once in the mammalian host bloodstream, the metacyclic form cells re-enter the cell cycle. Associated with the onset of replication, the cells undergo morphological changes, differentiating into the proliferative long slender bloodstream forms (BSF) that actively divide (Figure 1.3), establishing and maintaining an extracellular infection in the bloodstream and lymphatic system. BSF stage cells express variant surface glycoproteins (VSGs) on their surface membrane, which form a dense coat that is recognised by the host immune system. The parasite is able to evade the host's anti-VSG immune response by sequentially expressing distinct VSGs on its surface (via antigenic variation). As each new VSG is not immediately recognised by the host, such switching allows continued immune evasion (reviewed in McCulloch *et al.*, 2014; Glover *et al.*, 2013b), which contributes to the intermittent waves of parasitemia characteristic of trypanosomiasis (reviewed in MacGregor *et al.*, 2012). As long slender BSF numbers increase in the bloodstream, quorum sensing-like mechanisms are activated, and the parasite differentiates into a non-proliferative, cell cycle arrested stage pre-adapted for transmission to the fly, the short stumpy form (Figure 1.3) (Vassella *et al.*, 1997; Reuner *et al.*, 1997; MacGregor *et al.*, 2011; Dean *et al.*, 2009; Mony *et al.*, 2014). This differentiation allows the parasite to limit its numbers in the bloodstream, prolonging host survival and increasing the chances of transmission, as well as providing pre-adaptation to the fly's gut environment (reviewed in MacGregor *et al.*, 2012).

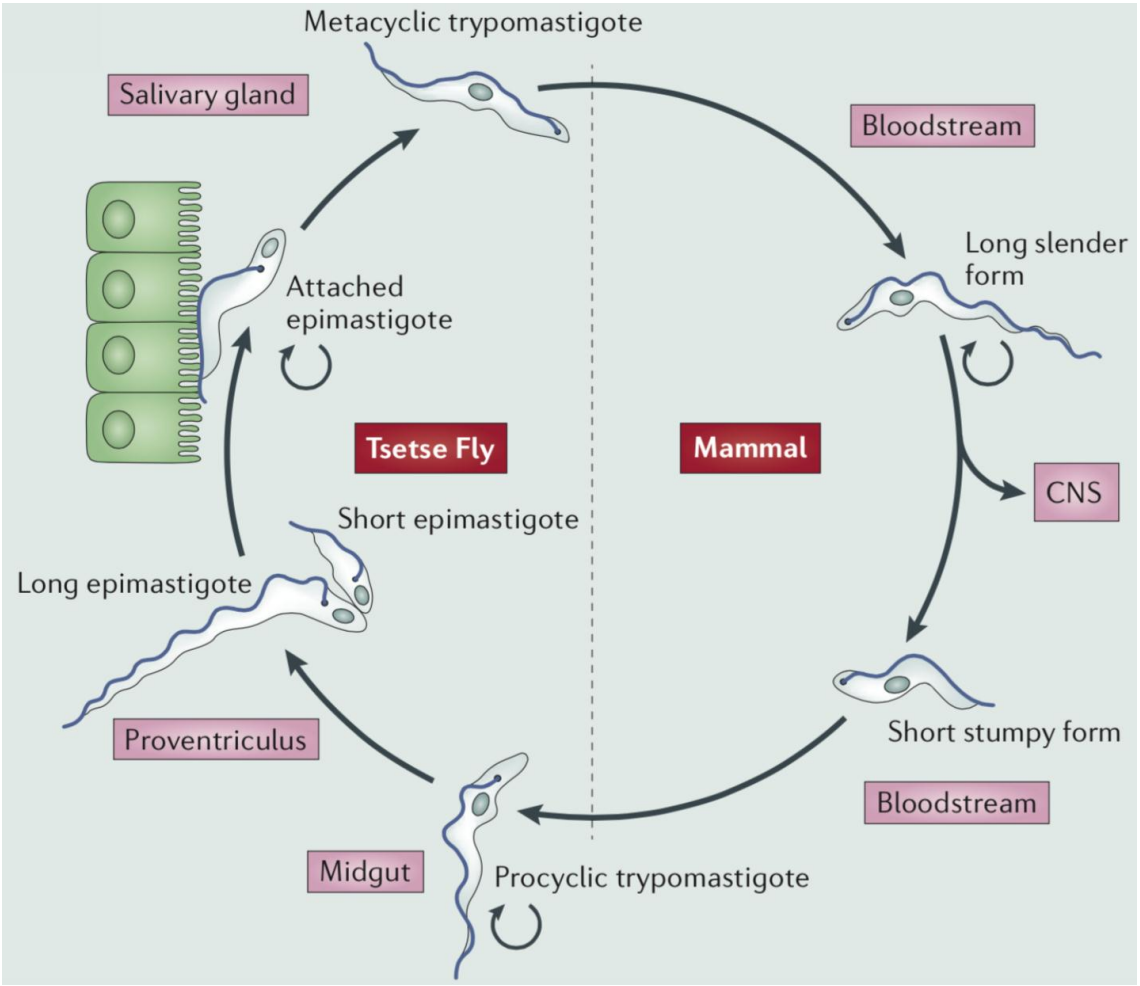


Figure 1.3. *T. brucei* life cycle.
Simplified diagram of the life cycle of *T. brucei* parasites. Detailed description in the main text. CNS refers to the central nervous system. Reproduced from (Langousis and Hill, 2014), with permission (license number 3586570870008).

When a tsetse fly probes its blood meal from an infected mammal, short stumpy forms are ingested and, in the fly's midgut, differentiate into replicative trypomastigote procyclic forms (PCF) (Figure 1.3) (Jones *et al.*, 2014; Szoor *et al.*, 2006; Szoor *et al.*, 2010; Matthews and Gull, 1997), thus establishing the next stage of the life cycle (reviewed in Dyer *et al.*, 2013; Ooi and Bastin, 2013). From here, the parasites migrate to the proventriculus, where they undergo extensive morphological changes into long trypomastigotes that then divide asymmetrically to generate one long and one short epimastigote (Figure 1.3) (reviewed in Dyer *et al.*, 2013; Ooi and Bastin, 2013). This latter form of the parasite then migrates to the salivary gland, where it attaches to the epithelial cells (Figure 1.3) and proliferates, colonising the salivary glands (Rotureau *et al.*, 2012). Eventually, this epimastigote form divides asymmetrically (Rotureau *et al.*, 2012) to generate non-proliferative metacyclic trypomastigote forms (Figure 1.3) (reviewed in Dyer *et al.*, 2013; Ooi and Bastin, 2013). Virtually all

fly stages of the parasite do not express VSG on their surface: procyclin is expressed in PCF cells (Acosta-Serrano *et al.*, 2001; Vassella *et al.*, 2001; Urwyler *et al.*, 2005) and BARP in epimastigotes (Urwyler *et al.*, 2007). However, metacyclic form cells express a single metacyclic VSG (MVSG) on their surface, selected from the MVSG repertoire (Ginger *et al.*, 2002; Vickerman, 1985; Barry *et al.*, 1998). A number of different MVSG-coated metacyclic cells are generated, providing a heterogeneous metacyclic population (Tetley *et al.*, 1987; Barry *et al.*, 1998). Together with other biological features (e.g. rate of endocytosis) (Natesan *et al.*, 2007), VSG expression renders the metacyclic forms adapted to infect the mammalian host (Vickerman, 1985; Barry *et al.*, 1998). Metacyclic forms are released into the salivary gland lumen (Tetley and Vickerman, 1985) and subsequently inoculated into the mammalian host, starting the life cycle again (Figure 1.3).

1.2 *Trypanosoma brucei* in more detail

1.2.1 Cell structure

Being a parasite with a complex life cycle in which it must survive and develop in two different hosts, *T. brucei* undergoes extensive biological modifications during the life cycle, including discernible morphological changes (Figure 1.3 and Figure 1.4). One morphological difference between the epimastigote and the trypomastigote cell types is the position of kinetoplast localisation relative to the nucleus: in epimastigotes, the kinetoplast is localised anteriorly to the nucleus, while in trypomastigotes it is positioned posteriorly (Figure 1.4, A) (Wheeler *et al.*, 2013a). Moreover, there are also considerable morphological differences between the various trypomastigote stages of the parasite (BSF, stumpy forms, PCF and metacyclic forms) (Figure 1.3). However, despite assuming various morphologies, the general cellular ultrastructure of the parasite is largely conserved (Wheeler *et al.*, 2013a). *T. brucei* cells have an elongated shape, with a single flagellum that is laterally attached to most of the length of the cell body (juxtaform morphology; only the distal tip is free of the cell body) (Wheeler *et al.*, 2013a). The length and lateral attachment of the flagellum controls the body length of the cell (Wheeler *et al.*, 2013a), besides

also playing an important role in cell motility and cell division (reviewed in Langousis and Hill, 2014; Field and Carrington, 2009). The flagellum has a canonical 9 + 2 axoneme structure with an associated paraflagellar rod, and protrudes from the posterior end of the cell, emerging from the flagellar pocket, the only site where endocytosis and exocytosis take place (reviewed in Overath and Engstler, 2004). Here, the flagellum originates from the basal body, which is linked to the kinetoplast through the mitochondrial membrane (reviewed in Langousis and Hill, 2014; Field and Carrington, 2009). The cell's shape is defined by a highly polarised sub-pellicular microtubule cytoskeleton, an array of uniformly distributed microtubules under the plasma membrane, which remains intact during all stages of the cell cycle (reviewed in Matthews, 2005). Cellular structures and organelles, such as the flagellar pocket, the flagellum, the Golgi complex, the mitochondrion, the kinetoplast, and the nucleus, are present in single copy, and are specifically positioned (with the exception of the mitochondrion that runs the length of the cell) between the posterior end and the centre of the cell (Figure 1.4, B), forming a highly reproducible and organised cell structure (reviewed in Matthews, 2005).

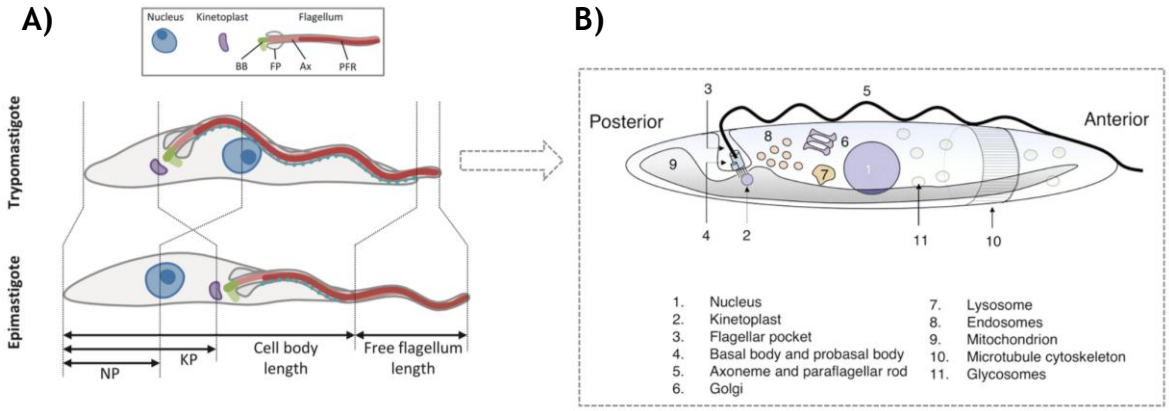


Figure 1.4. *T. brucei* general cell structure.
Simplified representations of *T. brucei* cells. A) Representation of a trypomastigote cell (top) and an epimastigote cell (bottom). Legend within the insert box: basal body (BB); flagellar pocket (FP); axoneme (Ax); paraflagellar rod (PFR); nucleus-posterior distance (NP); kinetoplast-posterior distance (KP). B) Trypomastigote general internal cellular structure in more detail. Legend is shown below the diagram. A) Reproduced from (Wheeler *et al.*, 2013a), PLoS ONE, with permission, © 2013 Wheeler et al. all rights reserved. B) Reproduced from (Matthews, 2005), with permission, © Company of Biologists Ltd., all rights reserved.

1.2.2 Cell cycle

Overall, *T. brucei* BSF and PCF proliferative trypomastigote forms follow the general lines of a typical eukaryotic cell cycle, although containing some unique features (reviewed in Matthews, 2005; Vaughan and Gull, 2008) and perhaps diverged regulatory mechanisms (reviewed in Hammarton, 2007; Li, 2012). During G1 phase, the first morphologic event of cell division is the duplication of the basal body and the beginning of the synthesis of a new flagellum (Woodward and Gull, 1990; Wheeler *et al.*, 2013b) (Figure 1.5, A, second row), which is immediately followed by the duplication of the Golgi complex (Ho *et al.*, 2006). In *T. brucei* the single-copy organelles are apparently duplicated by a template-based mechanism where a new organelle is generated just beside the old one (reviewed in Vaughan and Gull, 2008). One aspect in which *T. brucei*'s cell cycle differs from the canonical eukaryotic cell cycle is the S phase, as the parasite has two distinct, although highly coordinated, S phases: the kinetoplast S phase, where the kinetoplast DNA (kDNA) is replicated; and the nuclear S phase, when the nuclear chromosomal DNA is replicated (Figure 1.5, B). kDNA replication is initiated by factors (reviewed in Liu *et al.*, 2005; Jensen and Englund, 2012) distinct from the ones involved in nuclear DNA replication (reviewed in Tiengwe *et al.*, 2013). Kinetoplast S phase is started immediately before nuclear S phase, and it is considerably shorter (Woodward and Gull, 1990), meaning that the kinetoplast and nuclear cell cycle phases are not synchronised. During nuclear S phase, the kinetoplast G2 phase starts, culminating in the segregation of the old and newly formed kinetoplast networks (when the basal bodies are separated by a microtubule-mediated event) during nuclear G2 phase (reviewed in Liu *et al.*, 2005; Jensen and Englund, 2012) (Figure 1.5, A, third row). Alongside, the Golgi complex is also segregated (Hall *et al.*, 2006). Until the onset of nuclear mitosis, these cells possess one nucleus (N) and two kinetoplasts (K) (1N2K) (Woodward and Gull, 1990). Next, mitosis occurs as a closed process (Figure 1.5, A, third row): the nuclear envelope is not disrupted, and an intra-nuclear spindle mediates chromosomes segregation, resulting in a cell with two nuclei and two kinetoplasts (2N2K) (Ogbadoyi *et al.*, 2000). During all the previous phases, the new flagellum is continuously formed.

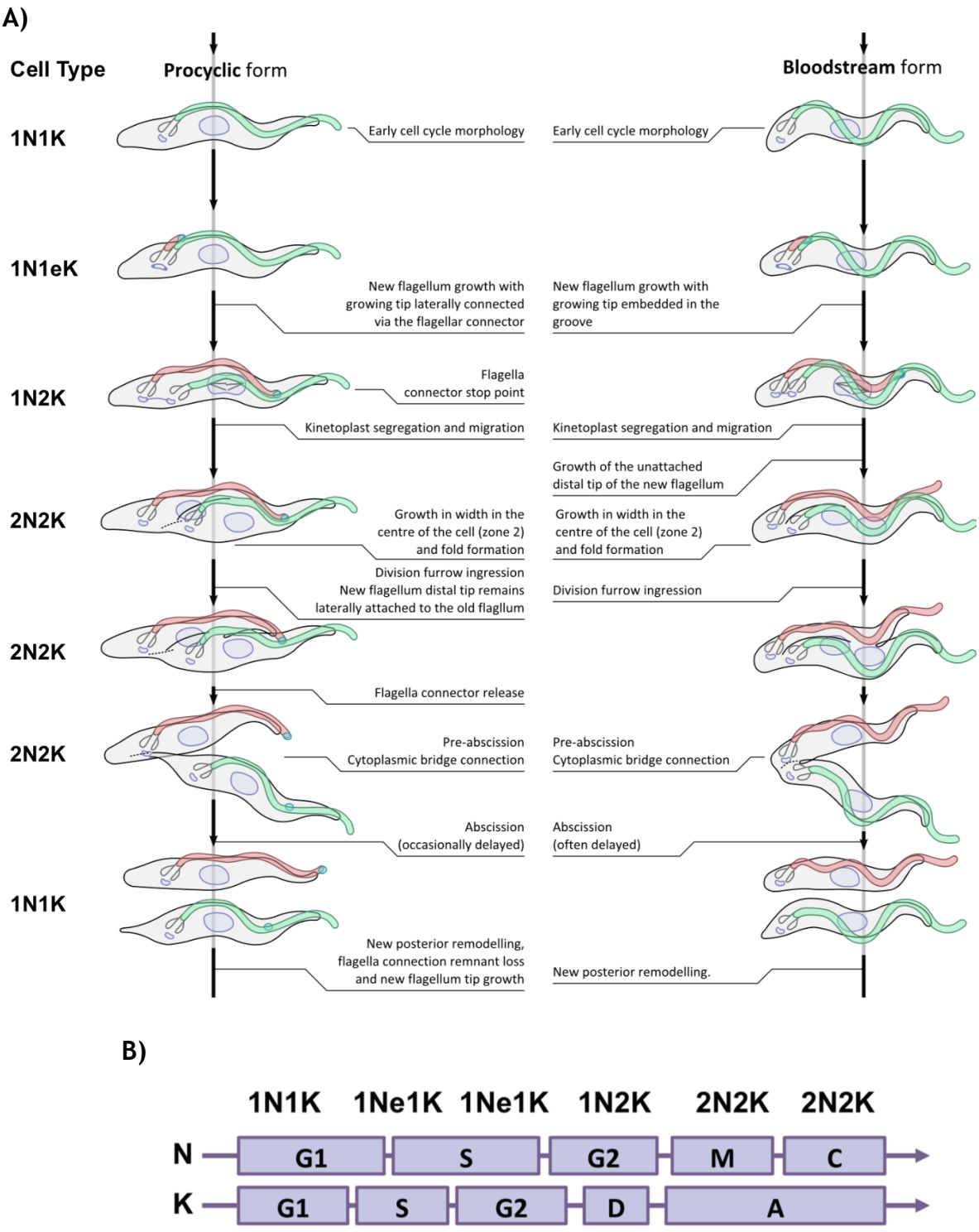


Figure 1.5. Cell cycle of *T. brucei* PCF and BSF cells.
A) Schematic representation of the different cell morphologies of the PCF cells (left) and BSF cells (right) during the cell cycle. On the far left, the cell type refers to the nucleus (N) and kinetoplast (K) configurations characteristic of each cell cycle stage shown for both PCF and BSF cells. In both representations of PCF and BSF cells, the new flagellum is represented in red, and the old flagellum is shown in green. Each major occurrence is described for each cell cycle stage. B) Simplified representation of the nuclear (N) and kinetoplast (K) S phases timing in relation to each other. M refers to nuclear mitosis, C to cytokinesis, D to kinetoplast division, and A to kinetoplast anaphase (segregation). The cell types according to their nucleus and kinetoplast conformations are shown above. A) Figure adapted from (Wheeler *et al.*, 2013b), © 2013 Wheeler et al. all rights reserved. B) Diagram generated based on the descriptions and schematics from (Woodward and Gull, 1990; McKean, 2003; Ploubidou *et al.*, 1999; Hammarton, 2007).

The non-synchronisation but tight coordination of the kinetoplast and nuclear S phases allow the classification of individual cells in an asynchronous cellular population into the cell cycle stages they are in: cells with 1N1K morphology are in nuclear G1; cells with one nucleus but one "elongated" (e) kinetoplast (1N1eK) are in nuclear late G1 or S phase; cells with 1N2K are in nuclear G2; while cells with 2N2K are post-mitotic but have not undergone complete cytokinesis (Woodward and Gull, 1990; Siegel *et al.*, 2008) (Figure 1.5, B).

Upon cytokinesis, which starts immediately after mitosis, a unidirectional cleavage furrow is formed (Figure 1.5, A, rows 4 and 5), anterior-posteriorly along the longitudinal axis of the parasite, passing between the two flagella of the dividing cell (Wheeler *et al.*, 2013b). This allows the division of the two cells, which remain connected by a discrete "cytoplasmic bridge" (thin section of the membrane) (Figure 1.5, A, sixth row) prior to abscission (Wheeler *et al.*, 2013b). After abscission the two daughter cells are released, each with a single set of organelles (Figure 1.5, A, seventh row).

Overall, the cell cycles of PCF and BSF cells are similar (Figure 1.5, A), though several morphological differences are seen (Wheeler *et al.*, 2013b), and functional studies have shown that the molecular mechanisms involved in cell cycle regulation might differ between these two life cycle stages (reviewed in Hammarton, 2007; Li, 2012). In nuclear G1 phase (1N1K cells), the morphology of BSF and PCF cells is similar, although the latter is longer and wider, with the kinetoplast localised further from the posterior end, and with a shorter protruding portion of the flagellum (Wheeler *et al.*, 2013b) (Figure 1.5, A, top row). The major morphological differences arise after mitosis, most likely due to the way the new flagellum is displayed in the cell during its synthesis and cytokinesis in the two cell types (Wheeler *et al.*, 2013b). A key difference lies in the presence of a flagella connector in PCF cells that attaches the distal tip of the growing new flagellum to the old flagellum until it reaches a "stop point", from where the new flagellum tip does not move forward relative to the old one, though its growth continues and it is maintained completely attached to the cell body (reviewed in Langousis and Hill, 2014) (Figure 1.5, A, rows 2 to 4). The flagella connector is absent from BSF cells and the new flagellum grows with the distal tip located in the cell-body groove (Hughes *et al.*, 2013) until it reaches

the end of the cell body, continuing to grow unattached to the cell body during cytokinesis (Wheeler *et al.*, 2013b) (Figure 1.5, A, rows 2 to 4). The new flagellum growth coincides with the movement of the base of the flagellum, the flagellar pocket and consequently the kinetoplast, resulting in the differing placement of the nuclei and kinetoplasts in post-mitotic 2N2K cells, with a conformation of 1K1N1K1N in PCF cells but 1K1K1N1N in BSF cells (Wheeler *et al.*, 2013b) (Figure 1.5, A, rows 4 and 5). Differences between the cell cycle controls of these two stages of the parasite are, however, more profound. Various functional studies investigating the cell cycle effects resulting from the depletion of certain factors by RNA interference (RNAi) or from blocking certain cell cycle stages with drugs have revealed different phenotype outcomes in PCF and BSF cells. For example, chemical inhibition of DNA replication and mitosis revealed that PCF cells still progress through cytokinesis, suggesting that this life cycle stage does not possess the mitosis to cytokinesis cell cycle checkpoint (Ploubidou *et al.*, 1999). This was supported by RNAi studies investigating the role of TbCYC6, a mitotic cyclin (Hammarton *et al.*, 2003), the mitotic cyclin-dependent kinase TbCRK3 (Tu and Wang, 2004), and the aurora-like kinase TbAUK1 (Li and Wang, 2006). However, the depletion of any of these factors in BSF cells resulted in the blocking of cytokinesis, suggesting that this life cycle stage, in contrast to PCF cells, possesses the mitosis to cytokinesis checkpoint (Hammarton, 2007; Li, 2012). A similar disparity in phenotypes was also obtained with the depletion of the DNA replication initiator factor TbORC1/CDC6 (see below) (Benmerzouga *et al.*, 2013; Tiengwe *et al.*, 2012b).

1.2.3 Unusual genome organisation and unconventional transcription

The nuclear genome of *T. brucei* is composed of eleven diploid megabase-sized chromosomes (0.9 to 6 Mbp), one to five aneuploid intermediate chromosomes (150 to 900 Kbp), and around 100 minichromosomes (50 to 150 Kbp) (Berriman *et al.*, 2005; Melville *et al.*, 1998; Turner *et al.*, 1997) (Figure 1.6). All the three types of chromosome possess telomeric repeats (TTAGGG) with the same base composition as mammals (reviewed in Dreesen *et al.*, 2007). The megabase chromosomes (Figure 1.6) display a portioned organisation, with a core region containing all the RNA polymerase II (RNA Pol II)-transcribed housekeeping genes (~7,500) flanked by large hemizygous subtelomeres (between the telomere and

the first housekeeping gene) containing arrays of VSG genes and VSG pseudogenes (~1,500) (Marcello and Barry, 2007; Cross *et al.*, 2014). Mammalian VSGs are exclusively expressed from telomeric bloodstream expression sites (BES), found on the megabase and intermediate chromosomes and where the VSG and expression site associated genes (ESAGs) are co-expressed by RNA Pol I (reviewed in McCulloch *et al.*, 2014; Glover *et al.*, 2013b; Daniels *et al.*, 2010; Ersfeld, 2011). Each BSF cell expresses only one BES, a process termed monoallelic expression (reviewed in Gunzl *et al.*, 2014). *T. brucei* is the only known organism that uses RNA Pol I to transcribe not only rRNA genes, but also protein-coding genes (Gunzl *et al.*, 2003): this is not limited to the *VSGs*, but also and the procyclin genes, encoding the surface proteins on PCF cells (Haenni *et al.*, 2006; Haenni *et al.*, 2009). To date, the centromeres of only megabase chromosomes 1 to 8 have been mapped (Obado *et al.*, 2007), and were recently shown to localise with some protein components of the highly diverged, putative *T. brucei* kinetochore (Akiyoshi and Gull, 2014). The role of both the intermediate and minichromosomes appears to be entirely dedicated to antigenic variation. These chromosomes encode intact and functional VSGs or VSG pseudogenes (that can recombine to generate functional new “mosaic” VSGs), contributing to the parasite’s VSG archive, and providing the parasite with a repertoire of transcriptionally silent VSGs that can be activated (reviewed in McCulloch *et al.*, 2014; Glover *et al.*, 2013a). Structurally, the minichromosomes consist of a central uninterrupted repetitive palindromic core region (20-80 Kbp per chromosome, with a single inversion point in the centre) of 177 bp repeats, flanked by short subtelomeric regions containing non-repetitive DNA, and the telomeres (Wickstead *et al.*, 2004). The stretches of 177 bp repeats are also present in the intermediate chromosomes, though at the subtelomeres, flanking a non-repetitive core region (Wickstead *et al.*, 2004). These intermediate chromosomes also contain BESs (Hertz-Fowler *et al.*, 2008) that can be activated upon antigenic variation (reviewed in McCulloch *et al.*, 2014; Glover *et al.*, 2013b). Overall, around 20% of the *T. brucei* genome is dedicated to the VSG archive and antigenic variation (Berriman *et al.*, 2005).

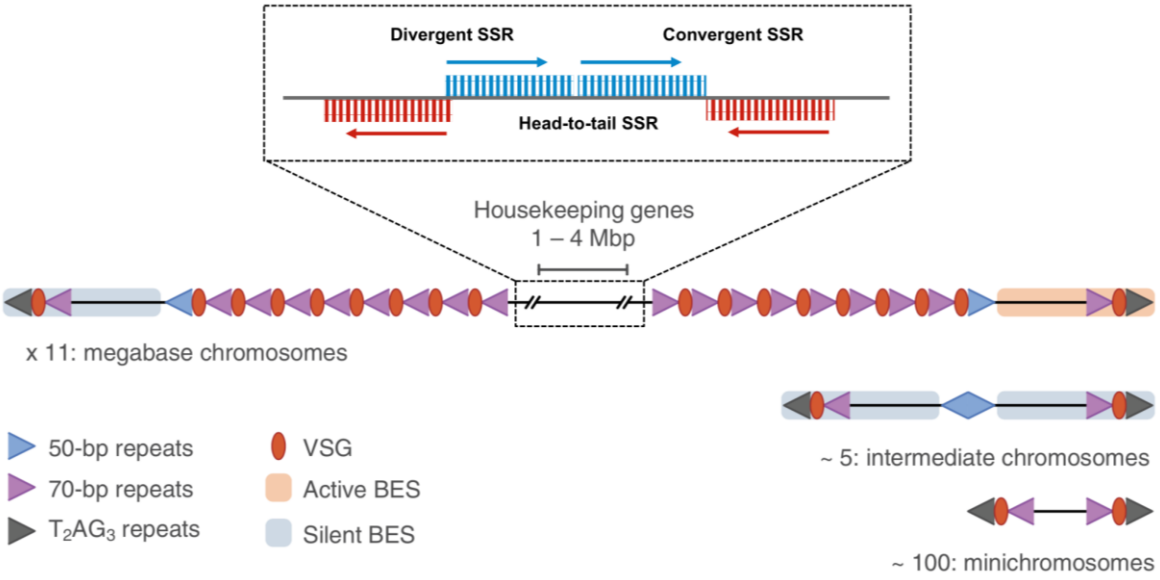


Figure 1.6. *T. brucei* megabase, intermediate and minichromosome structures. Simplified representation of chromosome structure and organisation of the three types of chromosomes composing the genome of *T. brucei*. The megabase chromosomes contain a core region harbouring all the parasite's housekeeping genes, arranged in large directional gene clusters (DGCs) or large polycistronic transcription units (shown in red and blue, the transcription orientation highlighted by an arrow). Two DGCs are separated by a region called the strand switch region (SSR), which can be divergent, convergent or head-to-tail, depending on the orientation of the DGCs, as shown and described in the main text. The subtelomeric region, containing part of the VSG archive, flanks this core section of the chromosomes. At the extremities are the telomeres, where the bloodstream expression sites (BESs) are localised. Only one BES is active per cell at a given time (represented in orange). The telomeres of the megabase, intermediate and minichromosomes are made up by TTAGGG (T_2AG_3)-repeats. The intermediate and minichromosomes harbour the remaining of the VSG archive, with some of the BES localising to the intermediate chromosomes. The core region of the minichromosomes consists of palindromic 177 bp repeats. Not to scale. Diagram adapted from (Glover *et al.*, 2013b), and descriptions from (Ersfeld, 2011; Daniels *et al.*, 2010).

Sequencing of the *T. brucei*, *T. cruzi* and *L. major* (also known as TriTryps) genomes 10 years ago (El-Sayed *et al.*, 2005b; El-Sayed *et al.*, 2005a; Berriman *et al.*, 2005; Ivens *et al.*, 2005) showed that these parasites share a conserved core proteome (~6200 genes; *T. brucei* and *T. cruzi* genomes share an average of 57% identify, while *T. brucei* and *L. major* share about 44%), encoded in highly syntenic (i.e. the order of the genes is conserved) gene blocks within the chromosome cores, with, for example, ~68% of *T. brucei* and ~75% of *L. major* genes being found in the same genomic context (El-Sayed *et al.*, 2005b). The organisation of the core genome into 11 large megabase chromosomes contrasts with the more fragmented genomes of *T. cruzi* (41 chromosomes) (Weatherly *et al.*, 2009), and *L. major* (36 chromosomes) (Ivens *et al.*, 2005). Due to the high levels of synteny, it has been suggested that the *T. brucei* genome derived from the fusion of a more fragmented ancestral genome (El-Sayed *et al.*, 2005b). The main difference between these genomes lays in the non-syntenic subtelomeric

regions, which are short in *L. major*, but large in both *T. cruzi* and *T. brucei* (El-Sayed *et al.*, 2005b). These contain species-specific features associated with disease mechanisms, such as antigenic variation in *T. brucei* (reviewed in McCulloch *et al.*, 2014; Glover *et al.*, 2013b). A characteristic of the trypanosomatids genomes is their unconventional gene organisation into large polycistronic gene clusters, also known as polycistronic transcription units or directional gene clusters (DGCs), in which the genes are arranged in a head-to-tail orientation (reviewed in Daniels *et al.*, 2010; Jackson, 2014; Stuart and Myler, 2006; Ersfeld, 2011). The various DGCs are separated by regions named strand switch regions (SSRs) (Figure 1.6), which can be divergent (dSSR; the genes in the two DGCs are oriented away from the SSR in opposing directions), convergent (cSSRs; the transcription the genes in the two DGCs are oriented towards the SSR in opposing directions), or head-to-tail (h-t SSR; two DGCs are oriented in the same direction) (reviewed in Daniels *et al.*, 2010). Each DGC is transcribed from a single RNA Pol II transcription initiation site, though no typical RNA Pol II promoters are found, generating polycistronic mRNAs (reviewed in Daniels *et al.*, 2010; Ersfeld, 2011). These polycistronic mRNAs are then processed into monocistronic mRNAs by co-transcriptional *trans*-splicing of a capped short spliced-leader (SL) RNA exon to the 5' untranslated region (5' UTR), coupled with polyadenylation of the upstream gene, producing mature mRNAs (reviewed in Daniels *et al.*, 2010; Clayton, 2002; Michaeli, 2011). Transcription is initiated at dSSRs and h-t SSRs sites enriched in modified or variant histones, such as histone H4 acetylated at lysine 10 (H4K10ac), histone H3 trimethylated at lysine 4 (H3K4me3) (Wright *et al.*, 2010) and histone H2 variants H2AZ and H2BZ (Siegel *et al.*, 2009) in *T. brucei*, and acetylated histone H3 in *L. major* (Thomas *et al.*, 2009). Transcription termination is still not completely understood (reviewed in Ersfeld, 2011), but transcription appears to terminate at cSSRs and h-t SSRs that are enriched in histone variants H3V and H4V in *T. brucei* (Siegel *et al.*, 2009), and base J (an hypermodified thymidine base in kinetoplastids) (reviewed in Borst and Sabatini, 2008) in *L. major* (Reynolds *et al.*, 2014). As a consequence of transcription into large polycistronic mRNAs, gene expression in trypanosomatids is mainly controlled at the post-transcriptional level (reviewed in Clayton, 2002; Siegel *et al.*, 2011; Kramer, 2012).

1.3 Initiation of nuclear DNA replication as we know it

Throughout evolution, cellular life has diverged into what is now classified as the three domains of life: *Bacteria*, *Archaea* and *Eukarya* (reviewed in Pace *et al.*, 2012). Despite their extreme differences, basic molecular mechanisms underlying vital processes, such as gene transcription (reviewed in Werner and Grohmann, 2011) and DNA replication (reviewed in O'Donnell *et al.*, 2013), have been kept considerably functionally conserved. However, the molecular machineries and associated regulatory processes involved in these core processes present different degrees of complexity between the three domains, being apparently simpler in bacterial and archaeal organisms than in eukaryotic ones (reviewed in Masai *et al.*, 2010; O'Donnell *et al.*, 2013).

The complete faithful duplication of the genome before cellular division is central to the propagation of life, and one of the most complex and tightly regulated processes in the cell. Normally, the genomic DNA is replicated once, and only once, per cell cycle (reviewed in O'Donnell *et al.*, 2013; Masai *et al.*, 2010). A key step in this is, therefore, the initiation of the process, which involves the tightly controlled and timely recruitment and activation of various factors. Simplistically, initiation of DNA replication consists in the recruitment and binding of the initiator factor(s) to the replication starting sites (termed origins of replication), which leads to the recruitment and loading of the replicative helicase, that is necessary for opening of the DNA helix during replication (Figure 1.7) (reviewed in Masai *et al.*, 2010; O'Donnell *et al.*, 2013). After the helicase is loaded, the components of the replication machinery, including the replicative DNA polymerases, are then recruited to the origin and consequently, start of DNA synthesis.

In the next sections, the molecular machineries involved in the initiation of DNA replication, the origins of replication and how they are defined, and the regulation of the initiation steps of DNA replication in the three domains of life, are explored.

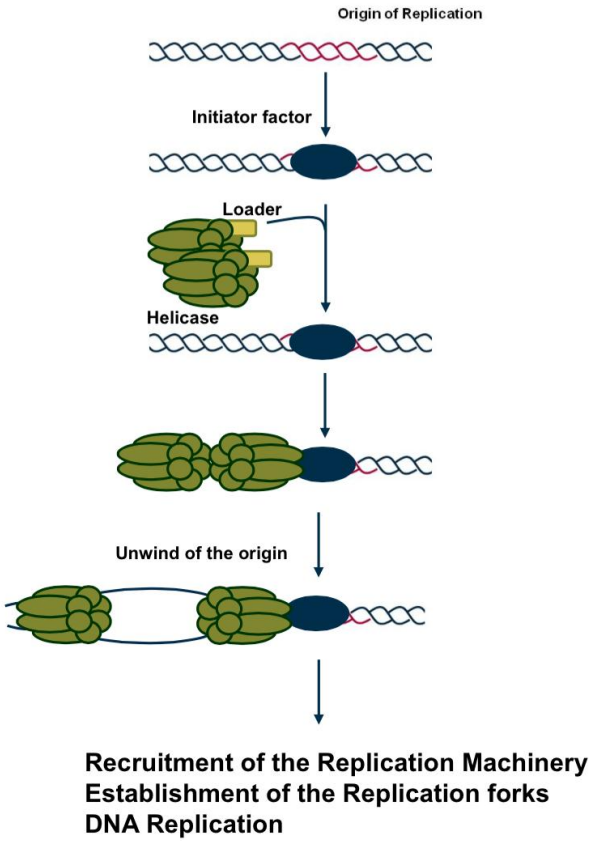


Figure 1.7. Initiation of DNA replication.
Simplified diagram representing the steps of initiation of DNA replication. An initiator factor recognises and binds the origin of replication. This allows the further recruitment of the replicative helicase to the origin, either through a helicase loader or directly. The helicase is then activated, and melts the dsDNA, allowing the recruitment of the remaining components of the replication machinery (not shown), and thus the beginning of DNA synthesis. Note than in bacteria the origin is not melted by the helicase but by the initiator factor. Details of the machineries and process in bacteria, archaea and eukaryotes are detailed in the main text. Diagram based on the descriptions made in various review articles (Costa *et al.*, 2013; Masai *et al.*, 2010).

1.3.1 Initiation molecular machineries

Recognition of the origins of replication by the initiator factors is the first step in DNA replication (reviewed in O'Donnell *et al.*, 2013; Masai *et al.*, 2010; Costa *et al.*, 2013). The replication initiator factors found in bacteria, archaea and eukaryotes belong to the AAA+ superfamily of ATPases, more specifically, to the DnaA/CDC6/ORC clade, which encompasses two sequence-divergent families: the bacterial DnaA family and the archaeo-eukaryotic CDC6/ORC family (reviewed in Iyer *et al.*, 2004; Davey *et al.*, 2002b; Erzberger and Berger, 2006; Kawakami and Katayama, 2010). Nevertheless, these AAA+ ATPase factors are molecularly conserved and perform analogous roles in replication, working both in the identification and binding of replication origins (through their DNA recognition domains) and as molecular switches, in which ADP/ATP are used as

driving forces of initiator factor assembly and helicase loading onto the origins (reviewed in Robinson and Bell, 2005; Wigley, 2009; Duncker *et al.*, 2009; Kawakami and Katayama, 2010).

1.3.1.1 General structure of AAA+ ATPases

The amino acid ATP-binding (AAA+) domain's basic structure is shown in Figure 1.8. Motifs within domain I (Figure 1.8) are also found in other P-loop NTPases, while motifs in domain II define the AAA+ ATPase family (reviewed in Davey *et al.*, 2002b). In domain I, the signature Walker A and Walker B motifs, fundamental parts of the AAA+ ATP-binding site, are found (Figure 1.8). The Walker A (the phosphate-binding loop, or P-loop) is usually conserved as $GX_4GK[S/T]$[a], and forms a distinctive loop that interacts directly with the phosphates of the ATP molecule (for which the lysine (K) residue is crucial). The Walker B motif is typically defined by hhhhDE[b], in which the hydrophobic amino acids are essential for ATPase activity, the aspartate (D) co-ordinates the magnesium ions (Mg^{2+}) needed for ATP hydrolysis, and the glutamate (E) residue's role is to prime a water molecule for the ATP hydrolysis reaction (reviewed in Iyer *et al.*, 2004; Davey *et al.*, 2002b; Erzberger and Berger, 2006; Matte and Delbaere, 2010; Hanson and Whiteheart, 2005). Though it is now recognised, as more protein structures are analysed, that the canonical sequences of the Walker motifs are not as conserved as previously thought (reviewed in Matte and Delbaere, 2010), some specific modifications to the Walker A and B motifs are known to render the proteins unable to bind ATP (e.g. the replacement of the lysine – K, in the Walker A by an alanine – A) and devoid of ATPase activity (change from the glutamate – E , to glutamine – Q, in the Walker B motif) (reviewed in Iyer *et al.*, 2004; Hanson and Whiteheart, 2005). A distinguishable characteristic of the initiator factors clade is the insertion of an additional α-helix between the helix α2 and strand ß2 (thus between the Walker A and Walker B motifs). This insertion is called the initiator-specific motif (ISM), and is known to be important for oligomerisation of the initiator factor as well as origin recognition (Iyer *et al.*, 2004; Erzberger and Berger, 2006; Duderstadt and Berger, 2008). Still within domain I, is the sensor 1 motif (Figure 1.8). This motif

[a] X refers to any amino acid.

[b] h refers to an hydrophobic amino acid.

has a conserved polar residue (usually asparagine – N, though it can also be serine – S, threonine – T, or aspartate – D), which is important for ATP hydrolysis (reviewed in Iyer *et al.*, 2004; Erzberger and Berger, 2006). Physically, this polar residue is placed between the Walker A and Walker B motifs, where it interacts with the second acidic residue of the Walker B motif to properly orient the water molecule for ATP hydrolysis (reviewed in Hanson and Whiteheart, 2005; Erzberger and Berger, 2006). Next to the sensor 1 motif, localising to Box VII or SRC, is the conserved arginine finger (Figure 1.8), which is involved in ATP hydrolysis not of the molecule it is part of, but of the adjacent AAA+ ATPase subunit in the complex they are both part of (see example of Orc1 and Orc4, below). This is achieved because, structurally, the arginine of one subunit is docked into the ATP-containing active site of the adjacent subunit in the complex, where it interacts directly with the ATP there bound (reviewed in Iyer *et al.*, 2004; Hanson and Whiteheart, 2005). Finally, in the C-terminal region of the AAA+ domain (domain II) is the sensor 2 motif (Figure 1.8) (reviewed in Davey *et al.*, 2002b; Hanson and Whiteheart, 2005), containing a frequently conserved arginine residue that interacts directly with ATP (reviewed in Iyer *et al.*, 2004; Erzberger and Berger, 2006). This element is involved in both ATP-binding and hydrolysis, and undergoes conformational changes whether the protein is bound to ATP or ADP (reviewed in Hanson and Whiteheart, 2005; Iyer *et al.*, 2004).

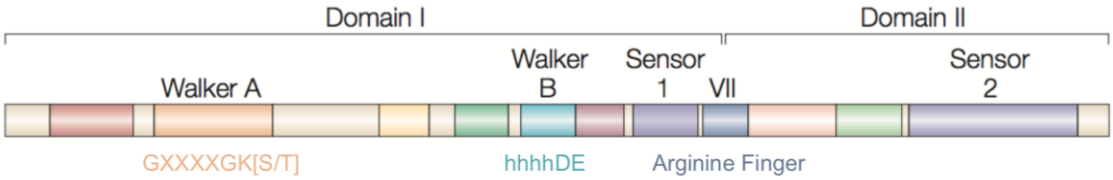


Figure 1.8. AAA+ ATPase domain: characteristic motifs and elements.
Simplified representation of the AAA+ domain. The signature motifs and elements are shown: domain I, enclosing the Walker A, Walker B, Sensor 1, and Box VII or SRC (containing the arginine finger), and domain 2, encompassing Sensor 2. The canonical sequences of Walker A and Walker B are shown below. Diagram reproduced and adapted from (Davey *et al.*, 2002b), with permission (license number 3591430587500).

1.3.1.2 Bacterial DnaA

DnaA is the replication initiator factor present, typically as a single copy gene, in all known bacteria (reviewed in Iyer *et al.*, 2004; Messer, 2002). DnaA is structurally arranged in four functionally distinct domains (Figure 1.9) (reviewed in Mott and Berger, 2007; Messer *et al.*, 1999). The amino-terminal domain (domain I) is involved in the oligomerisation of DnaA (Simmons *et al.*, 2003) as well as in the interaction of DnaA with the bacterial replicative helicase, DnaB (see below) (Seitz *et al.*, 2000). Adjacent is a flexible and poorly conserved linker element (domain II), whose length and sequence content diverges between different bacterial species (Messer *et al.*, 1999), or can even be absent, as in *Aquifex aeolicus* (Erzberger *et al.*, 2002). Downstream is the signature AAA+ domain (domain III), containing the binding site for ATP and ADP, and is responsible for the protein's ATPase function, which is critical for the initiation of DNA replication (Messer *et al.*, 1999). Both ATP-bound (ATP-DnaA) and ADP-bound (ADP-DnaA) forms of DnaA are present in the cell and are involved in the initiation of DNA replication, although only the ATP-DnaA oligomers are able to effectively initiate DNA replication from the bacterial origin (*oriC*) (Sekimizu *et al.*, 1987). Moreover, domain III also contains another DnaB interaction site (Messer *et al.*, 1999). At the carboxy-terminus is domain IV, which includes a helix-turn-helix (HTH) DNA-binding motif, the DnaA signature sequence that mediates the specific recognition of the DnaA box sequences (DnaA sequence-specific binding sites, see below) in *oriC*, and a basic loop that is essential for DNA binding (Blaesing *et al.*, 2000).

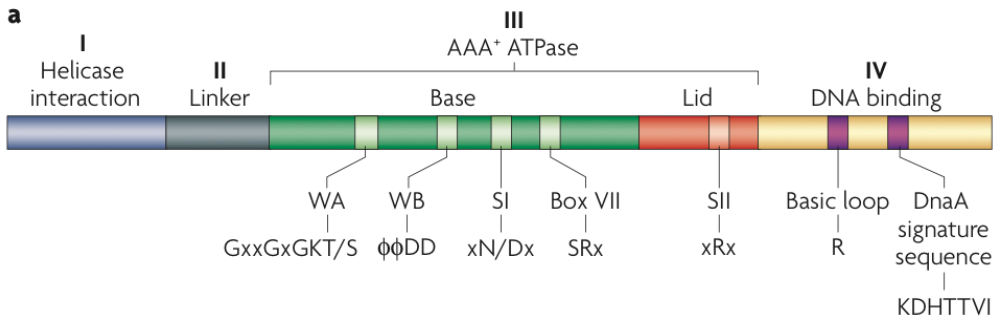


Figure 1.9. Bacterial DnaA protein domains.
Schematic representation of the four protein domains of the bacterial DnaA. The sequences shown for the basic loop and the DnaA signature sequence are the ones from *E. coli* DnaA. Detailed description in the main text. Diagram reproduced from (Mott and Berger, 2007), with permission (license number 3590960451859).

1.3.1.3 Archaeal Orc1/Cdc6

In archaea, the identified initiator factors are distinct from DnaA, and are instead related to the Orc1 subunit of the eukaryotic ORC (see below), as well as one of the eukaryotic ORC-MCM mediators, Cdc6 (reviewed in Kelman and Kelman, 2014). Eukaryotic Orc1 and Cdc6 appear to be paralogues (see below), and archaeal homologues are commonly referred to as the Orc1/Cdc6 proteins (reviewed in Robinson and Bell, 2005). In archaea, the machinery involved in initiation of DNA replication appears to differ greatly amongst the various archaeal lineages, where different groups of organisms may use various combinations of key players in the process. For instance, with the exception the analysed Methanococcales and Methanopyrales species, in which no homologue of Orc1/Cdc6 could be found, meaning a non-orthologous protein is used as the initiator factor (Raymann *et al.*, 2014), all archaeal genomes studied to date contain at least one *orc1/cdc6* gene, with most lineages containing two or three genes encoding Orc1/Cdc6 (Barry and Bell, 2006; Raymann *et al.*, 2014). Indeed, some Halobacteriales species possess large numbers of *orc1/cdc6* genes (Raymann *et al.*, 2014), with as many as 14 in *Haloferax volcanii* (Norais *et al.*, 2007). Because of its homology with eukaryotic Orc1 and Cdc6 proteins, it is generally considered that the archaeal initiation machinery is a simplified version of the eukaryotic one (see below), and it has been hypothesised that the archaeal Orc1/Cdc6 proteins are involved in both origin recognition (a role that has been established) and replicative helicase loading, though how this is achieved remains unclear (reviewed in Kelman and Kelman, 2014).

Studies developed in two of the three Orc1/Cdc6 proteins found in *Sulfolobus solfataricus*, Orc1/Cdc6-1 and Orc1/Cdc6-3 (Dueber *et al.*, 2007; Dueber *et al.*, 2011), and one of the two from *Aeropyrum pernix*, Orc1/Cdc6-1 (Gaudier *et al.*, 2007), revealed that archaeal Orc1/Cdc6 protein structure comprises two domains: an amino-terminal AAA+ ATPase domain, which is where sequence homology is found with bacterial DnaA, and a carboxyl-terminal DNA-binding domain of the winged helix domain (WHD) family. The WHD domain, which allows the direct interaction between Orc1/Cdc6 with the replicative helicase, is the primary DNA-binding element of these proteins, recognising and binding specifically to the origin recognition box (ORB) elements present at the origins (below) (Dueber *et al.*, 2007; Gaudier *et al.*, 2007). However, it has been shown

that DNA interaction is achieved through relatively few base-specific contacts, suggesting that the binding of Orc1/Cdc6 proteins to ORBs may rely on sequence specificity as well as DNA structure (Dueber *et al.*, 2007; Dueber *et al.*, 2011; Gaudier *et al.*, 2007). Surprisingly, these studies have also revealed that the AAA+ domain acts beyond its canonical activity as an ATPase and mediator of the further higher-ordered assembly of initiator factors to the origin: when in its ADP-bound form, the Orc1/Cdc6 AAA+ domain interacts directly with DNA via the ISM in an apparently sequence-independent fashion. This event was shown to cause a bend in the DNA and, together with the WHD domain, to be able to significantly deform and unwind the DNA target (Dueber *et al.*, 2007; Dueber *et al.*, 2011; Gaudier *et al.*, 2007). These observations reinforce the idea that origin recognition by archaeal Orc1/Cdc6 proteins is not solely dependent on specific sequence recognition, as observed in bacteria DnaA interaction with DnaA boxes in *oriC*, but is also influenced by the DNA shape and deformation. This may be analogous to origin binding by ORC in eukaryotes, where there is no evidence for sequence specific DNA binding in any organism other than *Saccharomyces cerevisiae* (see below), placing archaea Orc1/Cdc6 as a possible evolutionary link between the origin binders of bacteria and eukaryotes (Dueber *et al.*, 2011).

1.3.1.4 The Eukaryotic machine: the Origin Recognition Complex

In eukaryotes, it has been assumed that a six-subunit complex, named the origin recognition complex (ORC), first identified in the budding yeast *Saccharomyces cerevisiae* (Bell and Stillman, 1992), acts universally as the replication initiator factor (reviewed in Duncker *et al.*, 2009; Costa *et al.*, 2013). This is mainly because most of the present knowledge of eukaryotic DNA replication has been obtained from studies in model organism such as *S. cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster*, *Xenopus laevis*, mouse and human, all members of the Opisthokonta supergroup (Figure 1.1) (Costa *et al.*, 2013). Recently, however, the fast growing availability of sequenced genomes from organisms across the six eukaryotic supergroups (Figure 1.1) has allowed a more representative analysis of the diversity of the eukaryotic replication machineries (Aves *et al.*, 2012; Tiengwe *et al.*, 2012b). Analysis of 36 genomes has revealed that although the six subunits of ORC (Orc1 to Orc6) were found across the six eukaryotic supergroups, there was not a single ORC subunit

that was present in every single eukaryote species analysed (Aves *et al.*, 2012). More surprisingly, two of these (Orc3 and Orc6), were not found in all the eukaryotic supergroups (Aves *et al.*, 2012). This suggests that there may be variability in the composition of ORC between eukaryotes, or at least greater sequence flexibility that previously appreciated in the six-subunit complex constituted of Orc1 to Orc6 (see below). Nevertheless, because available studies focus almost exclusively on organisms possessing the conserved six-subunit ORC, this is described here in detail.

Five of the ORC subunits, Orc1 to Orc5, appear to be structurally similar, each having a predicted AAA+ ATPase domain, as well as at least one WHD domain (Figure 1.10) (reviewed in Duncker *et al.*, 2009; Costa *et al.*, 2013). However, while the AAA+ ATPase domains in the Orc1, Orc4 and Orc5 subunits are usually conserved and closely related with the one found in the bacterial DnaA, Orc2 and Orc3 appear to have highly divergent AAA+ folds, possessing non-canonical Walker A and Walker B motifs (Figure 1.10, insert box) that are somewhat conserved between the Orc2 and Orc3 subunits of the various species that have been analysed (Speck *et al.*, 2005; Clarey *et al.*, 2006). An insertion has recently been identified in Orc3 (Figure 1.10, only shown in the *D. melanogaster* diagram), between the protein's AAA+ fold and WHD domain, and this was shown to interact with Orc6 within ORC (Bleichert *et al.*, 2013). Orc1, the largest subunit of ORC, possesses an extra domain, absent from all the other Orc proteins, as well as from Cdc6 and the archaeal Orc1/Cdc6 proteins: the bromo-adjacent homology (BAH) domain (reviewed in Li and Stillman, 2012). Though considered a universal feature of all eukaryotic Orc1 subunits (reviewed in Costa *et al.*, 2013; Duncker *et al.*, 2009), where it plays roles in gene silencing and potentially origin recognition (see below), it is now known that this domain is absent from Orc1-like proteins in kinetoplastids (reviewed in Tiengwe *et al.*, 2013), and thus it might not be as universal among eukaryotes as initially thought. Interestingly, Orc1 is more closely related in terms of sequence to the conserved replication factor cell division cycle 6, Cdc6, than to the other Orc proteins, and phylogenetic analysis suggests that the two genes may be paralogues that evolved from an ancestral archaeal-like Orc1/Cdc6 protein (Giraldo, 2003). Like Orc1, Orc4, and Orc5, Cdc6 has a predicted AAA+ ATPase domain as well as a WHD domain (Figure 1.10) (reviewed in Duncker *et al.*,

2009), and plays an important role in the initiation process of DNA replication (see below). In contrast, Orc6, the least conserved of the Orc proteins in both sequence and function (Chesnokov *et al.*, 2003; Prasanth *et al.*, 2002; Semple *et al.*, 2006; Balasov *et al.*, 2007; Chen and Bell, 2011; Liu *et al.*, 2011; Takara and Bell, 2011) does not possess any of these domains, nor shows any evidence of a common evolution with Orc1-5 (reviewed in Duncker *et al.*, 2009). Despite this divergence, Orc6 is central to DNA replication and ORC function (see below).

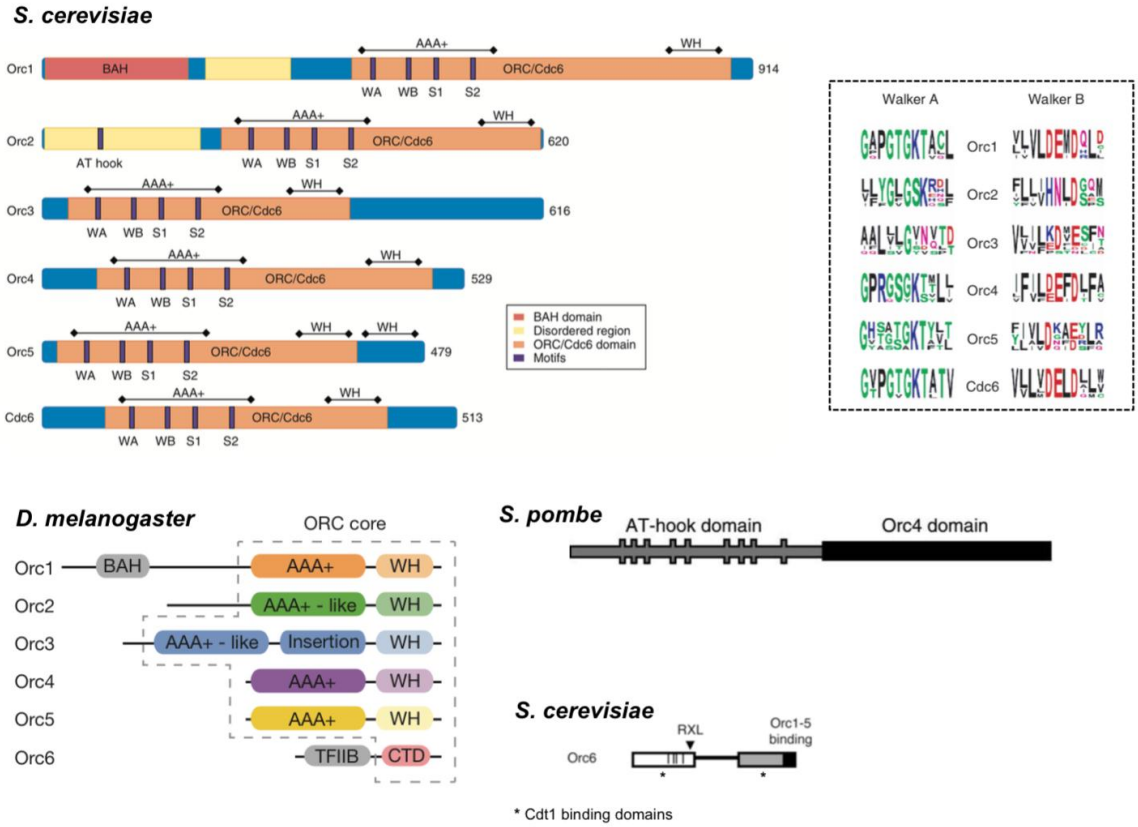


Figure 1.10. ORC subunits.
Top left, schematic representation of Orc1-Orc5 and Cdc6 subunits of *S. cerevisiae*. Below, diagrams representing the domains present in the Orc1-Orc6 subunits of *D. melanogaster*. Next, Orc4 subunit of *S. pombe*, highlighting the 9 AT-hook motifs at its N-terminus. Below, a schematic representation of *S. cerevisiae* Orc6 is shown. Within the insert dashed line box, the sequence conservation within the Walker A and Walker B motifs from various species (*S. cerevisiae*, *S. pombe*, *Homo sapiens*, *Mus musculus, D. melanogaster,* and *X. laevis*) are shown as a sequence logo format (Speck *et al.*, 2005). Note that Orc1, Orc4, and Orc5 sequences coincide with the canonical sequences of these motifs, while Orc2 and Orc3 show highly divergent motifs, though these differences are conserved between the Orc2 and Orc3 sequences analysed (Speck *et al.*, 2005). *S. cerevisiae* Orc1 to Orc5, and Cdc6 diagram reproduced from (Duncker *et al.*, 2009), with permission, © Duncker et al., 2009 BioMed Central Ltd. Sequence conservation diagram reproduced from (Speck *et al.*, 2005) with permission (license number 3592720894270). *D. melanogaster* Orc1 to Orc6 schematics, reproduced from (Bleichert *et al.*, 2015), with permission (license number 3592670461587). *S. pombe* Orc4 diagram reproduced from (Chuang and Kelly, 1999), with permission, © by the National Academy of Sciences. *S. cerevisiae* Orc6 schematics reproduced from (Chen and Bell, 2011), with permission, © Chen and Bell, 2011 by Cold Spring Harbor Laboratory Press.

The most well characterised ORCs are those from *S. cerevisiae*, *S. pombe*, *D. melanogaster* and human, in which all are composed of the canonical six Orc subunits (reviewed in Li and Stillman, 2012), with the Orc proteins of *S. cerevisiae* being usually used to portrait the general structure of the eukaryotic Orc proteins, as shown in Figure 1.10 (reviewed in Duncker *et al.*, 2009). However, there is some diversity in function, regulation and even protein domains between the homologues of these species (reviewed in Li and Stillman, 2012). For instance, the *S. pombe* Orc4 subunit is unique among the Orc4 subunits studied so far: it has an N-terminal extension harbouring nine AT-hook motifs (Figure 1.10), which are responsible for the binding of ORC to the DNA (Chuang and Kelly, 1999; Lee *et al.*, 2001; Gaczynska *et al.*, 2004). The human and *D. melanogaster* Orc6 subunits are structurally related with the transcription factor TFIIB, possessing a TFIIB-like domain (containing a conserved helix-turn-helix motif) at their N-terminal regions (Figure 1.10), potentially acting as a DNA-binding domain (Liu *et al.*, 2011; Balasov *et al.*, 2007; Chesnokov *et al.*, 2003), while its C-terminal domain (CTD) is responsible for interactions with other factors, e.g. Orc3 (Bleichert *et al.*, 2013) and Pnut (Chesnokov *et al.*, 2003) in *D. melanogaster*. In turn, though *S. cerevisiae* Orc6 also appears to have TFIIB-like domains (Bleichert *et al.*, 2013), it does not bind to DNA, but possesses two Cdt1 (helicase loader, see below) binding domains, and a C-terminal region that mediates its association with the other ORC subunits (Figure 1.10) (Chen and Bell, 2011; Takara and Bell, 2011).

Recently, the crystal structure of *D. melanogaster* ORC has been solved (Bleichert *et al.*, 2015), showing that the Orc proteins organise within ORC as follows: Orc1-Orc4-Orc5-Orc3-Orc2 (Figure 1.11, B), with Orc6 bound by its C-terminus to the insertion domain of Orc3 (Bleichert *et al.*, 2015; Bleichert *et al.*, 2013). In human cells Orc6 has also been shown to interact with ORC via Orc3 (Siddiqui and Stillman, 2007), while in budding yeast it is thought that Orc6 interacts with both Orc3 and Orc2 (Bleichert *et al.*, 2013; Sun *et al.*, 2012). This new structural model reorganises Orc2 and Orc3 within the complex, which was previously predicted, by electron microscopy-based techniques, to be Orc1-Orc4-Orc5-Orc2-Orc3 in *D. melanogaster* (Bleichert *et al.*, 2013) as well as in *S. cerevisiae* (Sun *et al.*, 2012), which had suggested a structural similarity between yeast and metazoan ORC complexes (Figure 1.11, A). In the latest

model (Figure 1.11, B), the AAA+ ATPase and the WHD domains co-associate, with the WHD domain of one subunit being packed against the AAA+ ATPase domain of its adjoining partner within the ORC body (the Orc1 WHD is packed into Orc4 AAA+ domain; Orc4 WHD is packed into Orc5 AAA+ domain; Orc5 WHD into Orc3 AAA+ domain; and Orc3 WHD packed into Orc2 AAA+ domain), forming a 'collar' of WHD domains that is rotationally offset from the AAA+ ATPase domains, containing a central channel (Figure 1.11, C) (Bleichert *et al.*, 2015). It has been suggested that this WHD collar of the Orc1-5 proteins interacts with the replicative helicase once this is loaded onto the origin (Figure 1.11, C) (Bleichert *et al.*, 2015), in contrast with the observations made in *S. cerevisiae* where the AAA+ ATPase domains of the Orc proteins latch onto the C-terminal AAA+ motor domains of the MCM2-7 helicase (Sun *et al.*, 2013). Though Orc1, Orc4 and Orc5 are known to bind to ATP in various organisms, Orc1 is the main source of ATPase activity of the complex, and requires the conserved arginine finger of Orc4, which is inserted into Orc1 ATP-binding cleft within its AAA+ ATPase domain, for its catalytic activity (Randell *et al.*, 2006; Bowers *et al.*, 2004; Siddiqui and Stillman, 2007; Klemm *et al.*, 1997; Chesnokov *et al.*, 2001). Similar to what has been reported in yeast (Sun *et al.*, 2012), the *D. melanogaster* ORC structure shows that the arginine finger of Orc4 and the ATP-binding site of Orc1 are too far away from each other to allow co-operative ATP-hydrolysis, suggesting that ORC must undergo conformational changes in order to bring these two regions together and consequently allow Orc1 ATPase activity (Sun *et al.*, 2012; Bleichert *et al.*, 2015). Indeed, the latest model (Figure 1.11, C) (Bleichert *et al.*, 2015) suggests that ORC is first assembled as an auto-inactive/inhibited ATP-bound form that is limited in its capabilities to bind DNA within its central channel or to bind Cdc6 to its ring. Upon activation (possibly through the recognition of origin DNA by Orc6, via its TFIIB-like domains), it is proposed that ORC undergoes conformational changes, including the block movement of the Orc1 AAA+ ATPase domain. This would allow interaction of Orc1 with the Orc4 arginine finger and unlatching of the Orc2 WHD domain, opening a gap in the Orc1-5 ring, and consequent loading of ORC onto the DNA. Cdc6 would then dock between Orc1 and Orc2, closing the gap and trapping the origin DNA within the complex central channel (Bleichert *et al.*, 2015). It has been shown that ORC binding to origin DNA is enhanced by the binding of Cdc6 to Orc1: Cdc6 induces conformational changes on ORC, and its ATPase activity

stabilises the further interaction of ORC specifically with origin DNA, allowing only initiation of replication from these sites (Speck and Stillman, 2007).

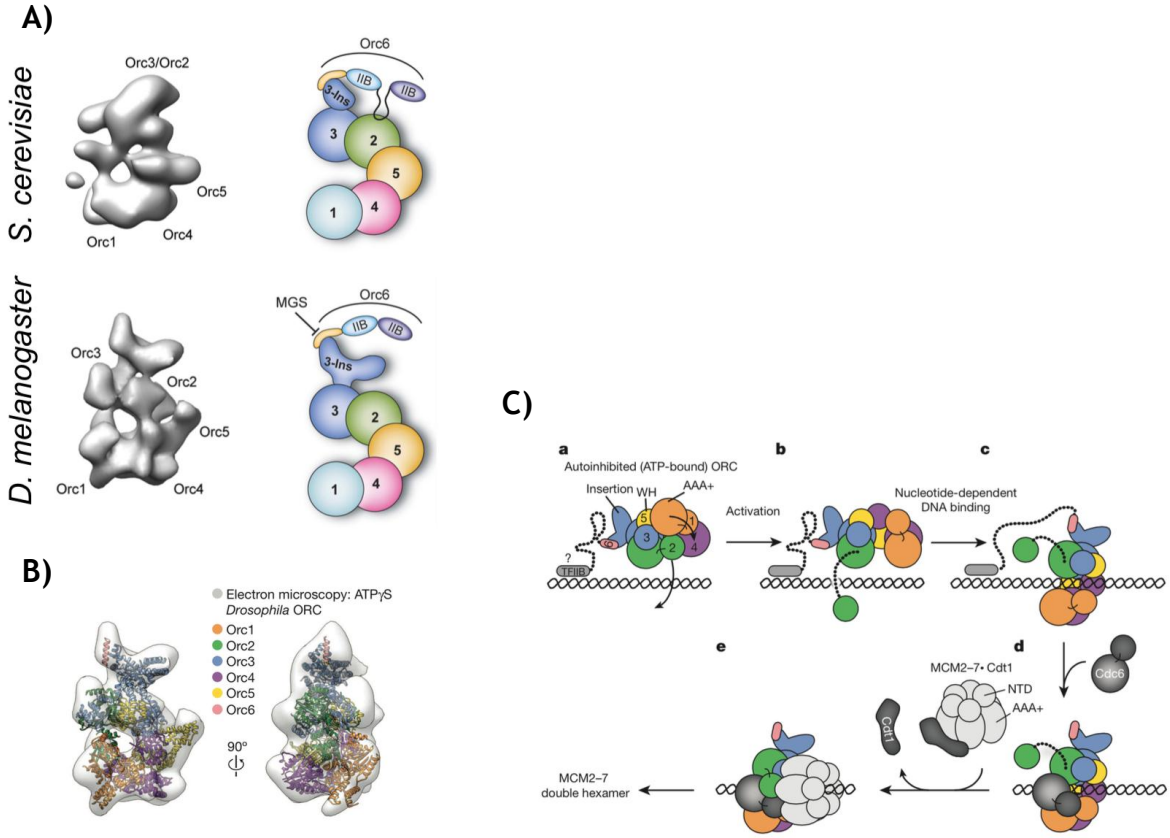


Figure 1.11. ORC models.
A) Comparison of the ORC structures of *S. cerevisiae* (top) (data originally from Speck *et al.*, 2005) and *D. melanogaster* (bottom) (Bleichert *et al.*, 2013) obtained by electron microscopy techniques. Next are shown the models of the different Orc subunits distribution within the ORC complex. Note that the most recent model has shown that, at least in *D. melanogaster* (Bleichert *et al.*, 2015), Orc2 and Orc3 are swapped. B) Crystal structure of *D. melanogaster* ORC, outlined by the electron microscopy image shown in A). C) Model proposed in (Bleichert *et al.*, 2015) for the ORC complex loading in *D. melanogaster*: a) ORC, in its auto-inhibited form, is recruited to the origin DNA, which is potentially recognised by Orc6's TFIIB-like domains; b) ORC is activated, changing its conformation; this repositions Orc2's WHD domain, exposing a gap in the ORC ring; c) the open ring encloses the origin DNA; d) Cdc6 is recruited to the origin where it binds ORC between Orc1 and Orc2, trapping the origin DNA within the complex; e) Cdt1-bound to the replicative helicase MCM2-7 is recruited and loads the helicase onto the origin DNA, using the exposed WHD of the ORC subunits. A) Reproduced from (Bleichert *et al.*, 2013), with permission, © 2013, Bleichert et al. B) and C) Reproduced from (Bleichert *et al.*, 2015), with permission (license number 3593200048352).

1.3.1.5 The replicative helicases

In the three domains of life, the replicative helicase is a six-subunit ring-shaped complex with a central channel through which it unwinds the parental duplex DNA during replication and upon which the replication fork machineries are assembled. The replicative helicase is the first component of the replication fork to be recruited to initiator-bound origins of replication, and its recruitment as

an inactive complex and eventual activation are two of the most tightly regulated events in DNA replication.

In bacteria, the best-studied replicative helicase is the homohexameric DnaB from *E. coli*, and all other bacteria examined so far possess DnaB-like orthologues (reviewed in Li and Araki, 2013; Kaguni, 2011). DnaB monomers have an N-terminal region containing a primase-binding domain, a linker region, and a C-terminal RecA-like ATPase domain (Leipe *et al.*, 2000), enclosing Walker A and Walker B motifs, as well as an arginine finger, which is responsible for ATP binding and association of DnaB with the helicase loader DnaC (see below) (reviewed in Li and Araki, 2013; Kaguni, 2011). The DnaB helicase consists of a ring shaped structure of six DnaB subunits (Bailey *et al.*, 2007; Wang *et al.*, 2008; Lo *et al.*, 2009), oriented in the same direction and aligned side-by-side, through which the single-stranded DNA (ssDNA) passes (Jezewska *et al.*, 1998; Kaplan and O'Donnell, 2002; Gupta *et al.*, 2010). Unlike the archaeal and eukaryotic replicative helicases, the DnaB helicase travels in the 5'-to-3' direction, with the C-terminal regions of each subunit oriented toward the duplex side of the fork, meaning that it unwinds the DNA by translocating along the parental DNA lagging-strand (reviewed in Li and Araki, 2013; Kaguni, 2011).

In archaea, the homohexameric minichromosome maintenance complex (MCM) acts as the replicative helicase (reviewed in Kelman and Kelman, 2014). Like the initiator factors, the archaeal MCM proteins are also AAA+ ATPases, but together with the eukaryotic helicase, MCM2-7 (below), are classified into the Helix-2 insert clade, NtrC/MCM group (Iyer *et al.*, 2004; Erzberger and Berger, 2006). The similarities between the archaeal and eukaryotic MCM subunits suggest that these have most likely evolved from a common ancestor, though eukaryotes have evolved greater complexity, with six distinct MCM subunits (MCM2-7) compared with a single archaeal MCM subunit making up the homohexameric MCM complex (reviewed in Slaymaker and Chen, 2012). All archaeal species studied to date encode for at least one member of the MCM family, but certain archaeal groups, such as the Methanococcales, contain 2 to 8 *MCM* genes (Walters and Chong, 2010), though only one of these appears to be essential and to encode the MCM helicase subunit that acts on DNA replication (Pan *et al.*, 2011; Ishino *et al.*, 2011; Raymann *et al.*, 2014). MCM proteins have a N-terminus domain (NTD)

connected to a C-terminus domain (CTD) by an N-C linker (reviewed in Slaymaker and Chen, 2012; Barry and Bell, 2006; Kelman and Kelman, 2014). The NTD is poorly conserved between MCM proteins in terms of primary sequence, but appears to be structurally and functionally conserved (Fletcher *et al.*, 2003), and plays a role in higher-order structure assembly (reviewed in Slaymaker and Chen, 2012; Bell and Botchan, 2013). The NTD domain is divided into three subdomains: A, which plays a role in regulating the MCM complex (Miller *et al.*, 2014); subdomain B contains a zinc-binding domain, and is involved in DNA binding and in the helicase activity of the MCM complex; and subdomain C that is connected to the AAA+ domain via the N-C linker (reviewed in Slaymaker and Chen, 2012; Bell and Botchan, 2013; Barry and Bell, 2006; Costa and Onesti, 2009). The CTD, also known as the motor domain, as it is responsible for the chemo-mechanical motion of the MCM complex, encompasses the AAA+ ATPase domain and a small subdomain containing a predicted degenerate WHD fold. The AAA+ ATPase domain harbours conserved Walker A (though with minor variations from the canonical sequence, 1.3.1.1) (Iyer *et al.*, 2004) and Walker B motifs, an arginine finger, as well as sensor 1 and 2 motifs (reviewed in Slaymaker and Chen, 2012; Costa and Onesti, 2009). The arginine finger appears to be crucial, as its mutation results in the MCM complex being unable to unwind DNA and hydrolyse ATP (Moreau *et al.*, 2007). In addition, the AAA+ domain harbours DNA-interacting β-hairpin motifs (reviewed in Bell and Botchan, 2013; Costa and Onesti, 2009; Slaymaker and Chen, 2012), including the H2I motif that places the MCM proteins into the Helix-2 insert clade of AAA+ proteins (Iyer *et al.*, 2004). As a complex, the six MCM subunits are organised into a two-tiered ring-shaped structure, with a positively charged central channel wide enough to accommodate either ssDNA or dsDNA. One tier of the structure is made up of the NTDs of the subunits, and the other tier by the AAA+ ATPase domains (reviewed in Costa and Onesti, 2009; Costa *et al.*, 2006). The MCM helicase uses the energy from ATP hydrolysis to translocate over the DNA in the 3'-to-'5 direction (McGeoch *et al.*, 2005; Rothenberg *et al.*, 2007; Graham *et al.*, 2011), and thus translocates along the parental DNA leading strand, displacing the other, and therefore melting the DNA (there are four models on MCM unwinding of the DNA, but they are not discussed here) (reviewed in Bochman and Schwacha, 2009; Slaymaker and Chen, 2012). Presently it is thought that the AAA+ ATPase domains of the six MCM subunits face the DNA duplex junction, with the NTDs

trailing behind (McGeoch *et al.*, 2005). Most of the archaeal MCM complexes appear to exist as single hexamers, though it has been shown that they can exist as head-to-head double-hexamers (like MCM2-7, see below), or even as heptamers and filaments (reviewed in Kelman and Kelman, 2014; Bell and Botchan, 2013; Slaymaker and Chen, 2012).

In contrast with the bacterial and archaeal systems, the core of the eukaryotic replicative helicase is a heterohexameric complex (MCM2-7) composed of six distinct, though evolutionarily related, MCM subunits: MCM2, MCM3, MCM4, MCM5, MCM6 and MCM7 (reviewed in Bochman and Schwacha, 2009; Bell and Botchan, 2013; Costa *et al.*, 2013). Analysis of genomes across the six subgroups of the eukaryotic tree (Figure 1.1) showed that all the six MCM subunits are present in every analysed species, which is consistent with MCM2-7 indispensable role as the eukaryotic replicative helicase (Aves *et al.*, 2012). In terms of protein domains, the eukaryotic MCM2-7 subunits have the basic domain structure as the archaeal MCM counterparts, with conserved AAA+ ATPase domains, but subunit-characteristic N- or C-terminal extensions that are important for the regulation and recruitment of MCM2-7 in eukaryotes (reviewed in Bochman and Schwacha, 2009; Bell and Botchan, 2013), allowing the readily identification of homologues in all eukaryotic species studied to date (Aves *et al.*, 2012). Studies suggest that the MCM2-7 subunits interact in a defined conformation – MCM5-MCM3-MCM7-MCM4-MCM6-MCM2 (Figure 1.12, A) (Davey *et al.*, 2003; Bochman *et al.*, 2008) – either to form a closed or a gapped ring structure (opened between MCM5 and MCM2), with a positively charged central channel (Costa *et al.*, 2011; Costa *et al.*, 2014). As in archaea, the MCM2-7 complex is a two-tier structure, corresponding to the subunits NTDs and the AAA+ ATPase domains (Figure 1.12, B) (Costa *et al.*, 2011). However, MCM2-7 is only the catalytic core of the larger complex that is the active eukaryotic replicative helicase, the CMG complex, which also contains the accessory activating/stimulatory proteins Cdc45 and the four members of the GINS complex (Sld5, Psf1, Psf2 and Psf3) (Moyer *et al.*, 2006; Ilves *et al.*, 2010). Like other ATPases, the ATPase activity of one of the MCM2-7 subunits is dependent on the donation of the arginine finger of the adjacent subunit to its active site (reviewed in Vijayraghavan and Schwacha, 2012). It is thought that Cdc45 and GINS binding to the gapped-ring conformation of MCM2-7 (Figure 1.12, C) (Costa *et al.*, 2011) results in the latter undergoing

conformational changes that enable the combination of the ATP-binding sites and the arginine fingers of the different adjacent subunits to form the active sites, and thus allow ATP-hydrolysis (reviewed in Bell and Botchan, 2013; Vijayraghavan and Schwacha, 2012). Nevertheless, the MCM2-7 complex is first loaded onto ORC-bound origin dsDNA as an inactive head-to-head double-hexamer (Evrin *et al.*, 2013a; Sun *et al.*, 2013; Evrin *et al.*, 2009; Remus *et al.*, 2009) by the helicase loader Cdt1 (Takara and Bell, 2011), and only later does it interact with the remaining members of the CMG complex to initiate DNA replication (see below). Though it is initially bound to dsDNA, after activation by kinases (see below), the origin dsDNA is melted, and the two CMG complexes split apart in opposite directions (Yardimci *et al.*, 2010), where each translocates along ssDNA in the 3'-to-5' direction (meaning along the DNA template leading strand, with the AAA+ ATPase domains of the MCM2-7 subunits facing the DNA duplex) (Figure 1.12, D) as the replication fork progresses (Costa *et al.*, 2014; Fu *et al.*, 2011; Froelich *et al.*, 2014), unwinding the DNA like the archaeal MCM helicase.

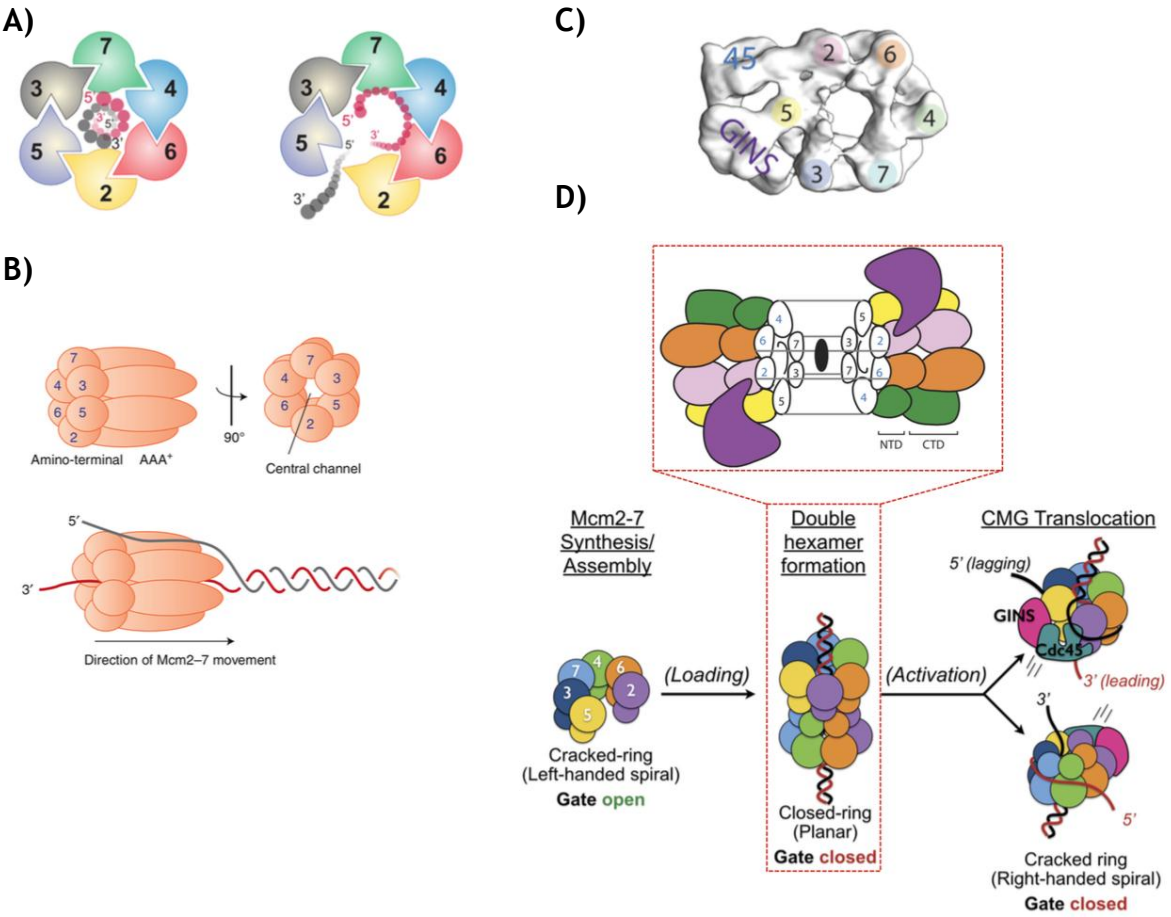


Figure 1.12. The MCM2-7. helicase.
A) Diagram showing the organisation of the MCM subunits within the MCM2-7 complex. Both closed (left) and gaped (right) ring conformations are represented. B) Schematic representation of

the MCM2-7 complex. Circles represent the NTD, while elipses represent the CTD of each subunit containing the AAA+ ATPase domain. Below, diagram representing the directionality of the MCM2-7 complex movement along ssDNA, with the AAA+ ATPase domains facing the DNA helix, and the NTD domains trailing behind. C) CMG structure obtained by electron microscopy, view from the NTD domains of MCM2-7. Cdc45 binds to MCM2, while the GINS complex interacts with MCM5 and MCM3. D) Model of the double-hexamers of MCM2-7 assembly onto dsDNA. The gaped ring conformation is assembled onto the dsDNA as an inactive double-hexamer (accessory proteins are not shown), where each MCM2-7 complex assumes the closed conformation. Upon binding of Cdc45 and the GINS complex, and activation by cell cycle kinases (not shown), the dsDNA melts, and the two CMG complexes split in opposite directions, unwiding the DNA in the bi-directional replication forks. A) reproduced from (Froelich *et al.*, 2014), with permission, © 2013, Froelich et al, all rights reserved. B) reproduced from (Bell and Kaguni, 2013), with permission, © Bell and Kaguni, 2013 by Cold Spring Harbor Laboratory Press, all rights reserved. C) and D) reproduced from (Costa *et al.*, 2014), © 2014 Costa et al., all rights reserved.

1.3.1.6 Initiator factors: more than replication factors

Besides playing essential roles in the initiation steps of DNA replication, several studies have suggested that initiator factors do not act exclusively on DNA replication, and can contribute to further cellular events, such as gene expression, cell cycle progression, chromosome segregation and cytokinesis (reviewed in Scholefield *et al.*, 2011). Before considering the action of origin binding factors in replication initiation, this section addresses some of these further functions.

In Bacteria, DnaA has been shown to be an important transcription regulator by binding to specific DnaA binding sites found upstream, or within, the promoter regions of numerous genes (Goranov *et al.*, 2005). One of these genes is *dnaA* itself, where DnaA acts as a transcription repressor, auto-regulating its expression and thus initiation of replication (Braun *et al.*, 1985; Ogura *et al.*, 2001). DnaA was also demonstrated to regulate, both by activation and repression, the *nrdAB* operon (Olliver *et al.*, 2010) that encodes the ribonucleotide reductase (RNR) enzyme. RNR catalyses the final step in the deoxyribonucleotides (dNTPs) synthesis, and is thus necessary for DNA replication (reviewed in Nordlund and Reichard, 2006), demonstrating that DnaA is involved in the coordination between initiation of replication and nucleotide synthesis. Furthermore, DnaA appears to play a role in the coordination between DNA replication and cell division: DnaA was shown to repress *ftsL* gene transcription upon DNA replication disturbance (Goranov *et al.*, 2005), effectively blocking cell division, a process dependent on FstL (reviewed in Errington *et al.*, 2003).

In Eukaryotes, ORC subunits have been extensively associated with other functions besides DNA replication initiation (reviewed in Duncker *et al.*, 2009; Sasaki and Gilbert, 2007; Scholefield *et al.*, 2011). For example, in human cells Orc1 is involved in the control of centriole and centrosome copy number (Hemerly *et al.*, 2009). Moreover, Orc1 has also been associated with gene silencing through the formation of heterochromatin in diverse organisms. In *S. cerevisiae*, Orc1 interacts with the silent chromatin protein Sir1, recruiting it to the transcriptionally silent mating type loci, HMR and HML, whose silencing is essential for proper haploid cell identity (Hou *et al.*, 2005; Ozaydin and Rine, 2010). Similarly, in *D. melanogaster*, *Xenopous* sp. and mammalian cells, Orc1 interacts with heterochromatin protein 1 (HP1) to promote the spreading of heterochromatin (Pak *et al.*, 1997; Auth *et al.*, 2006). It has also been shown in mammals that a number of ORC subunits (most likely, ORC as a complex) form a complex with the telomere repeat binding factors TRF1 and TRF2 (as well as HP1 and other factors), and bind telomeres via an RNA intermediate, TERRA, influencing heterochromatin formation and telomere capping by the shelterin complex (Deng *et al.*, 2009; Flynn *et al.*, 2011). Orc1 does not solely contribute to transcriptional silencing, however, since *Arabidopsis thaliana* Orc1 (perhaps uniquely amongst eukaryotes) possesses a homeodomain that binds histone H3 trimethylated on lysine residue 4 (H3K4me3) at a number of genes, activating transcription (de la Paz Sanchez and Gutierrez, 2009). Orc6 has been shown to be important in human and *D. melanogaster* cytokinesis, where it appears to coordinate this process with DNA replication: Orc6 depletion by RNA interference (RNAi) leads to a reduction in DNA synthesis and an increase in multinucleated cells that have completed mitosis without cytokinesis (Prasanth *et al.*, 2002). Studies showed that by the end of mitosis, Orc6 localises to the cytokinetic furrow, and directly interacts with the septin protein Pnut, a filament-forming GTPase essential for a successful cytokinesis (Huijbregts *et al.*, 2009). Other subunits of ORC, such as Orc5 or even the whole ORC complex, have also been associated with other cellular events, namely chromosome organisation and segregation (reviewed in Scholefield *et al.*, 2011).

1.3.2 Origins of Replication – where to start?

DNA replication starts at specific genomic sites called origins of replication, and it is to these sites that the initiator factors bind to (reviewed in Leonard and Mechali, 2013). As discussed above, initiator factors are functionally conserved and possess both ATPase and DNA binding domains that are essential for recognition and binding to origins of replication. However, bacterial, archaeal and eukaryotic origins of replication are considerably different: while bacterial and archaeal origins are defined by specific DNA sequences, eukaryotic origins (with the exception of *S. cerevisiae*) are not, and therefore binding of ORC appears to be dependent on other, poorly defined factors, such as chromatin status and structure (reviewed in Leonard and Mechali, 2013; Masai *et al.*, 2010; Mechali, 2010; Mott and Berger, 2007; Costa *et al.*, 2013).

1.3.2.1 The single origin in Bacteria

In all studied bacteria, a single origin of replication, *oriC*, is present per chromosome (reviewed in Messer, 2002; Robinson and Bell, 2005; Mott and Berger, 2007), often adjacent to the *dnaA* gene (Mackiewicz *et al.*, 2004). Typically, *oriC* comprises multiple, repetitive non-palindromic 9 bp sequences – the DnaA boxes, to which DnaA specifically binds. In addition, there is an AT-rich DNA-unwinding element (DUE) (Figure 1.13), whose melting is a key event in the initiation of replication, by facilitating the opening of the dsDNA (Mackiewicz *et al.*, 2004). The *oriC* regions can vary considerably between bacteria: the *oriC* size can fluctuate from 200 bp up to 1,000 bp; the consensus sequence, number and arrangement of DnaA boxes, as well as of repeats present in the AT-rich region, may be quite different; and other DnaA binding sites (e.g. I-sites) may be present, as well as specific binding sites for other regulatory factors (reviewed in Zawilak-Pawlik *et al.*, 2005; Leonard and Mechali, 2013; Costa *et al.*, 2013; Mott and Berger, 2007; Messer, 2002). For example, the *E. coli oriC* is approximately 250 bp and contains five DnaA boxes (Figure 1.13), while *Mycobacterium tuberculosis* and *Streptomyces coelicolor* have, respectively, 550 bp and 930 bp-sized *oriC*s comprising 13 and 19 DnaA boxes, respectively, with completely different spatial arrangements (Zawilak-Pawlik *et al.*, 2005). Interestingly, the differences in DnaA box numbers and distribution along the *oriC* have been demonstrated to be optimally adjusted to their equivalent DnaA proteins, with a

specific DnaA being only able to act on the cognate origin, thus suggesting that both the *oriC* and the initiator factor have evolved towards optimal interaction (Zawilak-Pawlik *et al.*, 2005).

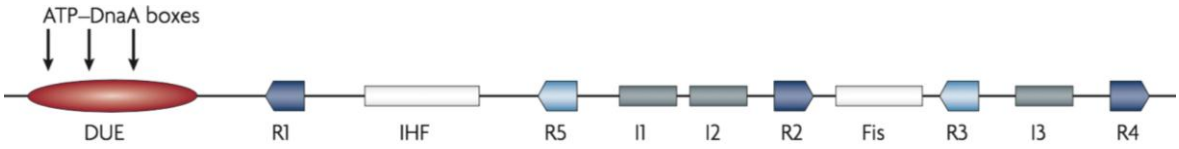


Figure 1.13. Bacterial *oriC*.
Schematic representation of the *E. coli oriC*. The main features such as the DUE, R boxes, I-sites, and ATP-DnaA boxes are shown, and are described in the main text. The τ-sites are not shown. Adapted from (Mott and Berger, 2007), with permission (license number 3594771446008).

In more detail, *E. coli oriC* (Figure 1.13), which has been extensively studied as a model, contains five DnaA boxes (R1-R5) defined by the consensus sequence 5'-TT(A/T)TNCACA-3' (Schaper and Messer, 1995). Of these, R1, R2 and R4 show high affinity for both ATP- and ADP-DnaA forms (Schaper and Messer, 1995), while R3 and R5 show preferentially affinity to ATP-DnaA (Kawakami *et al.*, 2005; Keyamura *et al.*, 2007). In addition, other weaker DnaA binding sites such as the three 9 bp I-sites and the τ-sites, interspersed between the DnaA-boxes (McGarry *et al.*, 2004), and the six 6 bp ATP-DnaA boxes, localised within the AT-rich region (DUE) of *oriC*, are also preferably recognised by ATP-DnaA (Speck and Messer, 2001). Furthermore, *E. coli oriC* also contains recognition sites for initiation of replication regulatory proteins such as Fis (repressor) and IHF (stimulator), which upon binding change the origin DNA conformation, thus affecting DnaA binding (Ryan *et al.*, 2004).

1.3.2.2 A single or multiple origins in Archaea

It is becoming apparent that DNA replication in archaeal organisms possesses a blend of attributes found in bacteria and eukaryotes, and the origins of replication are an example of this variability: although most archaeal circular genomes contain a single origin, several genera have now been shown to replicate their genome from multiple origins of replication (reviewed in Barry and Bell, 2006; Kelman and Kelman, 2014; Leonard and Mechali, 2013). For example, the first origin of replication in archaea was identified in *Pyrococcus abyssi*, and termed *oriC* (Myllykallio *et al.*, 2000). Like in bacteria, *P. abyssi* possesses a single origin containing several repeated sequences (see below) and an AT-rich region, from where replication is initiated (Matsunaga *et al.*, 2001)

and proceeds at a similar rate to that of bacteria (Myllykallio *et al.*, 2000). Moreover, the origin is localised immediately upstream of *P. abyssi*'s single *orc1/cdc6* gene (Matsunaga *et al.*, 2001), analogous to the co-localisation of the bacterial *oriC* and *dnaA*. This would suggest that DNA replication in archaea might be similar to the process in bacteria. However, this view was challenged by studies in *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius* (Robinson *et al.*, 2004; Lundgren *et al.*, 2004), where three origins of replication, named *oriC1*, *oriC2*, and *oriC3*, where identified in their single circular chromosomes. In these archaeal species, three *orc1/cdc6* genes are found, *cdc6-1*, *cdc6-2*, and *cdc6-3* (She *et al.*, 2001). Like in *P. abyssi*, in these *Sulfolobus* species two of the origins are located immediately adjacent to a *cdc6* gene, although the third origin was not associated with a *cdc6* gene, and the *cdc6-2* gene does not co-localise with any origin (Robinson *et al.*, 2004; Lundgren *et al.*, 2004). In addition, it was shown that the replication rate in these *Sulfolobus* species is more similar to rates described in eukaryotic organisms, and slower than that of *E. coli* (Lundgren *et al.*, 2004). Other archaea were also found to possess multiple origins, such as *Aeropyrum pernix*, which has two (Robinson and Bell, 2007), and *Haloferax volcanii*, which has three to five (Norais *et al.*, 2007; Hawkins *et al.*, 2013). It has been proposed that the development of origin multiplicity in archaea might have occurred via the acquisition of extrachromosomal elements, by horizontal gene transfer (Robinson and Bell, 2007).

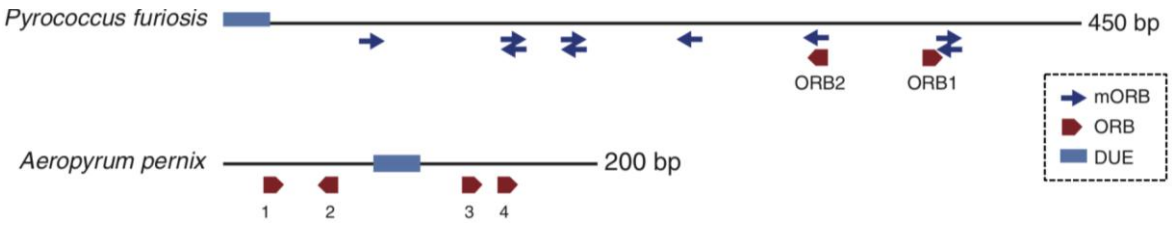


Figure 1.14. Archaeal origins of replication.
Schematic representation of the origin of replication of *P. furiosus* and one of the origins of *A. pernix*. The DUE, ORBs and mini-ORBs (mORB) are shown. Reproduced from (Leonard and Mechali, 2013), with adaptations,

Generically, archaeal origins localise to intergenic regions (frequently near genes of proteins involved in DNA replication, in most cases upstream of a *orc1/cdc6* gene). They are composed of one or more AT-rich sections acting as DUEs, flanked by a variable number (species-specific) of inverted repeated

motifs, named origin recognition box (ORB) elements, which are specifically recognised by the Orc1/Cdc6 proteins and are essential for DNA replication initiation (Figure 1.14) (Myllykallio *et al.*, 2000; Robinson *et al.*, 2004; Berquist and DasSarma, 2003; Grainge *et al.*, 2006). Additional sites have also been found in *Sulfolobus* spp. origins: mini-ORBs (motifs that correspond only to the minimal ORB sequence and are recognised with lower affinity by the Cdc6-1 factor), and C2 and C3 sites (whose sequence is distinct from that of the ORB elements, and are binding sites for the Cdc6-2 and Cdc6-3 factors) (Robinson *et al.*, 2004). Overall, the presence of specific sequences that are recognised by Orc1/Cdc6 proteins, as well as of a DUE, resemble the origins of replication in bacteria, while origin multiplicity in several archaeal genera resemble the eukaryotic systems (see below).

1.3.2.3 The amazing repertoire and plasticity of Eukaryotic origins

In contrast with bacteria and archaea, DNA replication of the eukaryotic large linear chromosomes is initiated from hundreds, and even thousands, of potential origins that, so far, were rarely shown to possess any characteristic sequence elements (reviewed in Leonard and Mechali, 2013; Mechali, 2010; Masai *et al.*, 2010). Origins in *S. cerevisiae* and its close yeast relatives are rare exceptions among eukaryotes: *S. cerevisiae* origins contain specific sequence elements, named autonomous replication sequences (ARS). These contain a specific 11 bp consensus sequence (5'-[A/T]TTTA[T/C][A/G]TTT[A/T]-3') known as the autonomous consensus sequence (ACS), to which ORC specifically binds (Bell and Stillman, 1992), besides other elements close to the ACS that contribute to its activity (reviewed in Leonard and Mechali, 2013). In contrast, in all other eukaryotes studied so far, no consensus sequence elements have been identified that define origins or ORC-binding sites, and the rules that drive specification of initiation sites, as well as the determinants of origin activation timing, are still poorly understood (reviewed in Leonard and Mechali, 2013; Mechali, 2010; Masai *et al.*, 2010; Rhind and Gilbert, 2013). Indeed, though it is clear that ORC binding is needed to designate a sequence as a potential origin, it is insufficient to ensure its activation: only a subset of the ORC binding sites, and thus potential origins, is used per cell cycle to replicate the whole genome. It appears that at the beginning of S phase potential origins are present in large excess, organised in groups that define replicons. Within such groups about one

in five origins is activated per cell cycle in an apparently stochastic manner (Cayrou *et al.*, 2011), providing origin flexibility within each replicon (reviewed in McIntosh and Blow, 2012) and thus adaptation to possible perturbations of the normal cell cycle (see below). Another characteristic of eukaryote replication is that the genome is replicated in a defined temporal sequence, as origins are activated asynchronously during S phase, with some being fired early in S phase and others later (Figure 1.15, A), following a replication timing programme (reviewed in Rhind and Gilbert, 2013) that appears to be specific to different cell types (Ryba *et al.*, 2010). Finally, origin activation follows a tightly regulated temporal plan (reviewed in Jackson *et al.*, 2012), stretching from ORC binding during the G1 phase, to replication initiation in S phase (see below). Overall, many factors appear to contribute to this complex picture, including chromatin structure and status, transcriptional activity, and epigenetic factors such as histone acetylation (Figure 1.15, C) (reviewed in Aladjem, 2007; Masai *et al.*, 2010; Mechali, 2010; Leonard and Mechali, 2013).

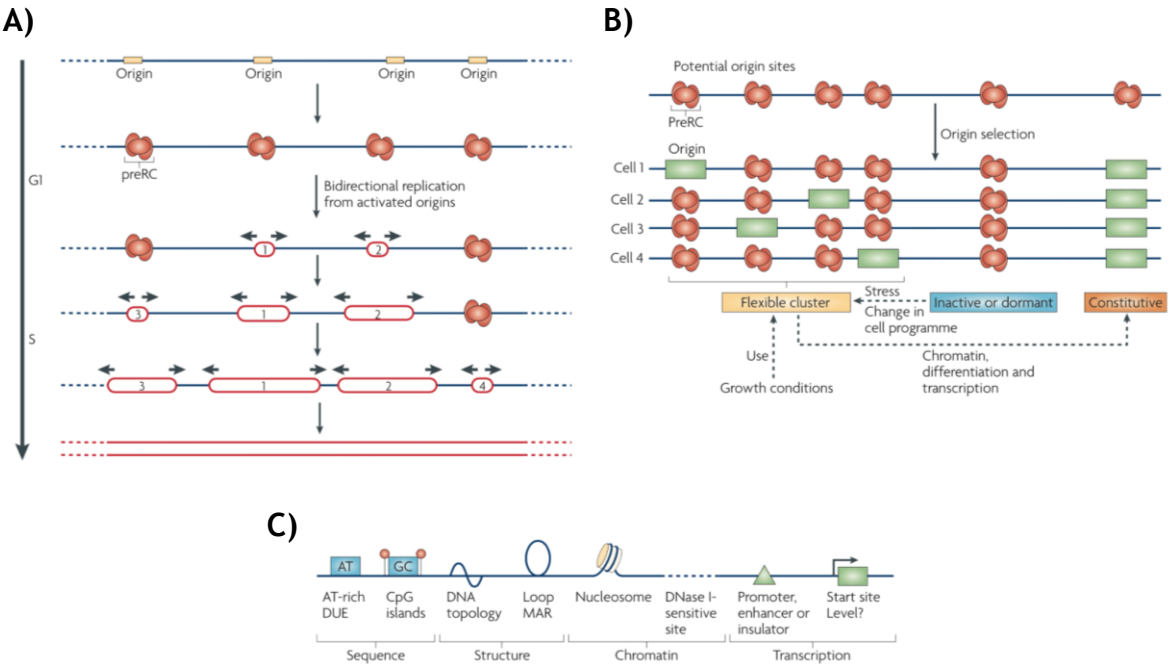


Figure 1.15. Eukaryotic origins of replication.
A) Origins of replication in eukaryotes are activated at different times during S phase. Though all four origins are recognised by the ORC complex, and the pre-replication complex (pre-RC) is established, the origins are activated at different times, with origins 1 and 2 being early firing, while 3 and 4 are late firing origins. The activation of the four origins allows the complete replication of the shown genome segment. B) Only a subset of origins are activated per cell cycle. Though six potential origins are shown (bound by the pre-RC), only some are activated (green rectangles). The different types of origins are depicted: constitutive origins, flexible origins, and dormant/inactive origins. Detailed description in the main text. C) Simplified representation of the different features thought to have a role in defining a DNA region as an origin: sequence features; chromatin structure cues; chromatin status environment; active transcription elements. Description in the main text. All reproduced from (Mechali, 2010), with permission (license number 3595050459007).

DNA replication in eukaryotes presents a high level of plasticity in which origin selection and timing allows the cell to respond to internal and external factors in order to maintain genomic stability. This is possible because not all the potential origins are activated in each cell cycle, or simultaneously. Therefore, depending on their usage, origins of replication can be classified as follows (Figure 1.15, B): constitutive origins, which are consistently used in every cell cycle; flexible origins, which can be used stochastically in different cell cycles; and dormant or inactive origins, which are barely used in normal growth conditions, but can be activated during cell stress conditions, such as replication fork blockage and stalling, or specific development programmes (reviewed in McIntosh and Blow, 2012; Masai *et al.*, 2010; Mechali, 2010). In normal growth conditions, replication forks formed from nearby activated origins passively replicate the origins that have not been activated in that cell cycle (Woodward *et al.*, 2006; Ge *et al.*, 2007).

Although no origin or ORC-binding specific sequences have been found, some features appear to be present at eukaryotic origins (Figure 1.15, C). For instance, in *S. pombe* and *D. melanogaster*, origins localise in intergenic regions and appear to possess characteristic AT-rich islands to where ORC binds (Segurado *et al.*, 2003; MacAlpine *et al.*, 2004; Chuang and Kelly, 1999). Interestingly, studies in *S. pombe* revealed that the origin AT-rich regions are over-represented between divergent transcription units (Segurado *et al.*, 2003), while in *D. melanogaster* two-thirds of ORC-binding sites are present in promoter regions, and some overlap with a subset of RNA polymerase II-binding sites (MacAlpine *et al.*, 2010; MacAlpine *et al.*, 2004). Studies in mouse (Delgado *et al.*, 1998; Sequeira-Mendes *et al.*, 2009) and human cells (Delgado *et al.*, 1998; Cadoret *et al.*, 2008) have also reported origins to localise within or near CpG islands, GC-rich unmethylated regions (where G and C are bound by a phosphodiester (p) bond) that enclose promoters of many mammalian genes (Larsen *et al.*, 1992). Indeed, studies have shown that origins present in or near CpG islands replicate more efficiently than origins situated in GC-poor regions (Sequeira-Mendes *et al.*, 2009; Watanabe *et al.*, 2002). Together, these observations suggest a link between DNA replication initiation and gene transcription. However, this relationship is frequently contradictory, and perhaps more complex than initially thought (reviewed in Helmrich *et al.*, 2013):

in some cases transcription appears to enhance origin activation, while in others seems to inhibit it, or even do not show correlation at all, like in *S. cerevisiae* (Raghuraman *et al.*, 2001; Saha *et al.*, 2004; Norio *et al.*, 2005). Whether the effect of transcription on replication involves protein-protein interaction between transcription and replication factors remains elusive, and some studies suggest that transcription induces alterations in the DNA topology and chromatin status, which may allow replication machinery access to the origin and further activation (reviewed in Aladjem, 2007; Masai *et al.*, 2010; Mechali, 2010; Leonard and Mechali, 2013; Sequeira-Mendes and Gomez, 2012). For instance, in *D. melanogaster* the topology of the DNA appears to influence ORC binding, as it shows a higher affinity for negatively supercoiled DNA (Remus *et al.*, 2004), known to be induced by transcription (Dayn *et al.*, 1992).

Potential origin localisation and activation appears to be greatly influenced by chromatin structure and status. For example, in *S. cerevisiae,* origins are localised in intergenic regions that have an open, nucleosome-free chromatin structure (Yuan *et al.*, 2005) and, when nucleosomes are forcedly positioned over the ACS, origin function can be inhibited (Simpson, 1990). Histone post-translational modifications, such as acetylation, have also been associated with replication origins, influencing the timing of activation (reviewed in Masai *et al.*, 2010). It has also been reported that acetylation of histones H3 and H4 enhances origin activation (Unnikrishnan *et al.*, 2010), while Sir2, a histone deacetylase, was shown to inhibit origin activation from five different origins (Crampton *et al.*, 2008). Similarly, in human cells the histone acetyltransferase HBO1 was shown to be essential for the loading of the replicative helicase to the chromatin (Iizuka *et al.*, 2006; Miotto and Struhl, 2010). Interestingly, it has also been suggested that nuclear positioning of origins during G1 phase may play a crucial role in the determination of the replication timing (Jackson *et al.*, 2012; Aladjem, 2007; Mechali, 2010). In *S. cerevisiae* cells, for instance, origins that are activated late in S phase localise preferentially at the nuclear periphery, while the ones that are activated early localise randomly (Heun *et al.*, 2001). Ultimately, eukaryotic origins appear, unlike origins in bacteria and archaea, not to be clearly defined by sequence elements, but instead to depend on the combination of a variety of less defined features for ORC binding and activation.

1.3.3 Initiation of DNA replication – helicase loading and origin melting

Chromosomal DNA replication must occur strictly once per cell cycle; otherwise, under- or over-replication of genomic DNA will result in genomic instability (reviewed in Abbas *et al.*, 2013), with deleterious effects to the cell. Therefore, DNA replication is a tightly regulated process (reviewed in Siddiqui *et al.*, 2013; Masai *et al.*, 2010; Mott and Berger, 2007), mainly due to the activation and inactivation of the initiator factors and assembly of the pre-replication complexes (pre-RCs). It must also be coordinated with other major cell cycle events in order to ensure complete replication of the genome. As depicted in Figure 1.7, initiation of DNA replication generally refers to the steps from origin recognition to helicase activation. These steps are summarised in the next few sections for the three domains of life.

1.3.3.1 Initiation of DNA replication in Bacteria

Throughout the cell cycle, DnaA is bound to the high affinity R1, R2 and R4 DnaA boxes at *oriC*, and only binds to the remaining DnaA boxes, I-sites and ATP-DnaA boxes within the DUE, immediately before the initiation of DNA replication (Cassler *et al.*, 1995; Speck *et al.*, 1999), when the ATP-DnaA levels increase abruptly, after remaining low for most part of the cell cycle (Kurokawa *et al.*, 1999). Regulation of initiation of replication in bacteria is rather complex, involving multiple intricate mechanisms, from regulation of *dnaA* gene expression, proteins that regulate DnaA activity, *oriC* sequestration by SeqA, methylation of *oriC* DNA, as well as active transcription at *oriC* (not further explored here, but recently extensively reviewed in Skarstad and Katayama, 2013). The cooperative assembly of the DnaA molecules to *oriC* leads to the formation of a large nucleoprotein initiation complex (a helical oligomer capable of ATP-hydrolysis as DnaA molecules donate their arginine fingers to the active site of adjacent DnaA subunits) (Mott *et al.*, 2008; Duderstadt *et al.*, 2010) that promotes the melting of the AT-rich DUE region by destabilizing it (Figure 1.16), generating an open strand-separated complex that is stabilised by the additional binding of ATP-DnaA to the ATP-DnaA boxes therein present (Speck *et al.*, 1999).

After DNA melting, two DnaB hexamers are recruited to the ssDNA in the opened region of *oriC* (Figure 1.16) to allow the formation of bidirectional replication forks (reviewed in Bell and Kaguni, 2013; Robinson and Bell, 2005; Costa *et al.*, 2013). Although DnaA interacts physically with DnaB (Seitz *et al.*, 2000), helicase loading requires the action of DnaC, the helicase loader that, alongside DnaA, belongs to the DnaA/CDC6/ORC clade of the AAA+ superfamily of ATPases (Iyer *et al.*, 2004). Briefly, ATP-DnaC inhibits DnaB helicase activity (Mott *et al.*, 2008; Davey *et al.*, 2002a) and has a high affinity for ssDNA, playing an important role in the loading of DnaB to the melted AT-rich DUE and subsequent expansion of the ssDNA region at the *oriC*. Afterwards, both DnaB and ssDNA presence lead to ATP hydrolysis by DnaC, that in the ADP-DnaC form releases DnaB, relieving the inhibition from the helicase (Davey *et al.*, 2002a; Makowska-Grzyska and Kaguni, 2010). Next, other replication-fork factors, such as the DnaG primase and the DNA polymerase III holoenzyme, are placed on the ssDNA region, starting the bidirectional polymerization of the DNA around the circular genome of *E. coli* (reviewed in Johnson and O'Donnell, 2005).

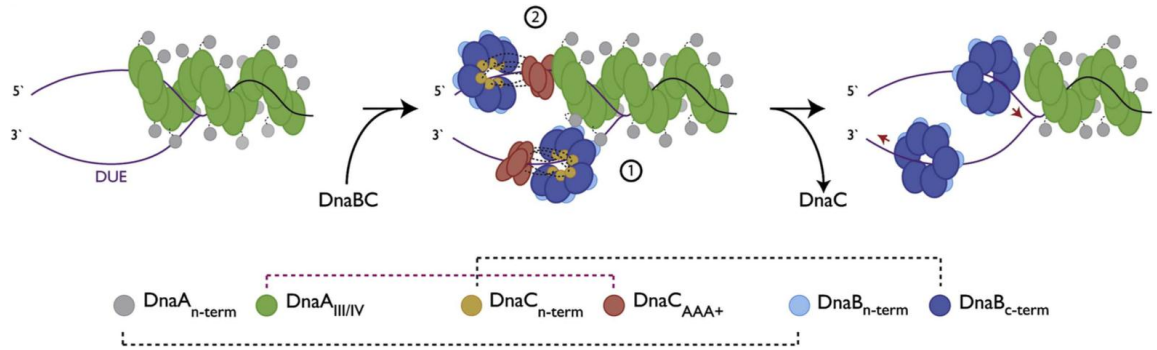


Figure 1.16. Origin melting and helicase loading in bacteria.
Schematic representation of *oriC* melting and helicase loading. Various DnaA molecules are loaded onto the *oriC*, leading to the melting of the DUE (left). DnaB is loaded onto the ssDNA at the melted DUE by DnaC (middle): in (1), DnaB is loaded onto the "bottom" strand where it interacts directly with DnaA; in (2), DnaB is loaded onto the "top" strand via interaction of DnaC with DnaA. DnaC is released, and the DnaB hexamers migrate to their positions in the DUE to allow the assembly of the replication forks (right). Below is represented the colour-coded legend; dashed lines highlight the interactions between the different factors domains. Reproduced from (Mott *et al.*, 2008), with permission (license number 3595340631956).

1.3.3.2 Initiation of DNA replication in Archaea

In archaea, the mechanisms underlying origin binding and helicase recruitment are still being established, and may differ between the various archaeal taxonomic groups, perhaps reflecting differences in the number of encoded

Orc1/Cdc6 factors and origins per chromosome. Studies in *Pyrococcus furiosus* have shown that Orc1/Cdc6 binds to the ORB and mini-ORB elements (Matsunaga *et al.*, 2007), where it induces alterations in the topological structure of the DNA that leads to the melting of the AT-rich region of *oriC*, which is required for the further loading of the MCM helicase (Matsunaga *et al.*, 2010). Furthermore, it was also shown that the recruitment of MCM is dependent on Orc1/Cdc6 (Akita *et al.*, 2010), suggesting that in *P. furiosus* helicase loading might be independent from a loading factor, such as DnaC in bacteria, or Cdc6 and Cdt1 (see below) in eukaryotes. Conversely, studies in multi-origin archaea have revealed a more complex process. For instance, in *S. solfataricus*, replication initiation appears to be extremely complex, with interplay between the three Orc1/Cdc6 proteins with the three origins, each having different and even overlapping binding sites (Robinson *et al.*, 2004; Lundgren *et al.*, 2004; Dueber *et al.*, 2007). Orc1/Cdc6 binding to overlapping ORB sites results in contacts between the ATPase domains of the two different initiator proteins, which is reminiscent of the multimeric interactions seen in eukaryotic ORC (Dueber *et al.*, 2007). In addition, the binding causes DNA distortions, though whether this aids MCM loading is unclear. The overall process, however, appears to be simpler in *A. pernix*, where the AT-rich region is flanked by two pairs of ORBs, to each of which Orc1/Cdc6-1 binds (Gaudier *et al.*, 2007). When the four sites have been bound by Orc1/Cdc6-1, a higher-order assembly of Orc1/Cdc6 is thought to occur at the origin and initiates unwinding of the DUE, allowing further MCM loading (Grainge *et al.*, 2006).

Despite the fact that there is evidence for direct Orc1/Cdc6-MCM interactions in some archaea, an interesting gene has been identified adjacent of one of the origins of both *A. pernix* and *S. solfataricus*. This gene encodes a protein with distant sequence homology to the eukaryotic helicase loader Cdt1. The product of this gene was name winged-helix initiator protein (WhiP), as it has two winged helix-turn-helix domains, like Ctd1 (Robinson and Bell, 2007). The presence of a putative helicase loader in these organisms may place them once again as potential evolutionary links between bacteria and eukaryotes. However, despite the intriguing location of its gene, and evidence that WhiP binds to origins (Robinson and Bell, 2007), functional evidence that it acts as a helicase loader is currently lacking (reviewed in Kelman and Kelman, 2014).

Archaeal MCM has also been shown to interact with a chromatin protein, Alba, which in its deacetylated form strongly inhibits MCM activity (Marsh *et al.*, 2006). Depending on its acetylation state, Alba is able to influence the degree of chromatin packaging (Wardleworth *et al.*, 2002), an additional observation that may reinforce the idea that archaeal origin activation may be susceptible to other factors, such as chromatin structure, like in eukaryotes.

Besides the clear homology with the eukaryotic Orc1 and Cdc6 subunits as well as with the MCM subunits of the helicase, homologues of other eukaryotic replication-associated factors have also been found in archaea, strengthening the idea that the archaeal replication machinery involved in the initial steps of DNA replication might be a simplified version of the eukaryotic one. These factors include orthologues of some subunits of the eukaryotic GINS complex and Cdc45, which in eukaryotes form the active replicative helicase together with MCM2-7 (see below) (reviewed in Kelman and Kelman, 2014). Reflecting the heterogeneity of the archaeal group, some species appear to possess a single orthologue of GINS, labelled GINS15 (or GINS51) due to its similarity with the Psf1 and Sld5 subunits of the eukaryotic GINS, while others also possess another orthologue, GINS23, which shows similarity with both Psf2 and Psf3 subunits of the eukaryotic complex (Raymann *et al.*, 2014; Marinsek *et al.*, 2006). Nevertheless, independently of the presence of one or two GINS homologues, in all cases studied to date, the archaeal GINS proteins form tetrameric complexes (either an homotetramer of GINS15, or a heterotetramer of 2:2 ratio of GINS15 and GINS23) (Ogino *et al.*, 2011; Marinsek *et al.*, 2006), mimicking the tetrameric complex in eukaryotes. In addition, these GINS complexes have been shown to interact with the MCM helicase (Marinsek *et al.*, 2006; MacNeill, 2011; Li *et al.*, 2010) and the orthologue of the eukaryotic Cdc45 and the bacterial RecJ (the GINS associated nuclease – GAN) (Yuan *et al.*, 2013; Krastanova *et al.*, 2012; Li *et al.*, 2011b), among other replication-related factors (reviewed in Kelman and Kelman, 2014; MacNeill, 2010). To date, all analysed species of archaea appear to have a protein with some similarity to Cdc45 (Makarova *et al.*, 2012). However, the role of the Cdc45-like factor is not clear, since gene deletion studies suggest it is not essential for archaeal DNA replication (reviewed in Kelman and Kelman, 2014).

1.3.3.3 Initiation of DNA replication in Eukaryotes and its regulation

When compared to bacteria and archaea, initiation of DNA replication in eukaryotes is a much more complex process: not only are the molecular machineries involved more complex (sections 1.3.1.4 and 1.3.1.5), but there are hundreds to thousands of origins of replication per genome that are not defined by a specific sequence, and the number of players and overall regulation of the process is more intricate (Bell and Kaguni, 2013; McIntosh and Blow, 2012; Jackson *et al.*, 2012; Rhind and Gilbert, 2013; Siddiqui *et al.*, 2013; O'Donnell *et al.*, 2013). Very simplistically, initiation of replication in eukaryotes is regulated at three levels: though all potential origins are recognised by ORC ("where"), only a subset of these is activated ("which") in each individual cell per cell cycle during S phase ("when") (reviewed in Masai *et al.*, 2010). The main events in replication initiation in eukaryotes are the loading and then the activation of the MCM2-7 helicase. In order to guarantee that the genome is replicated once and only once per cell cycle, these two events are tightly regulated and are mutually exclusive, taking place in different stages of the cell cycle: helicase loading is restricted to G1 phase (referred to as pre-RC assembly and origin licensing), while helicase activation and subsequent replication fork assembly take place exclusively during the S phase. Helicase loading requires three factors: ORC, Cdc6 and the helicase loader Cdt1, and these are the steps described below.

The first step is the recognition of all potential origins by ORC (with Orc1 bound to ATP) (Klemm *et al.*, 1997), from late mitosis to the end of G1 phase (Figure 1.17, A). In *S. pombe*, ORC binding to origins occurs in an ATP-independent fashion through its Orc4 AT-hook motif, which recognises AT-rich DNA, and it is to date the only eukaryote where ORC binding does not occur in the ATP-bound form (Chuang and Kelly, 1999; Lee *et al.*, 2001; Kong and DePamphilis, 2001). At least in human cells, ORC assembly and recruitment to the nucleus is a dynamic process: Orc1, Orc6 and an ORC subcomplex comprising Orc2-5 are transported independently to the nucleus of the cell, and only then do they interact at the origins, in an ATP-dependent fashion (Ghosh *et al.*, 2011). Once ORC is bound to the origin, and upon entry into G1 phase, Cdc6 is recruited to the origin (Figure 1.17, A). Cdc6, in its ATP-bound form, binds to the Orc1 subunit (Wang *et al.*, 1999), stabilising the ORC-DNA interaction (Speck and Stillman, 2007), and leading to conformational changes in ORC (Sun *et al.*, 2012). In at least *S.*

cerevisiae, the Cdc6-ORC conformational changes result in the exposure of two Cdt1-binding domains in Orc6 that allow the recruitment and subsequent loading of the MCM2-7 helicase to the origin (Figure 1.17, A) (Takara and Bell, 2011; Chen *et al.*, 2007; Chen and Bell, 2011). Cdt1 is the MCM2-7 helicase loader, and though it is poorly conserved among eukaryotes (Aves *et al.*, 2012), it possesses two characteristic WHD domains at the centre and in the C-terminus (reviewed in Caillat and Perrakis, 2012; Bell and Kaguni, 2013). It is the most C-terminal WHD domain of Cdt1 that interacts with the MCM2-7 helicase (via MCM6, at least in human and mouse cells) (Wei *et al.*, 2010; Zhang *et al.*, 2010; Yanagi *et al.*, 2002). At least in *S. cerevisiae*, Cdt1 and MCM2-7 are only recruited to the origin as a complex (Tanaka and Diffley, 2002), where it binds to ORC via interaction between Cdt1 and Orc6 (Takara and Bell, 2011). It appears that the helicase is loaded onto the origin in pairs, as an inactive head-to-head double-hexamer (Figure 1.17, A), with each MCM2-7 complex being loaded by Cdt1 in a concerted fashion (Remus *et al.*, 2009; Takara and Bell, 2011; Evrin *et al.*, 2009; Fernandez-Cid *et al.*, 2013; Evrin *et al.*, 2013a; Gambus *et al.*, 2011). The assembly at the origin composed of ORC, Cdc6, Cdt1 and the MCM2-7 helicase is the pre-replication complex (pre-RC) (reviewed in Costa *et al.*, 2013).

ATP-binding by Cdc6, as well as its further hydrolysis (stimulated by the binding of Cdc6 to ORC and origin DNA) (Randell *et al.*, 2006; Evrin *et al.*, 2013b) not only induces structural changes in the origin-bound ORC and Cdc6 complex, but also in the structure of the MCM2-7 ring itself: though it is not known when, the MCM2-7 ring must be opened at the MCM2/MCM5 interface in order to encircle origin DNA (Figure 1.17, A) (reviewed in Bell and Kaguni, 2013). A proposed scenario is that when Cdt1 interacts with ORC, it leads to the opening of the MCM2/MCM5 gate in the two MCM2-7 complexes, forming a single channel/opening gate (Figure 1.17, A); subsequently, ATP-hydrolysis by Cdc6 leads to the release of Cdt1 from the pre-RC and consequently, the deposit of the two closed inactive MCM2-7 hexamers around dsDNA (reviewed in Bell and Kaguni, 2013). Next, ATP bound to Orc1 is hydrolysed, resulting in the release of the loaded MCM2-7 double-hexamers from ORC (Figure 1.17, A), and allowing the subsequent loading of more MCM2-7 double-hexamers (repeated loading) (Bowers *et al.*, 2004) to the origin after ADP to ATP exchange. In *S. cerevisiae*, the MCM3 subunit (specially its C-terminus) appears to have a role in the ATP

hydrolysis by Cdc6 and ORC, and is essential for the recruitment of all MCM2-7 subunits to the ORC-binding site (Frigola *et al.*, 2013). Nevertheless, in all eukaryotes, once the helicase has been loaded, the origin is considered licensed, and ready to be activated/fired. As already discussed in section 1.3.2.3, though all potential origins are licensed, only a subset of these are activated during S phase, following a specific activation timing programme, with the non-activated origin pool working as dormant origins that can be activated during S phase if needed. However, the cell must ensure that enough potential origins are licensed by the end of G1 phase before entering S phase (reviewed in McIntosh and Blow, 2012). For instance, in human cells, it appears that the licensing checkpoint delays or even inhibits progression into S phase if an insufficient number of origins have been licensed by the end of G1 phase (Shreeram *et al.*, 2002; Blow and Gillespie, 2008). It is thus crucial that the cell has enough licensed origins to completely replicate its genome, even in cases of replication perturbation.

Next, the helicase must be activated. This is achieved via the recruitment of additional replicative factors to the origin that assemble to form the pre-initiation complex (pre-IC), and is dependent on the levels and functions of two protein kinases, the cyclin-dependent kinase (CDK) and the Dbf4-dependent kinase (DDK, which consists in the catalytic Cdc7 subunit and the regulatory Dbf4 subunit) (Figure 1.17, B) (reviewed in Tanaka and Araki, 2013; Costa *et al.*, 2013). The pre-IC components include the remaining factors that, together with MCM2-7, constitute the replicative helicase (the CMG complex): Cdc45 and the GINS complex (Moyer *et al.*, 2006; Ilves *et al.*, 2010). This stage in the initiation of DNA replication has been extensively studied in *S. cerevisiae*, and it appears to be functionally conserved in metazoans, and potentially in other eukaryotic groups, though the factors and regulatory mechanisms involved might differ (reviewed in Tanaka and Araki, 2013). Taking *S. cerevisiae* as the current best understood model, at the G1 to S phase transition, the levels of Dbf4 increase (and are kept high during S phase before decreasing in late mitosis) (Cheng *et al.*, 1999; Oshiro *et al.*, 1999; Ferreira *et al.*, 2000), and thus DDK activity (Figure 1.17, B). DDK then phosphorylates MCM2-7 (Randell *et al.*, 2010; Bruck and Kaplan, 2009; Ramer *et al.*, 2013; Yeeles *et al.*, 2015), promoting the subsequent interaction of MCM2-7 with Sld3, Sld7 and Cdc45 at the origins

(Figure 1.17, A) (Heller *et al.*, 2011; Tanaka *et al.*, 2011). At the onset of S phase, CDK is activated (Figure 1.17, B), and recruits the GINS complex (Figure 1.17, A), via the interplay with other factors such as Sld2, Dpb11 and DNA Polymerase ε (Muramatsu *et al.*, 2010), to origin-bound MCM2-7 associated with Sld3, Sld7, and Cdc45, by phosphorylating Sld2 and Sld3 (Zegerman and Diffley, 2007; Tanaka *et al.*, 2007). MCM10 is also recruited to the origin (Figure 1.17, A), and appears to be necessary for the unwinding of the origin (van Deursen *et al.*, 2012; Kanke *et al.*, 2012), but its actual role or timing is not clear. The complete replicative helicase (CMG complex) is then activated – most likely via MCM2-7 undergoing conformational changes triggered by the binding of Cdc45 and GINS (Costa *et al.*, 2011; Ilves *et al.*, 2010) – leading to the unwinding of the origin DNA (Figure 1.17, A), separation of the MCM2-7 double-hexamers initially loaded onto the origin (Yardimci *et al.*, 2010; Kang *et al.*, 2012; Costa *et al.*, 2011), and establishment of the bidirectional replication forks (Figure 1.17, A). A key step in the initiation of DNA replication is the switch of the MCM2-7 helicase from encircling dsDNA in the pre-RC to enclosing leading strand ssDNA at the replication fork. However, how this is achieved is presently not clear (reviewed in Costa *et al.*, 2013; Tanaka and Araki, 2013).

It is imperative that any region of the genome is replicated only once per cell cycle, meaning that each origin is activated only once during S phase, and that the whole genome must be completely replicated by the end of S phase, by activating enough origins (reviewed in Diffley, 2011). As detailed above, initiation of DNA replication is separated into two major, non-overlapping, events: helicase loading and helicase activation. In order to ensure that these two steps do not occur simultaneously, thus preventing re-replication, eukaryotes have developed various redundant regulatory mechanisms (reviewed in Diffley, 2010), with different degrees of complexity, depending on the organism. These involve the cell cycle-dependent interplay of various protein kinases and the anaphase–promoting complex/cyclosome (APC/C), an E3 ubiquitin ligase that targets multiple factors for degradation by the proteasome, as well as nuclear exclusion of members of the pre-RC during specific stages of the cell cycle, or their cell cycle-dependent expression (reviewed in Arias and Walter, 2007; Costa *et al.*, 2013; Tanaka and Araki, 2013). In higher eukaryotes, Cdt1 is regulated by another, and additional, type of mechanism: it is

sequestered/inhibited by geminin in S and G2 phases, which prevents MCM2-7 loading to the origins (Wohlschlegel *et al.*, 2000; Lee *et al.*, 2004), and thus re-licensing and re-replication of the origins during S phase. Some of these mechanisms are explored in the discussion of Chapter 3, as they relate directly to the potential function of one *T. brucei* replication factor, TbORC1B.

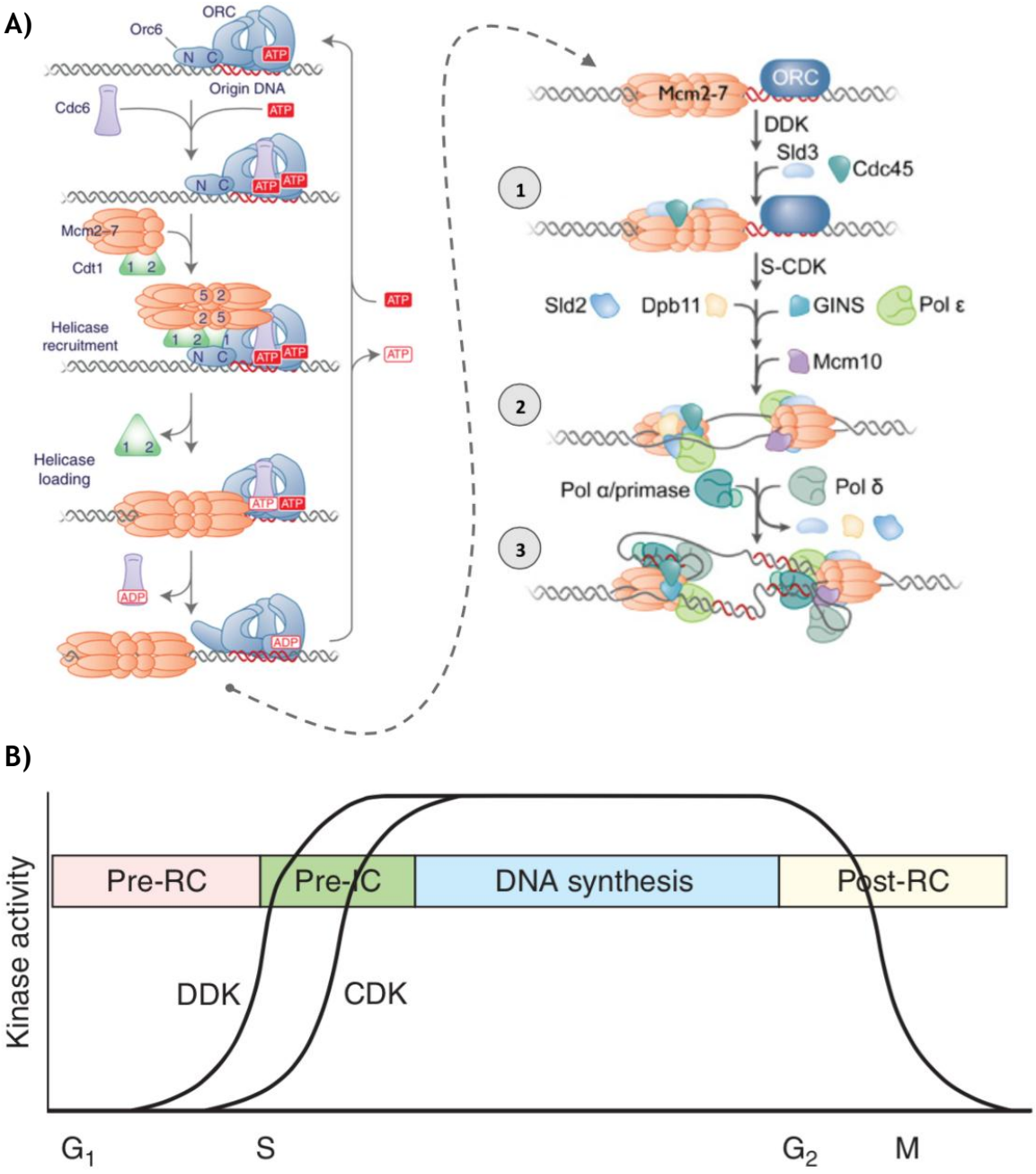


Figure 1.17. Initiation of DNA replication in eukaryotes.
A) Diagram depicting the two crucial events in initiation of DNA replication in eukaryotes (using *S. cerevisiae* as a model): helicase loading (left panel) and helicase activation (right panel). On the left panel, the origin is recognised and bound by ORC from late mitosis to the end of G1 phase. Next, Cdc6 is recruited and binds to ORC, leading to conformational changes in the complex. This exposes Orc6, and Cdt1-binding sites within. Next, the MCM2-7 helicase bound to Cdt1 is recruited to the origin (forming the pre-RC complex with ORC and Cdc6), and loaded onto the origin, in an ATP-dependent fashion, and by opening the gate between the MCM2 and MCM5 subunits. Once the helicase is loaded, the origin is considered licensed. The licensed origin (right panel, dashed arrow) is then activated at the onset (or during) S phase. The MCM2-7 helicase is phosphorylated

by DDK, leading to the recruitment and binding of Cdc45, Sld3 and Sld7 to the helicase (1). The action of CDK then recruits another factors including the GINS complex that together with Cdc45 and MCM2-7 forms the CMG complex, which is the active replicative helicase (2). Once active, the helicase unwinds the origin DNA, allowing the assembly of the replicative fork components and initiation of DNA synthesis (3). B) Diagram depicting the fluctuating levels of CDK and DDK (y-axis) throughout the cell cycle (x-axis). The gaps within which the different replication complexes can be formed are shown. In A) the diagram on the left was reproduced from (Bell and Kaguni, 2013) with permission, © Bell and Kaguni, 2013 by Cold Spring Harbor Laboratory Press, all rights reserved. Diagram on the right was reproduced, with adaptations, from (Heller *et al.*, 2011), Copyright © 2011 Elsevier Inc. All rights reserved. B) Diagram reproduced from (O'Donnell *et al.*, 2013) with permission, © O'Donnel et al., 2013 by Cold Spring Harbor Laboratory Press, all rights reserved.

1.4 DNA replication in *Trypanosoma brucei* and related kinetoplastids – still a field to explore

As evident in the previous sections, present knowledge of DNA replication initiation events in eukaryotes is largely derived from research based on model organisms, all members of the Opisthokonta supergroup (Figure 1.1). Although it has been long assumed that all eukaryotes possess a conserved six-subunit ORC, recent evidence from studies analysing large numbers of organisms from across the different eukaryotic supergroups (Aves *et al.*, 2012), as well as studies on DNA replication in protists such as *Tetrahymena thermophila* (Mohammad *et al.*, 2007; Mohammad *et al.*, 2003; Lee *et al.*, 2015), *Plasmodium falciparum* (Gupta *et al.*, 2008; Gupta *et al.*, 2009), and various kinetoplastids (reviewed in Tiengwe *et al.*, 2013), suggest that the six-subunit ORC might not be as conserved across the eukaryotic tree as initially thought.

One organism that appears to lack a conventional six-subunit ORC is *T. brucei*, where only recently has the process of initiation of DNA replication become the focus of research interest (reviewed in Calderano *et al.*, 2011a; Tiengwe *et al.*, 2013). The next few sections report the present knowledge of how nuclear DNA replication is initiated in *T. brucei* and related kinetoplastids.

1.4.1.1 Do kinetoplastids have an unconventional and divergent ORC?

Until quite recently, virtually nothing was known about how nuclear DNA replication is initiated in *T. brucei* and related kinetoplastids (reviewed in Tiengwe *et al.*, 2013). Upon sequencing of the TriTryp genomes (*T. brucei*, *T. cruzi*, and *L. major*) (Berriman *et al.*, 2005; El-Sayed *et al.*, 2005a; Ivens *et al.*, 2005; El-Sayed *et al.*, 2005b), it became clear that in these parasites the process

of initiation of DNA replication might be different from model eukaryotes. Although an initial genome-wide bioinformatic analysis of these genomes suggested that the molecular machinery involved in replication fork assembly and DNA synthesis resembled that found in model eukaryotes, with orthologues of the MCM2-7 helicase, Cdc45 and the sliding clamp proliferating cell nuclear antigen (PCNA) identified (El-Sayed *et al.*, 2005a), only one potential ORC subunit, a putative orthologue of Orc1 (which, like in other eukaryotes, also shows homology with Cdc6), was identified (El-Sayed *et al.*, 2005a). This protein was subsequently named TbORC1/CDC6 in *T. brucei* and TcORC1/CDC6 in *T. cruzi* (Godoy *et al.*, 2009), and Orc1 in *Leishmania* spp. (Kumar *et al.*, 2008). Together with the lack of identifiable orthologues of key factors such as Cdc6, Cdt1, the GINS complex, MCM10, and the DDK kinase, these observations lead to the hypothesis that initiation of DNA replication in these three kinetoplastids could be more similar to the process found in archaeal organisms rather than the one found in model eukaryotes (El-Sayed *et al.*, 2005a; Godoy *et al.*, 2009).

This hypothesis has sparked research interest, not just because the molecular machinery might be divergent and make the evolutionary link between archaea and eukaryotes, but also because the process of initiation of DNA replication might be a potential target for drug development against these pathogenic parasites (reviewed in Calderano *et al.*, 2011a; Tiengwe *et al.*, 2013). Consequently, in the last few years, orthologues of several factors involved in the late steps of the initiation of DNA replication have been identified and experimentally validated in *T. brucei* (reviewed in Tiengwe *et al.*, 2013). Interaction and functional analysis has been reported for the MCM2-7 complex (Tiengwe *et al.*, 2012b; Dang and Li, 2011), Cdc45 (Oyola *et al.*, 2009; Dang and Li, 2011) and the four members of the GINS complex (Dang and Li, 2011). Localisation of PCNA has been described (Kaufmann *et al.*, 2012), and sequence analysis of MCM10 and MCM9 reported (Liu *et al.*, 2009). Finally, analysis of MCM2-7 supplementary factors, MCM8 and the MCM-binding protein (MCM-BP), has been described in the context of antigenic variation (Kim *et al.*, 2013) and a recent study has identified the highly diverged components of the *T. brucei* kinetochores (Akiyoshi and Gull, 2014).

Identification of the earliest acting ORC components, as well as how MCM2-7 is recruited to the pre-RC, and thus testing if TbORC1/CDC6 alone provides ORC function, has proved challenging. Sequence-based searches have failed to identify clear kinetoplastid orthologues for many members of the putative pre-RC, as no orthologues for Cdc6 and Cdt1, as well as Orc2, Orc3, Orc4, Orc5 and Orc6, were detectable. Instead, four putative ORC-like factors have been identified in *T. brucei* and shown to interact with TbORC1/CDC6 in PCF cells (Tiengwe *et al.*, 2012b; Dang and Li, 2011). One such factor was initially identified through a search of the *T. brucei* genome using both human and yeast ORC proteins as queries, revealing a gene encoding a protein with very low level homology to Orc1; this was suggested to be a divergent, second Orc1-like factor, and was named TbORC1B (Dang and Li, 2011). Identification of TbORC1/CDC6 interacting partners by immunoprecipitation-based assays revealed one protein that displays low level homology with *D. melanogaster* Orc4 subunit, suggesting it may be an Orc4 orthologue, leading to its naming as TbORC4. The same interaction analysis also revealed two apparently kinetoplastid-specific factors, Tb7980 and Tb3120 (Tiengwe *et al.*, 2012b). Orthologues of these four TbORC1/CDC6 interacting factors are also found in both *T. cruzi* and *L. major* (Tiengwe *et al.*, 2012b; Dang and Li, 2011).

Protein sequence analysis of these factors revealed that both TbORC1/CDC6 and TbORC1B lack the Orc1-typical BAH domain, although both appear to possess an AAA+ ATPase domain, containing well conserved Walker A and Walker B motifs, though TbORC1B appears to lack the characteristic arginine finger motif (Figure 1.18) (Godoy *et al.*, 2009; Dang and Li, 2011). Experimental data using recombinant protein was able to confirm that TbORC1/CDC6 and TcORC1/CDC6 are ATPases (Godoy *et al.*, 2009), while TbORC1B did not display such activity (Dang and Li, 2011). In addition, a NLS motif was found in TbORC1/CDC6's N-terminus, while a WHD domain has also been predicted in its C-terminus (Figure 1.18) (Godoy *et al.*, 2009), though such motif and domain have not been found in TbORC1B or any of the other putative ORC-like factors (reviewed in Tiengwe *et al.*, 2013). Furthermore, in the case of TbORC1/CDC6 and TcORC1/CDC6, both were shown, through a yeast phenotypic complementation assay, to be able to replace Cdc6's function in *S. cerevisiae*, but were not able to replace Orc1 (perhaps because *S. cerevisiae* Orc1 recognises specific sequences at the origins

of replication that the kinetoplastid ORC1/CDC6 protein cannot). This result was taken as evidence that TbORC1/CDC6 and TcORC1/CDC6 might have a dual role (of both Orc1 and Cdc6) in DNA replication (Godoy *et al.*, 2009). TbORC4 also appears to have an AAA+ ATPase domain, although both Walker A and Walker B motifs appear degenerate (Figure 1.18) (Tiengwe *et al.*, 2012b). Nevertheless, TbORC4 seems to possess a conserved arginine finger motif commonly found in other organisms' Orc4 subunits (Tiengwe *et al.*, 2012b). Although never tested, these observations suggest that TbORC4 does not possess ATPase activity, and might instead stimulate TbORC1/CDC6 ATPase activity, similarly to what has been observed for the Orc4 subunits of other eukaryotes (reviewed in Davey *et al.*, 2002b; Duncker *et al.*, 2009). Analysis of Tb7980 and Tb3120 protein sequences did not reveal significant homology with any ORC factors (the detection of non-significant Orc5 and Orc2 motifs, respectively, is discussed in detail in Chapter 3), though Tb7980 appears to have a putative AAA+ ATPase domain with relatively conserved Walker A and Walker B motifs (Figure 1.18) (Tiengwe *et al.*, 2012b; Tiengwe *et al.*, 2013). From the available data, it is clear that TbORC1/CDC6 is the most conserved Orc-like protein identified to date in *T. brucei*, and might explain why the remaining Orc-like factors have not been identified by bioinformatic analysis performed in earlier studies (El-Sayed *et al.*, 2005a).

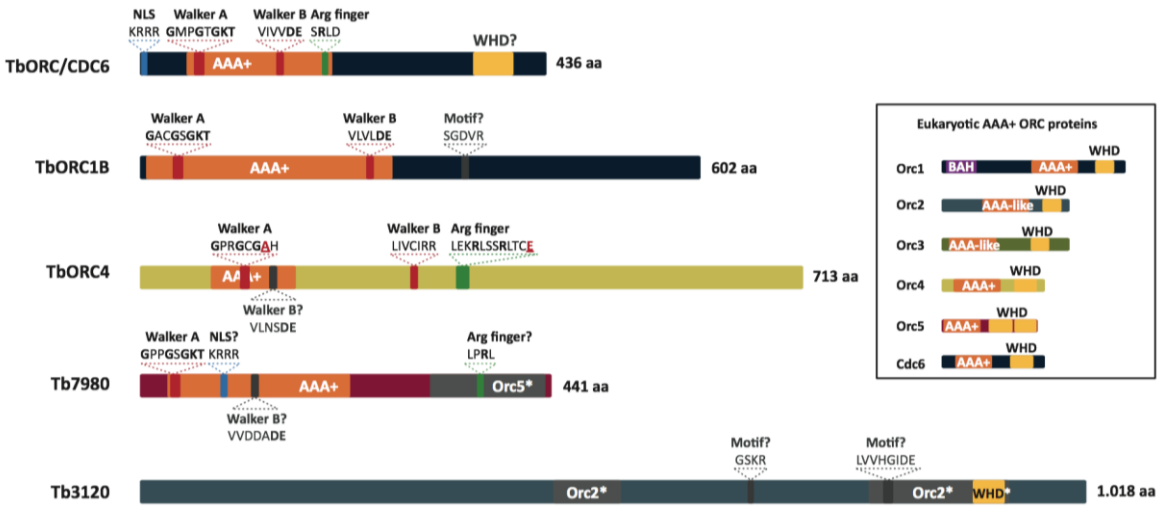


Figure 1.18. TbORC1/CDC6 and interacting factors putative domains.
Schematic representation of the different putative domains that have been identified to date in TbORC1/CDC6, TbORC1B, TbORC4, Tb7980 and Tb3120. Domains or motifs highlighted with (*) were identified as being non-significant. Within the insert box is a simplified diagram of the domains found in model eukaryotes ORC proteins, as in (Duncker *et al.*, 2009). Reproduced from (Tiengwe *et al.*, 2013),

Individual RNAi targeting TbORC1/CDC6 (Godoy *et al.*, 2009; Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013), TbORC4, Tb7980 or Tb3120 (Tiengwe *et al.*, 2012b), using tetracycline-inducible RNA interference (RNAi) systems (detailed in Chapters 3 and 4) in both *T. brucei* PCF and BSF cell types resulted in similar cell cycle and growth defects. However, TbORC1/CDC6 is presently the only factor that has been directly shown to be involved in the process of nuclear DNA replication (Benmerzouga *et al.*, 2013), and it will be essential to assess whether TbORC1B, TbORC4, Tb7980 and Tb3120 also have roles in DNA replication. Two aspects of the observed RNAi phenotypes are noteworthy. First, the effects seen after RNAi in PCF *T. brucei* are remarkably mild, with slowing of growth seen after around 4 days (Tiengwe *et al.*, 2012b; Godoy *et al.*, 2009), which is surprising, as replication factors would be predicted to be essential. Second, the phenotypes observed in PCF cells are distinct in severity and form to those reported for BSF cells (Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013). While in the former stage of the parasite's life cycle silencing of TbORC1/CDC6, TbORC4, Tb7980 and Tb3120 consistently results in the accumulation of enucleated cells (0N1K, termed zoids) (Godoy *et al.*, 2009; Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013), in BSF cells impairment of TbORC1/CDC6, TbORC4 and Tb7980 expression leads to the accumulation of cells with multiple nuclei and kinetoplasts (Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013). This disparity might result from differences between the two life cycle stages cell cycle checkpoints, as already discussed in section 1.2.2 (reviewed in Hammarton, 2007; Li, 2012), and are exhaustively discussed in Chapters 3 and 4. To date, the role of TbORC1B in DNA replication or the effects of its depletion have not been investigated.

Very little is known about these proteins' expression and dynamics throughout the cell cycle. Presently, sub-cellular localisation studies have been mostly limited to TbORC1/CDC6 in *T. brucei* PCF cells (Godoy *et al.*, 2009), TcORC1/CDC6 in *T. cruzi* epimastigote cells (Godoy *et al.*, 2009), and *L. major* Orc1 (herein referred to as LmORC1/CDC6) in promastigote cells (Kumar *et al.*, 2008), all replicative cell types in the insect host. Immunolocalisation of TbORC1/CDC6 has shown that, like TcOrc1/Cdc6 in *T. cruzi*, it is expressed throughout the cell cycle, during which it localises exclusively to the nucleus, where it is bound to chromatin (Godoy *et al.*, 2009). Likewise, LmORC1/CDC6

has been shown to localise to the nucleus of the cell throughout the cell cycle, but whether it is constantly bound to chromatin is presently unknown (Kumar *et al.*, 2008). More recent work, however, has shown that TcORC1/CDC6 localisation in the nucleus of *T. cruzi* epimastigotes is not static: during S phase TcORC1/CDC6 accumulates at the nuclear periphery, while throughout the other cellular phases is dispersed in the nucleus (Calderano *et al.*, 2011b). Whether this is also observed in both *L. major* and *T. brucei* counterparts requires further investigation, as this phenomenon has not been reported in microscopic observations of LmORC1/CDC6 and TbORC1/CDC6 made so far (Kumar *et al.*, 2008; Godoy *et al.*, 2009). This may suggest that TbORC1/CDC6 regulation, as well as its counterparts in *T. cruzi* and *L. major*, if existent, is not being controlled by proteosome-mediated degradation or nuclear exclusion, as it is observed in other eukaryotes (reviewed in Costa *et al.*, 2013; Arias and Walter, 2007; Siddiqui *et al.*, 2013), and therefore, other factors might be involved in the activation/repression of origin firing. Recently, TcORC1/CDC6 has also been localised in different life cycle stages of *T. cruzi*: whereas it is nuclear in replicative stages, it is absent from the nucleus in non-replicating cell types (Calderano *et al.*, 2014). Presently, there is no cellular localisation data for any of the remaining ORC-like factors, and it is unknown whether any of them might be performing a regulatory role. Nevertheless, evidence from *T. brucei* factors acting in the late steps and upstream of the pre-RC assembly might shed some light on the regulation of DNA replication in *T. brucei* (see below).

It has been proposed that TbORC1B, TbORC4, Tb7980 and Tb3120 interact with TbORC1/CDC6 (Tiengwe *et al.*, 2012b; Dang and Li, 2011) and form a diverged, potentially 5-subunit ORC-like complex (Li, 2012). However, to date, such putative ORC-like complex has not been isolated in *T. brucei*, and neither have TbORC1B, TbORC4, Tb7980 and Tb3120 been shown to interact between each other. Therefore, it remains possible that these factors provide distinct, non-ORC functions (Tiengwe *et al.*, 2013).

1.4.1.2 The *T. brucei* replicative helicase and other members of the pre-IC complex

As already mentioned, with the exception of ORC, Cdc6 and Cdt1, most of the factors involved in steps downstream of ORC recognition and binding to the

origin of replication have been identified in *T. brucei*. Consequently, the orthologues in both *T. cruzi* and *L. major* were also identified, suggesting that the main difference between kinetoplastids and other eukaryotes relies on the early acting components of the initiation of DNA replication. Below is a description of the present knowledge available on the kinetoplastid orthologues of the pre-IC and few early recruited members of the replication fork.

Orthologues of the six subunits of the MCM2-7 helicase were readily identified in an initial *in silico* analysis (El-Sayed *et al.*, 2005a), and were later experimentally confirmed by two independent studies (Tiengwe *et al.*, 2012b; Dang and Li, 2011). Together, these suggest that *T. brucei* has a typical eukaryotic replicative helicase, with the six subunits (TbMCM2 to TbMCM7) appearing to possess conserved AAA+ ATPase domains, as well as the signature arginine finger (Tiengwe *et al.*, 2012b; Dang and Li, 2011). As discussed in section 1.3.1.5, the MCM2-7 hexamer is organised in a highly defined geometry: MCM5-MCM3-MCM7-MCM4-MCM6-MCM2, and that it can exist in both closed or open ring conformations, with a gate between the MCM2 and MCM5 subunits (Davey *et al.*, 2003; Costa *et al.*, 2011; Costa *et al.*, 2014; Bochman *et al.*, 2008). Therefore, it would be predicted that *T. brucei* TbMCM2-7 subunits interact in a similar fashion. Using both immunoprecipitation (IP) assays in PCF cells and yeast-2-hybrid techniques, which have been used in previous studies to understand the interactions between the MCM2-7 subunits within the hexameric complex in other eukaryotes (Yu *et al.*, 2004; Crevel *et al.*, 2001), the two aforementioned studies were able to map the interactions between the *T. brucei* TbMCM2-7 subunits (Tiengwe *et al.*, 2012b; Dang and Li, 2011), which overall appear to match the ones mapped in other eukaryotes (reviewed in Bochman and Schwacha, 2009). However, the interaction patterns obtained in the two studies were not completely coincident (Tiengwe *et al.*, 2012b; Dang and Li, 2011), and therefore, at present, the interacting network within the TbMCM2-7 complex is not completely clear. For instance, in both studies, there were at least two adjacent subunits in the MCM2-7 ring that did not interact in one study, namely TbMCM3 and TbMCM7 (Tiengwe *et al.*, 2012b), and TbMCM7 and TbMCM4 (Dang and Li, 2011), but were shown to interact in the other, perhaps simply due to technical differences between the two studies, rather than biological differences (Figure 1.19, A and C) (Tiengwe *et al.*, 2012b; Dang and Li,

2011). A detailed network of the interactions mapped in both studies (Tiengwe *et al.*, 2012b; Dang and Li, 2011) is shown in (Figure 1.19, A and C), and is not further discussed here in detail.

Because to date no homologues of Cdt1 have been found in *T. brucei*, *T. cruzi* or *L. major*, it is possible that, like in some archaeal organisms, ORC may interact directly with the helicase, recruiting it to the origin independently of a loading factor. Besides mapping the interactions within the TbMCM2-7 complex, these studies have also mapped the interactions of TbORC1/CDC6 and TbORC1B with the TbMCM2-7 complex in *T. brucei* PCF cells via co-immunoprecipitation (co-IP) assays (Tiengwe *et al.*, 2012b; Dang and Li, 2011). Again, the results presented by both works are not congruent. In (Tiengwe *et al.*, 2012b), no interactions were observed between TbORC1/CDC6 and TbMCM3, TbMCM6 or TbMCM7, possibly suggesting that an unknown factor, functionally similar to Cdt1 but highly divergent in terms of protein sequence, may be mediating the interaction between the initiator factor at the origin and the core replicative helicase. On the other hand, both TbORC1/CDC6 and TbORC1B were shown to interact with TbMCM3 in (Dang and Li, 2011). It is, however, not known if either TbORC1/CDC6 or TbORC1B interact with other subunits of MCM2-7, as this was not tested (Dang and Li, 2011). Nonetheless, whether the interactions between initiator factors and the helicase occur already at the origin or if TbORC1/CDC6 or TbORC1B act as the helicase loader, it is not known. Again, the discrepancies between the two studies are most likely due to technical differences, but still, require further investigation for consistency. It is not known if other TbORC1/CDC6 interacting partners (TbORC4, Tb7980 and Tb3120) also interact with the helicase, or whether any of these has the potential to act as a helicase loader, and thus functionally replace Cdt1.

In addition to the six MCM2-7 subunits, orthologues of the four member of the GINS complex (Sld5, Psf1, Psf2 and Psf3), as well as Cdc45 have also been identified, and their interactions within the CMG complex mapped in *T. brucei* (Dang and Li, 2011). Via a yeast-2-hybrid assay, it was possible to infer the following interactions: TbMCM5 interacts with the four subunits of TbGINS; all four TbGINS subunits interact with each other, perhaps forming a tight complex; TbSLD5 interacts with the six subunits of TbMCM2-7; and TbPSF1 interacts only

with TbMCM3, TbMCM5 and TbMCM6 (Figure 1.19, C) (Dang and Li, 2011). This complex network of protein-protein interactions may suggest that both TbMCM5 and TbSLD5 are possibly located in the interface between the two complexes (Dang and Li, 2011). Indeed, it has been shown in *D. melanogaster* that the GINS complex binds to MCM2-7 via MCM3 and MCM5 (Costa *et al.*, 2011). TbCDC45 interactions with TbMCM2-7 and TbGINS were further inferred through a GST-pulldown assay (Dang and Li, 2011). This showed that TbCDC45 interacts with all the four subunits of TbGINS, as well as TbMCM2, TbMCM4 and TbMCM5 (Figure 1.19, C), analogous to what was reported in *D. melanogaster* (Costa *et al.*, 2011; Costa *et al.*, 2014), where Cdc45 has been described as being the bridge between the MCM2-7 and GINS complexes within the CMG complex, thus closing the MCM2-MCM5 gate. Furthermore, using recombinant proteins, the *T. brucei* CMG complex was shown, in contrast to TbMCM2-7 alone, to possess helicase activity, suggesting that association of both TbGINS and TbCDC45 with TbMCM2-7 is required to activate the latter's helicase activity, at least *in vitro* (Dang and Li, 2011).

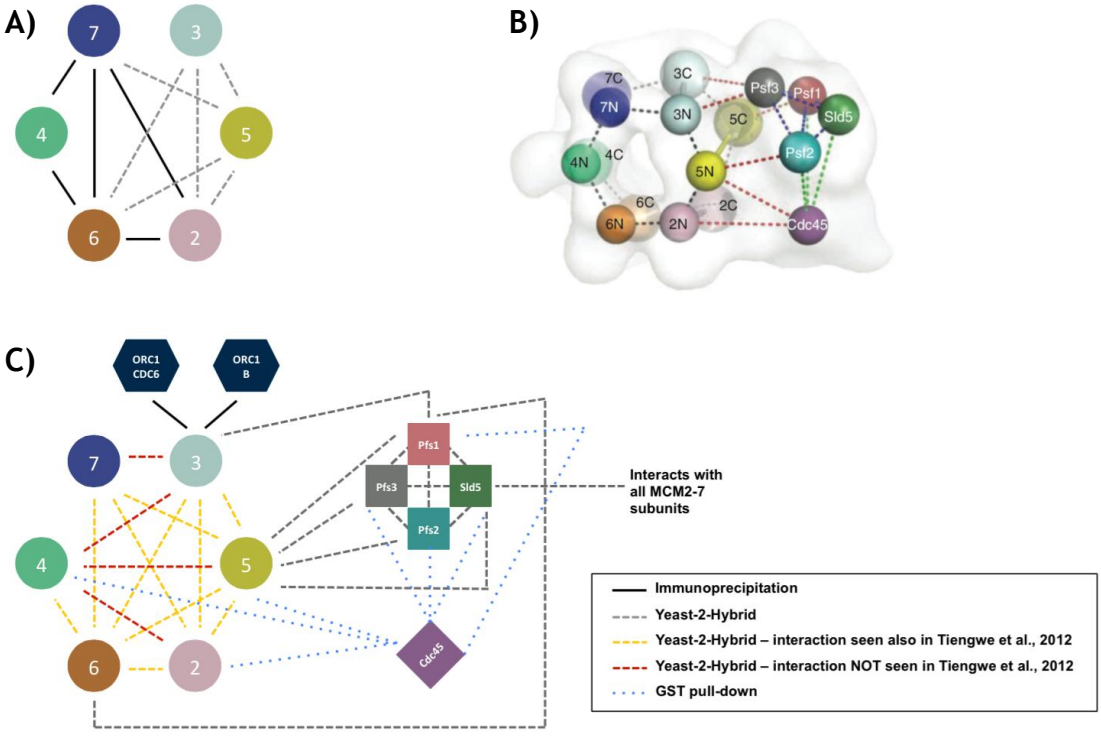


Figure 1.19. Protein-protein interactions within TbMCM2-7 and the TbCMG replicative helicase.
A) Interactions within the TbMCM2-7 complex as described in (Tiengwe *et al.*, 2012b); filled lines represent interactions mapped by IP, while grey dashed lines represent interactions mapped via yeast-2-hybrid assays. B) Representation of the interactions within the *D. melanogaster* CMG complex. Reproduced from (Costa *et al.*, 2011), with permission (license number 3597820018916). C) Schematic representation of the interactions within the *T. brucei* CMG complex as mapped in (Dang and Li, 2011). Legend of the interactions in A) and C) is shown within the insert box.

In addition to mapping the interactions between the *T. brucei* CMG subunits, the effects of RNAi, as well as their localisation throughout the cell cycle were also investigated (Dang and Li, 2011). Induction of RNAi targeting, individually, TbMCM2, TbMCM3, TbMCM5, TbMCM7, TbSLD5, TbPSF2, TbPSF3 and TbCDC45 in PCF cells resulted in a growth defect, accompanied by an increase in number of enucleated cells (0N1K, or zoids) and a reduction in replicating cells (Dang and Li, 2011), similarly to what has been observed for TbORC1/CDC6 (Benmerzouga *et al.*, 2013), and thus supporting their role in DNA replication in *T. brucei*. Despite a reduction in ~90% of the targeted mRNA, no growth defects were observed for TbMCM4, TbMCM6 or TbSLD5 (Dang and Li, 2011). Like TbORC1/CDC6 (Godoy *et al.*, 2009), all members of the TbMCM2-7 and TbGINS complexes localise to the nucleus of PCF cells throughout the cell cycle (Dang and Li, 2011), in contrast to what has been observed in other organisms such as yeast, where MCM2-7 export to the cytoplasm after DNA replication acts as a mechanism to impede re-replication of already duplicated DNA (Nguyen *et al.*, 2000; Braun and Breeden, 2007). TbCDC45, the remaining member of the CMG complex, however, shows a dynamic localisation: it is nuclear from G1 to G2 phases (with an expression peak at S phase), but is then excluded from the nucleus upon mitosis, apparently through the CRM1/exportin-1 mediated pathway (Dang and Li, 2011). This is not observed in yeast, where Cdc45 localises to the nucleus throughout the cell cycle (Hopwood and Dalton, 1996). It is then possible that exclusion of TbCDC45 from the nucleus might be used as a regulatory mechanism in *T. brucei* to prevent the re-assembly of the CMG complex, and thus avoid activation of assembled pre-RCs at the origins, and therefore, re-replication.

In addition to the putative ORC-like and CMC complexes, another replication factors, involved in the early stages of DNA replication, have also been explored experimentally in *T. brucei*. These include PCNA. In eukaryotes, PCNA is typically the DNA sliding clamp responsible for tethering the replicative DNA polymerases to DNA, although it is also involved in multiple cellular processes, including DNA repair (reviewed in Moldovan *et al.*, 2007; De Biasio and Blanco, 2013). In *T. brucei* PCF cells, analysis of TbPCNA revealed that its expression is cell cycle-regulated, localising in distinct foci throughout the nucleus of the cells mainly during late G1 and S phase cells, while it is virtually absent from G2

phase or post-mitotic cells (Kaufmann *et al.*, 2012). Perhaps surprisingly, this phenotype is rather different from the one observed for PCNA orthologues in *Leishmania donovani* (LdPCNA) (Kumar *et al.*, 2009; Minocha *et al.*, 2011), and *Trypanosoma cruzi* (TcPCNA) (Calderano *et al.*, 2011b): both LdPCNA and TcPCNA localise to the nucleus throughout the cell cycle (though with different sub-nuclear dynamics). It is therefore possible that some aspects of DNA replication control and regulation might differ between *T. brucei*, *T. cruzi* and *Leishmania*. Another replication-associated factor already analysed in *T. brucei* is the orthologue of the MCM-binding protein (MCM-BP), which in model eukaryotes has been shown to form alternative replication complexes with the MCM3-7 subunits in a cell cycle-dependent fashion (Sakwe *et al.*, 2007; Li *et al.*, 2011a; Nguyen *et al.*, 2012). Moreover, some evidence appears to indicate that MCM-BP might also be acting as an unloader of the replicative helicase from the chromatin at the end of the S phase (Nishiyama *et al.*, 2011). Like in other organisms, the *T. brucei* orthologue of MCM-BP (TbMCM-BP) was shown to interact with TbMCM4-7 and, conversely to other eukaryotes, with TbMCM8 as well (Kim *et al.*, 2013). Using a conditional knock-out (KO) system, it was shown that TbMCM-BP is essential for BSF cells viability, resulting in the de-repression of silent VSGs, and cell cycle arrest in G2 phase with the consequent generation of zoids, a phenotype that has never been reported in BSF cells before (Kim *et al.*, 2013).

1.4.1.3 Origins of replication – not as many as expected

Like most eukaryotes, to date no specific sequence elements defining origins of replication have been found in *T. brucei* (Tiengwe *et al.*, 2012a). Moreover, TbORC1/CDC6-binding sites, and thus potential origins of replication, surpass the number of detectable active origins, which appear to be activated during S phase following a defined replication timing programme (Tiengwe *et al.*, 2012a), suggesting that as far as origins of replication are concerned, *T. brucei* appears to possess what has been embraced as universal eukaryotic DNA replication features (reviewed in Masai *et al.*, 2010; O'Donnell *et al.*, 2013).

TbORC1/CDC6-binding sites were mapped across the 11-megabase chromosomes of *T. brucei* PCF cells by chromatin immunoprecipitation coupled with microarray hybridisation (ChIP-chip) (Tiengwe *et al.*, 2012a). This revealed ~953

TbORC1/CDC6-binding sites, and thus potential origins of replication, which displayed sparse binding in the core regions of the 11-megabase chromosomes, and dense binding at the subtelomeric regions. Within the chromosome core (~362 binding sites), TbORC1/CDC6 accumulated at the SSR boundaries of the polycistronic transcription units: 38% at dSSRs, 54% at h-t SSRs, and 7% at cSSRs, with an average spacing between TbORC1/CDC6-binding sites of ~130 Kbp (Tiengwe *et al.*, 2012a). As mentioned in section 1.2.3, while cSSRs are assumed to be transcription termination sites, the dSSRs and h-tail SSRs are transcription start sites, marked by histone H4K10ac accumulation (Siegel *et al.*, 2009). TbORC1/CDC6-binding sites localised just upstream of the H4K10ac-binding sites (Tiengwe *et al.*, 2012a). Analysis of the TbORC1/CDC6-binding sites did not retrieve any common sequence element, and it is presently not known if any specific chromatin or epigenetic markers are involved in TbORC1/CDC6 binding to these specific genomic sites.

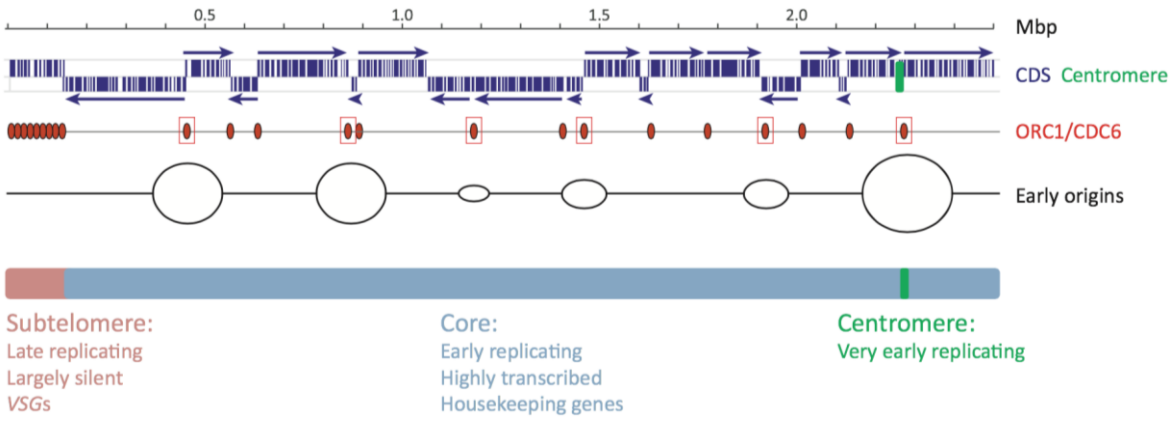


Figure 1.20. TbORC1/CDC6-binging sites and active origins of replication.
Schematic representation of *T. brucei* chromosome 8 (size in Mbp), with the different directional gene clusters shown in dark blue, and their transcription direction highlighted by dark blue arrows. TbORC1/CDC6 binding sites are represented in red, and the activated origins as open circles, where the size of the circles aims to reflect the different timing of origin activation – larger circles are earlier firing origins than smaller circles. Below is a representation of the chromosome regions: subtelomere and core. The centromere is highlighted in green. Reproduced from (Tiengwe *et al.*, 2013),

In order to map the origins of replication that are activated during early-mid S phase at the population level, marker frequency analysis (MFA; detailed in Chapter 4) coupled with deep sequencing (MFA-seq) was used (Tiengwe *et al.*, 2012a). From this analysis, 42 MFA-seq peaks (activated origins) were identified, all within the core region of the chromosomes. The number of origins per chromosome correlated with chromosome size, and the sites of the origins co-localised with TbORC1/CDC6-binding sites (~4.4% of the TbORC1/CDC6-binding

sites), thus confirming TbORC1/CDC6 as an initiator factor (or a member of an initiator complex) (Tiengwe *et al.*, 2012a). However, not all identified peaks had the same height, suggesting that the various origins are activated at different times during S phase, most likely following a replication timing programme, like in other eukaryotes (reviewed in Rhind and Gilbert, 2013). For instance, the most predominant peak in chromosomes 1 to 8 corresponds to the loci previously mapped as centromeres (Obado *et al.*, 2007), suggesting that these regions are replicated early, reflecting what has been observed in model eukaryotes such as yeast (Kim *et al.*, 2003). Nevertheless, these results suggest a low density of active origins, with one origin per 600 Kbp, which is much more dispersed than in other eukaryotes (estimated to have an origin per 50-150 Kbp) (Yamazaki *et al.*, 2013). Recently, the replication fork rate has been estimated for both *T. brucei* PCF and BSF cells to be of 3.7 Kbp.min^{-1} and 4.4 Kbp.min^{-1}, respectively (Calderano *et al.*, 2015), which is slightly faster than what has been reported in other eukaryotes (2 Kbp.min^{-1}) (reviewed in Mechali, 2010), but is theoretically enough to allow the complete replication of the *T. brucei* genome from the 42 mapped origins.

By overlapping with a subset of the TbORC1/CDC6-binding sites in the chromosomes' core region, the MFA-seq mapped origins overlapped with SSRs, thus suggesting a potential correlation between DNA replication and transcription (Tiengwe *et al.*, 2012a), as seen in other eukaryotes (reviewed in Barlow and Nussenzweig, 2014; Leonard and Mechali, 2013; Helmrich *et al.*, 2013). Indeed, this appears to be the case in *T. brucei*: depletion of TbORC1/CDC6 by RNAi results not only in the impairment of DNA replication (Benmerzouga *et al.*, 2013), but also in a widespread effect on transcription (Tiengwe *et al.*, 2012a), including the de-repression of silent metacyclic VSGs in PCF cells and silent VSGs in BSF cells (Tiengwe *et al.*, 2012a; Benmerzouga *et al.*, 2013). In the case of the silent VSG de-repression, it is possible that this is a result of the depletion of TbORC1/CDC6 densely accumulated at the subtelomeric regions of the chromosomes, which might, similarly to other eukaryotic Orc1 subunits (Sasaki and Gilbert, 2007; Duncker *et al.*, 2009), act in chromatin silencing, and thus be somehow involved in antigenic variation, supported by the report that TbORC1/CDC6 RNAi induction also results in an increase in VSG switching in BSF cells (Benmerzouga *et al.*, 2013).

1.5 Aims and Objectives

As detailed above, not much is known about how DNA replication is initiated in *T. brucei* and related kinetoplastids. However, it is now clear that though the replication fork and DNA synthesis molecular machineries appear to be conserved, the factors involved in the earliest stages of the process appear to be highly divergent. If this is verified, these diverged factors could be consequently used, in the long term run, as potential targets for drug development against *T. brucei*, *T. cruzi* and *L. major*.

The main aim of the work shown here was focused ultimately on testing whether *T. brucei* possesses a diverged ORC-like complex, or whether TbORC1/CDC6 assumes the role of the initiator factor alone, as previously suggested (Godoy *et al.*, 2009). To achieve this, the published TbORC1/CDC6-interacting factors TbORC1B, TbORC4, Tb7980 and Tb3120, as well as Tb1120, which has not been published, were analysed for homology with other eukaryotic factors, tested for a role in DNA replication, their subcellular localisation during the cell cycle analysed, and interactions with TbORC1/CDC6 in PCF cells examined. These results are shown and discussed in Chapter 3.

Due to the morphological and biological differences between PCF and BSF cells, the present work also aimed to investigate the role of the TbORC1/CDC6-interacting factors in BSF cells. Therefore, Chapter 4 explores the involvement of these factors in DNA replication, as well as their subcellular localisation throughout the cell cycle. Furthermore, the origins of replication were also mapped in BSF cells and compared with the ones identified in PCF cells.

Finally, Chapter 5 aimed to investigate origin conservation between two related organisms that share highly syntenic genomes, *T. brucei* and *L. major*, and thus explore which features define an origin of replication in kinetoplastids.

2 Materials and Methods

2.1 Bioinformatics

2.1.1 Gene and protein sequence retrieval

Gene and protein sequences were retrieved from the TriTrypDB online database (http://tritrypdb.org/tritrypdb/) (Aslett *et al.*, 2010), as FASTA format files. All gene IDs shown relate to TriTrypDB version 8.1. Whole genome sequences of *T. brucei* TREU 927, *L. major* Friedlin, and *L. mexicana* U1103 were retrieved from TriTrypDB version 6.0, while the *T. brucei* Lister 427 genome sequence was retrieved from TriTrypDB version 8.0.

2.1.2 Gene and protein sequence analysis and manipulation

Gene and protein sequence analysis, primer design, and plasmid cloning planning, were performed using CLC genomics Workbench, version 7.5.1 (QIAGEN Aarhus A/S), and are described in more detail in other sections when necessary.

2.1.3 Protein homology search and protein sequence alignments

For protein homology searches, protein sequences were analysed using the standard (default settings) protein-protein Basic Local Alignment Search Tool (BLAST) (blastp) (http://blast.ncbi.nlm.nih. gov/Blast.cgi) (Altschul *et al.*, 1990), with the non-redundant protein sequences database (nr) (Li *et al.*, 2013). For protein structure and function prediction, RaptorX (Källberg *et al.*, 2012) and I-TASSER (Zhang, 2008a; Roy *et al.*, 2010) online servers were used. Protein alignments were performed using CLC genomics Workbench, with a gap open cost of 10.0, gap extension cost of 1.0, end gap cost of “as any other”, and the very accurate (slow) option for alignment.

2.1.4 Protein domain and motif search

Protein domain and motif searches were performed using Pfam, version 27.0 (http://pfam.xfam.org/) (Finn *et al.*, 2014), InterPro, version 49.0 (http://www.ebi.ac.uk/interpro/) (Hunter *et al.*, 2012), GenomeNet Motif search tool (http://www.genome.jp/tools/motif/) (Kyoto University Bioinformatics Center), MyHits© Motif Scan tool (http://myhits.isb-sib.ch/)

(Pagni *et al.*, 2007), and ScanProsite online tool (http://prosite.expasy.org/scanprosite/) (Sigrist *et al.*, 2013).

2.2 General Molecular Biology Techniques

2.2.1 Plasmid cloning

2.2.1.1 Primer design

Primers were designed using CLC genomics Workbench, and ordered from Eurofins Genomics (Ebersber, Germany). For amplification of gene fragments to be cloned into vectors for endogenous tagging, gene deletion (labelled as KO plasmids), and RNAi in PCF cells (pLew RNAi plasmids), primers were designed to include the sequences for the restriction enzymes to be used for cloning, and are shown in Table 2-1. Primers designed for use in gateway cloning are described in section, 2.2.1.10.

2.2.1.2 Genomic DNA extraction and purification

Genomic DNA (gDNA) was extracted from approximately 3 x 10^6 PCF or BSF cells using the Blood and Tissue DNA extraction kit (Qiagen), according to the manufacturer's instructions. The gDNA was eluted in 50 µl of Buffer AE, and stored at 4°C until use.

2.2.1.3 Polymerase Chain Reaction (PCR) of products used for cloning

For PCR amplification of fragments to be used for cloning, Phusion® High-Fidelity DNA Polymerase (New England Biolabs), a high fidelity proofreading polymerase was used, according to the manufacturer's instructions. Reactions were set as follows: 10 µl of 5x Buffer HF, 1 µl of 10 mM dNTP mix (Promega), 2.5 µl of forward primer (10 pmol stock), 2.5 µl of reverse primer (10 pmol stock), 0.5 µl of Phusion®, 1 µl of gDNA, and 32.5 µl of MilliQ water, to a final volume of 50 µl per reaction. The basic PCR program was performed as follows: 98°C for 30 sec; 35 cycles of 98°C for 10 sec, 54-64°C for 30 sec (depending on the melting temperature of each pair of primers), and 72°C for 45 sec to 1 min 30 sec (depending on the size of the fragment to be amplified, 1 min per 1 Kb); and a final extension step of 72°C for 5 min. The resulting PCR products were then analysed by agarose gel electrophoresis (section 2.2.1.4).

Table 2-1. Primers used for cloning.
List of primer sequences used for cloning plasmids for endogenous tagging, gene deletions (KO) and pLew RNAi plasmids. 5' UTR and 3'UTR refer to the gene's intergenic regions at the 5' and 3' ends, respectively. The restriction site sequences are underlined and in red.

Number	Gene	Gene ID	Sense	Sequence	Restriction Sites	Purpose
CTOL[c]_65	TbORC4	Tb927.10.13380	Fwd	CCCAAGCTTCGTTTCTGCTGTCTTTGGGG	HindIII	C-terminus tag
CTOL_66			Rev	CCCTCTAGACACGAGGCTGCGTAATC	XbaI	
CTOL_69	Tb7890	Tb927.10.7980	Fwd	CCCAAGCTTCGAGCAAGTATCGTCACACAGGA	HindIII	C-terminus tag
C 103			Fwd	ATGGCGAGAGCTCTCGTCATCGGTCG	SacI	
CTOL_70			Rev	CCCTCTAGATCGTGGAATGAGGTCGT	XbaI	
C 120	Tb3120	Tb927.9.4530	Fwd	CCCAAGCTTAGTGCATGGTATAGACGAAC	HindIII	C-terminus tag
C 121			Rev	CCCTCTAGATGCCTCCACTGGAGCTCCAC	XbaI	
C 3	Tb1120	Tb927.6.1120	Fwd	GCATAAGCTTCTTCTGTTGCTTTCTGCGAG	HindIII	C-terminus tag
C 4			Rev	CCCTCTAGAGCTAAGCGCAAGCAAGGAGG	XbaI	
C 5	TbORC1B	Tb927.9.2030	Fwd	GCATAAGCTTACAACGAGACAGTCAAATCG	HindIII	C-terminus tag
C 6			Rev	GCATTCTAGACAGGGATAAAATGCCCTTGA	XbaI	
C 1	TbMCM10	Tb927.9.1730	Fwd	GACATGGCGCGCCCACATTACACGAGGATTCTG	AscI	C-terminus tag
C 2			Rev	GCATTCTAGACGCATCCTCTCGAAGGCTAA	XbaI	
C 184	Tb7890	Tb927.10.7980	Rev	GCATGGATCCACGACGGGGAAACAGAACG	BamHI	N-terminus tag (5'intergenic region)
C 185			Fwd	GCATGGTACCGACATGCCGTGACGAACTC	KpnI	
C 186			Fwd	GCATACTAGTGCAGCCCAAACACCACGCA	SpeI	N-terminus tag (ORF 5' region)
C 187			Rev	GCATGGTACCCACGACGAAGTGAAGCTCA	KpnI	
C 189	TbORC4	Tb927.10.13380	Fwd	GCATAAGCTTGCGGCCGCTACCCTCCTTTCTTCTAGT	HindIII/NotI	KO (5'UTR region)
C 190			Rev	GCATTCTAGACCACGCCTCGCAACCTCAAGAA	XbaI	
C 191			Fwd	GCATGAGCTCACCAAGGAGGTGGCGCGATT	SacI	KO (3'UTR region)
C 192			Rev	GCATATCGATGCGGCCGCGTTCTTCACATCATTAGCG	ClaI/NotI	
C 193	Tb7890	Tb927.10.7980	Fwd	GCATAAGCTTGCGGCCGCCGCCAGGAATCGTTTTAGTA	HindIII/NotI	KO (5'UTR region)
C 194			Rev	GCATTCTAGAAAGATGGAGATGTAACCGGA	XbaI	
C 195			Fwd	GCATGAGCTCCCACAGCACAAAGGGTTA	SacI	KO (3'UTR region)
C 196			Rev	GCATATCGATGCGGCCGCGACTTCTTCCACACTCAGTT	ClaI/NotI	
C 197	Tb3120	Tb927.9.4530	Fwd	GCATAAGCTTGCGGCCGCTCATCTCTTGGTGGTTCCTC	HindIII/NotI	KO (5'UTR region)
C 198			Rev	GCATTCTAGACATAACACCGAGCGCAAA	XbaI	

[c] Primers labelled as CTOL and CT were designed by Dr Calvin Tiengwe during the course of his PhD, and most can be found in Tiengwe et al., 2012b, as well as in his thesis Tiengwe, 2010.

C 199			Fwd	GCATGGAGCTCCAGTGGAGGCATAGGTTGA	SacI	KO (3'UTR region)
C 200			Rev	GCATATCGATGCGGCCGCCTCTCACCAGAAATGTACCC	ClaI/NotI	
C 205	Tb1120	Tb927.6.1120	Fwd	GCATAAGCTTGCGGCCGCAGTTGACTGTATGAGAGCG	HindIII/NotI	KO (5'UTR region)
C 206			Rev	GCATTCTAGATAGTGCATTGATTCCTTCGG	XbaI	
C 207			Fwd	GCATGGTACCAGAGTTTAAAGCGAGGTCC	KpnI	KO (3'UTR region)
C 208			Rev	GCATATCGATGCGGCCGCGTGCAGATGGGATTGAGGA	ClaI/NotI	
C 201	TbORC1B	Tb927.9.2030	Fwd	GCATAAGCTTGCGGCCGCGCACAACTCATTCGGTAAAA	HindIII/NotI	KO (5'UTR region)
C 202			Rev	GCATTCTAGACGTTATATTCTTGTTCCCGT	XbaI	
C 203			Fwd	GCATGAGCTCGGGAACTCAAGGGCATTTTA	SacI	KO (3'UTR region)
C 204			Rev	GCATATCGATGCGGCCGCAGTTTCTTTCAGTCTGCTCA	ClaI/NotI	
CT 1	TbORC1/ CDC6	Tb927.11.7216	Fwd	CCCCAAGCTTGAAGCCCACAGCTGTCTTTC	HindIII	pLew- RNAi
CT 2			Rev	CCCCCTCGAGTTCTCCGGCAACTTGTAACC	XhoI	
CT 3			Fwd	CCCCGGATCCGAAGCCCACAGCTGTCTTTC	BamHI	
CT 4			Rev	CCCCCTTAAGTTCTCCGGCAACTTGTAACC	AflII	
CT 5	TbORC4	Tb927.10.13380	Fwd	CCCCAAGCTTCACGTTGTATCCCCTTGCTT	HindIII	pLew- RNAi
CT 6			Rev	CCCCCTCGAGTTCAGTTTCGGCGAAGTTCT	XhoI	
CT 7			Fwd	CCCCGGATCCCACGTTGTATCCCCTTGCTT	BamHI	
CT 8			Rev	CCCCCTTAAGTTCAGTTTCGGCGAAGTTCT	AflII	
C 65	Tb7890	Tb927.10.7980	Fwd	GCATAAGCTTCTTCCCATGCGAGCAAGTAT	HindIII	pLew- RNAi
C 66			Rev	GCATCTCGAGTTCAGCAGGGACCGATAAAC	XhoI	
C 69			Fwd	GCATGGATCCCTTCCCATGCGAGCAAGTAT	BamHI	
C 70			Rev	GCATCTTAAGTTCAGCAGGGACCGATAAAC	AflII	
CT 9	Tb3120	Tb927.9.4530	Fwd	CCCCAAGCTTCTAACGGCTCAGTTTCTCGG	HindIII	pLew- RNAi
CT 10			Rev	CCCCCTCGAGTTGGCAAAAGATTCCTCACC	XhoI	
CT 11			Fwd	CCCCGGATCCCTAACGGCTCAGTTTCTCGG	BamHI	
CT 12			Rev	CCCCCTTAAGTTGGCAAAAGATTCCTCACC	AflII	
C 104	Tb1120	Tb927.6.1120	Fwd	CCCAAGCTTGGAGGATGGGAAAGAAGGAG	HindIII	pLew- RNAi
C 105			Rev	CCCCTCGAGCCGGGAGCAGAAGTACAGAG	XhoI	
C 106			Fwd	CCCGGATCCGGAGGATGGGAAAGAAGGAG	BamHI	
C 107			Rev	CCCCTTAAGCCGGGAGCAGAAGTACAGAG	AflII	
CT 13	TbORC1B	Tb927.9.2030	Fwd	CCCCAAGCTTCTATCGGCTGAGTACGCCTC	HindIII	pLew- RNAi
CT 14			Rev	CCCCCTCGAGTTTGCGATTTGACTGTCTCG	XhoI	
CT 15			Fwd	CCCCGGATCCCTATCGGCTGAGTACGCCTC	BamHI	
CT 16			Rev	CCCCCTTAAGTTTGCGATTTGACTGTCTCG	AflII	

2.2.1.4 Agarose Gel Electrophoresis

For PCR product analysis, 1% (w/v) UltraPure™ Agarose (Life Technologies) gels, diluted in 1x TAE buffer (40 mM Tris base, 19 mM acetic acid, 1mM EDTA) were prepared. To allow DNA detection by ultraviolet light, SYBR® Safe DNA Gel Stain (Life Technologies) was diluted 1:20000 into the warm melted agarose. To the PCR products, 6x type III DNA Gel-loading Buffer (Sambrook and Russell, 2001) was added to a final concentration of 1x, before these were loaded onto the gel, alongside the 1 Kb Plus ladder (Invitrogen™, Life Technologies). The gels were ran at 40-100 v in 1x TAE buffer, on Mini-Sub® Cell GT Cell tanks (Bio Rad). For visualisation, gels were imaged using a Gel Doc™ XR+ System (Bio Rad), equipped with the Quantity One 1-D v 4.6 Analysis Software (Bio Rad).

2.2.1.5 PCR products purification

PCR products were purified using one of two methods: either directly from the PCR reaction, using the QIAquick PCR purification kit (Qiagen), or from an agarose fragment containing the PCR product, using the QIAquick Gel purification kit (Qiagen), both according to the manufacturer's instructions. In the latter case, the gels were imaged under a Dark Reader blue Transilluminator (Clare Chemical Research), and the appropriate bands excised using a sterile scalpel blade.

2.2.1.6 Restriction Enzyme digestion

With the exception of cloning performed using the gateway system (below), all PCR products used for cloning were digested with the corresponding restriction enzymes (Table 2-1). All enzymes were from New England Biolabs (NEB), and used according to the manufacturer's instructions, with consultation of the Double Digest Finder application (https://www.neb.com/tools-and-resources/interactive-tools/double-digest-finder) for the planning of double enzyme digestion reactions. Reactions were set up as follows: 5 µl of the appropriate 10x buffer, 1 µl of each of the two restriction enzymes, 20 µl of the PCR product, and 23 µl of MilliQ water. The reactions were incubated for 1 h at the temperature recommended by NEB. The plasmids used as backbone vectors for the PCR product cloning were likewise digested using the same pairs of enzymes, as mentioned above. Both digested PCR products and plasmid vectors

were then separated on 1% (w/v) agarose gels (section 2.2.1.4), and purified with a QIAquick Gel purification kit (Qiagen), as aforementioned. The concentration of the purified products was then assessed by a NanoDrop 1000™ Spectrophotometer (Thermo Scientific), before ligation (below).

2.2.1.7 Ligation

Ligation of digested PCR products and plasmid vector was performed using the T4 DNA Ligase (NEB), according to the manufacturer's instructions. In order to assess the amount of PCR product to use, Equation 1 was used, considering 50 ng of plasmid vector. Two ligation reactions were prepared in parallel, using a 1:5 and 1:10 plasmid vector:PCR product ratios, each with a total volume of 20 µl, containing 1x Ligation Buffer and 1 µl of T4 DNA Ligase (NEB). The reactions were incubated at room temperature (21-25°C) for 30 min, and then transformed into competent bacteria (below).

Equation 1. Ligation components formula.

$$\frac{Plasmid\ weight\ (\mu g)\ \times\ PCR\ product\ size\ (bp)}{Plasmid\ size\ (bp)} = PCR\ product\ weight\ (\mu g)$$

2.2.1.8 Transformation of competent bacteria

Chemically competent *Escherichia coli* strain DH5α bacteria (prepared by Mr Craig Lapsley) were used. Aliquots stored at - 80°C were allowed to thaw on ice for 30 min. Next, 50 µl of bacteria were added to 10 µl of the ligation product, and left on ice for 30 min. Then, the cells were incubated at 42°C for 30 sec, and immediately placed on ice for 2 min. Next, 250 µl of liquid lysogeny broth (LB) medium (10 $g.L^{-1}$ of Tryptone, 5 $g.L^{-1}$ of yeast extract, and 10 $g.L^{-1}$ of NaCl) were added to the cells, and these incubated for 1 h at 37°C, with agitation. Around 100-200 µl of cells were then plated on LB agar (10 $g.L^{-1}$ of Tryptone, 5 $g.L^{-1}$ of yeast extract, 10 $g.L^{-1}$ of NaCl, and 15 10 $g.L^{-1}$ of agarose) supplemented with 100 $\mu g.ml^{-1}$ of Ampicillin (Sigma Aldrich), and incubated overnight at 37°C. Plates were inspected for colonies, which were then transferred onto 5 ml of liquid LB supplemented with 100 $\mu g.ml^{-1}$ of Ampicillin, and incubated overnight at 37°C, with agitation. The overnight culture was then centrifuged for 3 min at 6000 g, and used for plasmid purification (below).

2.2.1.9 Plasmid purification, and further confirmation by enzymatic digestion and sequencing

In order to purify the plasmids from the pellets of overnight grown liquid cultures of transformed *E. coli* DH5α bacteria, the QIAprep Spin Miniprep Kit (Qiagen) was used, according to the manufacturer's instructions. The concentration of the plasmids was then assessed by photospectrometry using a NanoDrop 1000™ Spectrophotometer (Thermo Scientific). Approximately 5 µl of plasmid was then digested with the appropriate restriction enzymes for confirmation (section 2.2.1.6), and separated on a 1% (w/v) agarose gel, as abovementioned (section 2.2.1.4). If correct, according to the sizes of the resulting fragments, the plasmids were then sent for sequencing either at the DNA Sequencing & Services™ University of Dundee, UK (www.dnaseq.co.uk), or Eurofins Genomics (Ebersber, Germany). Primers used are shown in Table 2-2. The resulting chromatogram files containing the sequence peak data (.ab1) were then analysed using CLC genomics Workbench, version 7.5.1 (QIAGEN Aarhus A/S), by assembling the sequence peak data to the reference file (the plasmid map) using the "Assemble Sequences to Reference" function.

Table 2-2. Primers used for sequencing of plasmids, for confirmation of the inserts.
(*) Primers provided by Eurofins sequencing services. (MP) primers designed by Dr Marko Prorocic; (AT) primers designed by Dr Anna Trenaman. The numbers refer to their individual primer lists.

Number	Name	Sequence	Plasmids seq.
MP 15	HA/myc Fwd	CGTTGGCCGATTCATTAATGC	C-terminal 12-myc and 6-HA tag
MP 16	HA/myc Rev	TAATGACGAACGGGAAATGC	C-terminal 12-myc and 6-HA tag
*	M13 uni (-43)	AGGGTTTTCCCAGTCACGACGTT	KO plasmids
*	M13 rev (-49)	GAGCGGATAACAATTTCACACAGG	KO plasmids
AT 120	N-tag Fwd	GTATACCAACAAGCCCGAAAAC	N-terminal 12-myc tag
AT 121	N-tag Rev	CCTTTCCACGGAAAAGACAC	N-terminal 12-myc tag

2.2.1.10 The specific case of Gateway® Cloning

Constructs designed for RNAi in bloodstream form cells (BSF) were generated using a derivation of the gateway® cloning technology (described in Jones *et al.*, 2014). Gene specific primers were designed to include the *att*B sites (Table 2-4), and the PCRs were performed using Phusion® High-Fidelity DNA Polymerase (NEB), as described above (section 2.2.1.3). The PCR products were separated on 1% (w/v) agarose gels and the bands extracted as described in previous sections (2.2.1.4 and 2.2.1.5). The BP reaction was then performed using the

Gateway® BP Clonase® II Enzyme mix kit (Life Technologies), as follows: 5 µl of purified PCR product, 1 µl of pGL2084 vector (Jones *et al.*, 2014), 3.75 µl of TE buffer, and 0.25 µl of BP Clonase® II, incubated for 1 h at room temperature (20-25°C). Next, 1 µl of Proteinase K solution (2 $\mu g.\mu l^{-1}$) was added to the mix, and incubated at 37°C for 1 h. The 10 µl of reaction was then added to 50 µl of *E. coli* MAX Efficiency® DH5α™ competent cells (Life Technologies), which were transformed as described in the previous section 2.2.1.8, using S.O.C medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose), instead of LB medium, for the bacteria recovery step. The plasmids resultant from the bacteria transfections were purified as described before (section 2.2.1.9), and digested with the following restriction enzymes: BamHI and XbaI to excise both inserts; StuI to excise the sense insert; ClaI to excise the anti-sense insert; and AscI to test the linearization sites. Reactions were prepared in a total of 20 µl as follows: 2 µl of Buffer 4 with BSA or CutSmart™ buffer (NEB), 0.25 µl of restriction enzyme (NEB), 3 µl of plasmid, and 14.75 µl of MilliQ water; and incubated for 1 h at 37°C. The resulting digested products were then separated on an agarose gel and analyse as described previously.

2.2.1.11 Preparation of plasmids for procyclic form and bloodstream form cells transfection

Prior to transfection into either PCF or BSF cells, 10 µg of plasmid were digested with the appropriate restriction enzyme overnight (as described above; the restriction enzyme used is mentioned in Table 2-3), for linearization. The digested sample was then purified by ethanol precipitation, as follows: the digestion reaction was centrifuged at maximum speed for 20 min at 4°C, and the supernatant discarded; 100 µl of 70% Ethanol were added to the pellet, and centrifuged at maximum speed for 2 min at room temperature; the supernatant was then removed, and the pellet was allowed to dry; the pellet was then re-suspended in 11 µl of MilliQ water. From this, 1 µl was separated, alongside the un-digested sample, on a 1% (w/v) agarose gel, as mentioned above, for confirmation. If correct, the remaining 10 µl were used for transfection as described in sections 2.3.4 and 2.3.5.

Table 2-3. Restriction sites used for plasmid linearization prior to transfection into PCF or BSF *T. brucei* cells.

Plasmid	Linearization restriction site (s)
TbORC1/CDC6-12mycBSD	XhoI
TbORC1B-12mycBSD	ClaI
TbORC4-12mycBSD	SmaI
Tb3120-12mycBSD	NsiI
Tb1120-12mycBSD	NsiI
TbORC1B-6HABLE	ClaI
Tb1120-6HABLE	NsiI
pLew RNAi (all plasmids)	NotI
pGL2084 RNAi (all plasmids)	AscI

2.2.2 Routine PCR for cell line confirmation

For routine PCR, Taq DNA Polymerase with ThermoPol® Buffer (NEB) was used. Primers were designed to confirm integration of the different plasmids, either KO or endogenous tag, into the *T. brucei* genome (Table 2-5). In more detail, reactions were set as follows: 2 µl of 10x ThermoPol® Buffer, 0.5 µl of 10 mM dNTP mix (Promega), 0.5 µl of forward primer (10 pmol stock), 0.5 µl of reverse primer (10 pmol stock), 0.125 µl of Taq DNA Polymerase, 4 µl of gDNA, and 12.4 µl of MilliQ water, to a final volume of 20 µl per reaction. The basic PCR program was performed as follows: 95˚C for 5 min; 35 cycles of 95˚C for 1 min, 53-57˚C for 30 sec (depending on the melting temperature of each pair of primers), and 68˚C for 1 min to 2 min and 30 sec (depending on the size of the fragment to be amplified, 1 min per 1 Kb); and final extension step of 68˚C for 10 min. The resulting PCR products were then analysed by agarose gel electrophoresis as in section 2.2.1.4.

Table 2-4. Primers used for Gateway cloning.
The *att*B sites are highlighted in red.

Number	Gene	Gene ID	Sense	Sequence
C 108	TbORC1/CDC6	Tb927.11.7216	Fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTGAAGCCCACAGCTGTCTTTC
C 109			Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCTCCGGCAACTTGTAACC
C 114	TbORC4	Tb927.10.13380	Fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTCACGTTGTATCCCCTTGCTT
C 115			Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTCAGTTTCGCCGAAGTTCT
C 116	Tb7890	Tb927.10.7980	Fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTAGTGTGGCTCCGGTTACATC
C 117			Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTGTTGCAAAGAGCGTGTTC
C 110	Tb3120	Tb927.9.4530	Fwd	GGGGACCACTTTGTACAAGAAAGCTGGGTTTGGCAAAAGATTCCTCACC
C 111			Rev	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTAACGGCTCAGTTTCTCGG
C 112	TbORC1B	Tb927.9.2030	Fwd	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTGCGATTTGACTGTCTCG
C 113			Rev	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTATCGGCTGAGTACGCCTC

Table 2-5. Primers used for cell line confirmation

Number	Gene	Gene ID	Sense	Sequence	Purpose
C 40	Blasticidin	-	Rev	AGAGATGGGGATGCTGTTGA	Endogenous tag integration
C 41	Bleomycin	-	Rev	AACGGCACTGGTCAACTT	
C 101	TbORC1/CDC6	Tb927.11.7216	Fwd	AAGACAGCGTCAGTGAAC	Endogenous tag integration/Wt allele
C 132	TbORC4	Tb927.10.13380	Fwd	AACGGTTACCTAACTCCAC	
C 218	Tb7890	Tb927.10.7980	Fwd	GTATACCAACAAGCCCGAAA	
C 220	Tb7890	Tb927.10.7980	Fwd	CTTATCGCTCACTGGTACTG	
C 135	Tb3120	Tb927.9.4530	Fwd	TACGGTTTCATCTCCAGACAT	
C 137	Tb1120	Tb927.6.1120	Fwd	TGTTGATAGTGGTGGTGGTG	
C 136	TbORC1B	Tb927.9.2030	Fwd	TTGAGTTTTCGGGTGATGTG	
C 215	Neomycin	-	Rev	CAGTCATAGCCGAATAGCC	KO construct integration
C 214	Puromycin	-	Rev	GCGTGAGGAAGAGTTCTTG	
C 213	TbORC1/CDC6	Tb927.11.7216	Fwd	TAGTGGTTTTAAGGGCGTAGG	
C 212	TbORC4	Tb927.10.13380	Fwd	TTTTTGGTGGTGCGTTGTG	
C 209	Tb7890	Tb927.10.7980	Fwd	TATCGCTCACTGGTACTGC	
C 210	Tb3120	Tb927.9.4530	Fwd	AGGTGGAGGAGAAAGAGGA	
C 221	Tb1120	Tb927.6.1120	Fwd	TGTGCAGGTGTGGGAATAAA	
C 211	TbORC1B	Tb927.9.2030	Fwd	GCCTGTTCTGTTCTCAATCT	

CTOL_12	TbORC1/CDC6	Tb927.11.7216	Rev	CCCCTCGAGCCTTCTTTCGGCTTTGGCTT	Wt allele
C 134	TbORC4	Tb927.10.13380	Rev	TCGGCGTTGGAACTAGAT	
C 219	Tb7890	Tb927.10.7980	Rev	TTAATACCCTCAGCACCCTC	
C 130	Tb3120	Tb927.9.4530	Rev	CGTGTGACGCTCCATATGAC	
C 138	Tb1120	Tb927.6.1120	Rev	CGCCTCTCCTTCCTCTATG	
C 131	TbORC1B	Tb927.9.2030	Rev	CTTCACCTTCACACAACGG	

2.2.3 RNA extraction and purification

A minimum of 2 x 10^7 of PCF cells was collected, centrifuged at 1620 g for 10 min, washed once in 1x phosphate-buffered saline (PBS) buffer solution, and the pellets stored at - 80°C. Total RNA was extracted using the RNeasy Mini Kit (Qiagen), following the instructions on the provided protocol "Purification of Total RNA from Animal Cells Using Spin Technology". An extra step of on-column DNase I (RNase-Free DNase Set, Qiagen) digestion was added after step 5 of the protocol. The concentration of the purified RNA samples was then assessed by photospectrometry using a NanoDrop 1000™ Spectrophotometer (Thermo Scientific), and while a portion of each sample was immediately converted into complementary DNA (cDNA) (below), the remainder RNA was stored at - 80°C.

2.2.4 cDNA preparation

For each sample, 1 µg of total RNA was converted into complementary DNA (cDNA) using the SuperScript™ III Reverse Transcriptase Kit (Invitrogen™, Life Technologies), according to the manufacturer's instructions regarding the "First Strand cDNA Synthesis" protocol. Random primers were used, and the final step of RNA removal was not performed. The resultant cDNA was stored at - 20°C, until analysis.

2.2.5 Protein Analysis

2.2.5.1 Whole cell extract sample preparation

For routine confirmation of the cell lines, approximately 2.5 x 10^6 cells were centrifuged for 10 min at 1620 g (PCFs) or 1000 g (BSFs), and re-suspended in 10 µl of 1x NuPAGE® LDS Sample Buffer (Life Technologies) diluted in 1x PBS. These were either immediately analysed (below) or stored at - 20°C.

2.2.5.2 Sodium-dodecyl-sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Samples were incubated for 5 min at 100°C, and loaded onto protein gels, either 10- or 12-well NuPAGE® Novex® 10% Bis-Tris Protein Gels 1.0 mm (Life Technologies). The Novex® Sharp Protein Standard was used as protein size marker. The gels were ran at 200 v for 50 min in a XCell SureLock™ Mini-Cell

Electrophoresis System (Life Technologies), and then were either transferred to a nitrocellulose membrane (Western Blot, section 2.2.5.3), or stained with gel staining dyes (sections 2.2.5.4), for further protein analysis.

2.2.5.3 Western Blot: gel transfer and immunodetection by chemiluminescence

After electrophoresis, the protein gel was transferred to a Hybond ECL Nitrocellulose blotting membrane (GE Helthcare Life Sciences) for 2 h at 100 v, in 1x transfer buffer (25 mM Tris pH 8.3, 192 mM Glycine, and 20% (v/v) methanol), using a Mini Trans-Blot Electrophoretic Transfer Cell (Bio Rad). The membrane was then washed 5 min in 1x PBS (pH 7.2) with 0.01% of Tween-20 (Sigma Aldrich), and incubated for 5 min with Ponceau S solution (Sigma Aldrich), to confirm the transfer of the proteins from the gel onto the membrane. The membrane was then washed 5 min with 1x PBS with 0.01% of Tween-20, and left overnight at 4°C, with agitation, in primary antiserum (Table 2-6) diluted in blocking solution (5% milk (v/v) in 1x PBS with 0.01% of Tween-20). The membrane was then washed three times, 15 min each, with 1x PBS with 0.01% of Tween-20, and incubated for 2 h, with agitation, with the secondary antiserum conjugated with horse radish peroxidase (HRP) (Table 2-6), diluted in blocking solution. Next, the membrane was washed three times, 15 min each, with 1x PBS with 0.01% of Tween-20, incubated for 4 min with SuperSignal West Pico Chemi-luminescent Substrate (Thermo Scientific) and exposed onto X-ray film (Kodak).

If more than one protein was to be detected per membrane, after detection of the first signal, the membrane was washed for 5 min in 1x PBS with 0.01% of Tween-20, followed by 30 min with 5 ml of Restore Western Blot Stripping Buffer (Thermo Scientific). The membrane was then washed once for 15 min with 1x PBS with 0.01% of Tween-20, before incubating it with the primary antiserum. The rest of the process was performed as aforementioned.

Table 2-6. Antisera and concentrations used for Western Blot

Name	Host	Target	Clone	Concentration	Manufacturer
α-myc	mouse	myc	4A6	1:7000	Millipore
α-HA	mouse	HA	HA-7	1:10000	Sigma Aldrich
α-Oligopeptidase B (OPB) form *L. major*	sheep	OPB	-	1:1000	(Munday *et al.*, 2011)
α-NOG1 from *T. brucei*	rabbit	NOG1	-	1:5000	(Jensen *et al.*, 2003)
α-EF-1α (elongation factor-1 alpha)	mouse	EF-1α	CBP-KK1	1:25000	Millipore
α-H2A (histone H2A) from *T. brucei*	rabbit	H2A	-	1:1000	Jeremy Mottram Lab
Goat α-mouse IgG (H+L) HRP[§] conjugate	goat	mouse	-	1:5000	Novex®
Goat α-rabbit IgG (H+L) HRP[§] conjugate	goat	rabbit	-	1:5000	Novex®
Donkey α-sheep IgG HRP[§] conjugate	donkey	sheep	-	1:5000	Santa Cruz

[§]HRP Horse Radish Peroxidase

2.2.5.4 SYPRO® Ruby staining of SDS-PAGE gels

After electrophoresis, the gel was transferred to a polypropylene box, and fixed in 100 ml of 50% (v/v) Methanol and 7% (v/v) Acetic Acid in MilliQ water, with agitation, for 30 min, twice. Later, 60 ml of SYPRO® Ruby gel stain (Molecular Probes™, Life Technologies) were added to the gel, and left overnight with agitation. The gel was then transferred to a new polypropylene container, and washed in 100 ml of 10% (v/v) Methanol and 7% (v/v) Acetic Acid in MilliQ water for 30 min with agitation, and then washed twice, for 5 min each, with MilliQ water. Next, the gel was placed onto a 3 mm glass plate, and scanned using a Typhoon 8600 Variable Mode Imager (Amersham Biosciences). The 457 mm Blue I and Blue II emission filters were used for imaging.

2.3 Trypanosoma brucei culture in vitro

2.3.1 Strains and cell lines used

For most of the methodologies involving the study of *T. brucei* PCF cells, the strains TREU (Trypanosomiasis Research Edinburgh University) 927 and Lister 427 (derived from MITat – Molteno Institute *T. brucei* antigenic type – 1.2 clone 221a) were used. For the specific purpose of RNA interference (RNAi) assays, the PCF Lister 427 pLew29-pLew13 cell line, here referred to as 29-13 for simplicity (Wirtz *et al.*, 1999), was used. For studies involving the bloodstream form cells,

the Lister 427 strain (derived from MITat – Molteno Institute *T. brucei* antigenic type – 1.2 clone 221a) was used. For assays involving RNAi, a genetically modified BSF Lister 427 cell line (named 2T1) (Alsford *et al.*, 2005) was used.

2.3.2 *In vitro* culture of procyclic cell forms

All used cell lines of *T. brucei* PCF cells were grown in non-vented flasks at 27°C in semi-defined medium, SDM-79 (Brun and Schonenberger, 1979) (Gibco®), supplemented with 10% (v/v) heat inactivated foetal bovine serum (FBS) (Sigma Aldrich for routine culture, or Gibco®, in the case of the tetracycline free FBS for RNAi assays), 5 μg.ml^{-1} of Haemin (Sigma Aldrich), and 1% (v/v) of 10.000 U.ml^{-1} penicillin-streptomycin solution (Gibco®). In the case of genetically modified cell lines, the respective selective drugs (InvivoGen) were added at the appropriated concentrations (Table 2-7). For cell culture maintenance, cells were seeded at a minimum of 5 x 10^5 cells.ml^{-1}, and used for experiments at concentrations ~1 x 10^7 cells.ml^{-1} (mid-log phase). Cell concentration was assessed using a Neubauer improved hemocytometer (Marienfeld-Superior, Germany), as standard.

Table 2-7. Antibiotics concentrations used for transfection and cell line maintenance, in both PCFs and BSFs.

	Concentration	
Name	Procyclic Cell Forms	Bloodstream Forms
Hygromycin	50 μg.ml^{-1}	5 μg.ml^{-1}
Neomycin (G418)	10 μg.ml^{-1}	2.5 μg.ml^{-1}
Zeocin	10 μg.ml^{-1}	-
Blasticidin	10 μg.ml^{-1}	10 μg.ml^{-1}
Phleomycin	2.5 μg.ml^{-1}	2.5 μg.ml^{-1}
Puromycin	1 μg.ml^{-1}	0.2 μg.ml^{-1}

In one occasion the doubling time of a cell line was calculated using the following mathematical formulae:

Equation 2. Doubling time formula.

$$Cell\ doubling\ per\ hour\ (K) = \frac{(Log10\ of\ the\ cell\ concentration\ at\ 48\ h) - (Log10\ of\ the\ cell\ concentration\ at\ 0\ h)}{(0.301\ \times\ 48\ hours)}$$

$$Doubling\ time\ (g) = \frac{1}{K}$$

2.3.3 *In vitro* culture of bloodstream forms

T. brucei BSF cells Lister 427 and derivative genetically modified cell lines were grown in vented flasks at 37°C and 5% CO_2, in HMI-9 (Hirumi and Hirumi, 1989) (Gibco®), supplemented with 20% (v/v) heat inactivated FBS (Sigma Aldrich) and 1% of penicillin-streptomycin solution (stock at 10.000 U.ml^{-1}) (Gibco®). 2T1 cell lines were grown in HMI-11, which differs from HMI-9 only in that it is supplemented with 10% (v/v) of FBS (Gibco®, tetracycline free). For RNAi assays which required the incorporation of the thymidine analogue 5-ethynyl-2'-deoxyuridine (EdU) (section 2.5.6), cells were cultured in HMI-11 thymidine-free media (Dr Gloria Rudenko laboratory, personal communication), consisting in Iscove's Modified Dulbecco's Medium (IMDM) (Gibco®), 10% (v/v) of FBS (Gibco®, tetracycline free), 1% of 10.000 U.ml^{-1} penicillin-streptomycin solution (Gibco®), 4% (v/v) of HMI-9 mix (0.05 mM of bathocuproine disulphonic acid, 1 mM of sodium pyruvate, and 1.5 mM of L-cysteine) (Sigma Aldrich), 1mM of hypoxanthine (Sigma Aldrich) and 0.0014% of 2-mercaptoethanol (Sigma Aldrich). In the case of genetically modified cell lines, the respective selective drugs (InvivoGen) were added at the appropriated concentrations (Table 2-7). For culture maintenance, cells were usually seeded at 1 x 10^4 cells.ml^{-1}, and used for experiments at concentrations ~1 x 10^6 cells.ml^{-1} (mid-log phase). Cell concentration was assessed using a Neubauer improved hemocytometer, as standard.

2.3.4 Stable transfection of procyclic cell forms

Approximately 2 x 10^7 PCF cells from a culture at ~1 x 10^7 cells.ml^{-1} were collected and centrifuged for 10 min at 900 g, room temperature. The supernatant ("used SDM-79 media") was filter-sterilised and saved at 4°C. The cell pellet was then re-suspended in 500 µl of ZMnoG buffer (132 mM of NaCl, 8 mM of KCl, 8 mM of Na_2HPO_4, 1.5 mM of KH_2PO_4, and 90 µM of $CaAc_2$), pre-warmed to 27°C and filter-sterilised, and transferred to an 0.4 cm electroporation cuvette (Bio Rad), containing 10 µg of linearised plasmid (prepared as described in section 2.2.1.11). The cells were then electroporated using a Gene Pulsor II electroporator machine (Bio Rad), and subjected to two electric pulses with a voltage of 1.5 Kv and 25 µF of capacitance. The cells were then transferred to 10 ml of complete SDM-79 media, and incubated overnight at

27°C. Conditioned media, comprising 10% (v/v) FBS (Sigma Aldrich), 15% (v/v) of "used SDM-79 media", 75% (v/v) of fresh SDM-79 media, and the appropriate amount of selective drug (Table 2-7), was then prepared, filter-sterilised, and warmed-up to 27°C. From the overnight-grown electroporated culture, 100 µl were taken and added to 40 ml of conditioned media, from which 20 ml were plated onto a 96-well plate, 175 µl per well (high dilution plate). To the remaining 20 ml of conditioned media, 1 ml of the culture was added, and the mixture plated as before (low dilution plate). The plates were then left at 27°C for 10 to 15 days, after which the surviving clones were isolated and grown in fresh SDM-79 media supplemented with the selective drug. The clones were further confirmed both by PCR and Western Blot, as described in sections 2.2.2 and 2.2.5.3, respectively.

2.3.5 Stable transfection of bloodstream forms

Approximately 3 x 10^7 cells from a culture at ~1 x 10^6 cells.ml^{-1} were collected and centrifuged for 10 min at 405 g, room temperature. The resulting pellet was transferred to a 1.5 ml tube and centrifuged for 2 min at 665 g. The cells were then immediately re-suspended in 100 µl of Amaxa Human T Cell Nucleofactor kit solution (Lonza) and transferred to an electroporation cuvette (Lonza), to which 10 µg of linearised plasmid (prepared as described in section 2.2.1.11) was added. The cells were then electroporated using an Amaxa Nucleofector II machine (Lonza), and the X-001 programme. The cells were then immediately transferred into a 50 ml tube containing 30 ml of complete HMI-9 (Tube A), and inverted several times. From this, 3 ml were transferred to a new 50 ml tube containing 27 ml of complete HMI-9 (Tube B) and inverted several times. The same process was then repeated in order to have tubes A (1 x 10^6 cells.ml^{-1}), B (1 x 10^5 cells.ml^{-1}) and C (1 x 10^4 cells.ml^{-1}). These were then plated onto 24-well plates, 1 ml per well, and incubated overnight at 37°C with 5% CO_2. Next, for each batch of three plates, 75 ml of HMI-9 medium containing the double of the drugs concentration used selection was prepared, and 1 ml was added to each well of the three transfected plates, resulting in the selective drugs appropriate concentration, as shown in Table 2-7. The plates were then incubated at 37°C with 5% CO_2 for 5 to 7 days, during which surviving clones were isolated and grown in HMI-9. The clones were then further confirmed both by PCR and Western Blot as described in sections 2.2.2 and 2.2.5.3.

2.3.6 Stabilate preparation

For long-term storage of both PCF and BSF cell lines, 800 µl of a mid-log phase culture were added to 200 µl of 50% (v/v) of glycerol in complete SDM-79 or HMI-9 media, respectively, and stored in a 2 ml cryotube (AlphaLaboratories), at -80°C overnight, wrapped in cotton wool to avoid snap-freezing. Subsequently, the tubes were transferred to liquid nitrogen tanks, and the respective information data, was submitted to the Wellcome Trust Centre for Molecular Parasitology (WTCMP) Freezerworks sample management (http://www.freezerworks.com/) database.

2.4 RNA interference Assays

2.4.1 RNAi assays in procyclic cell forms

For this set of experiments, the following cell lines were used: 29-13 wt (parental cell line), TbORC1/CDC6 RNAi Cla, TbORC1/CDC6 RNAi Clb, TbORC4 RNAi Cla, TbORC4 RNAi Clb, Tb3120 RNAi Cla, Tb3120 RNAi Clb, and TbORC1B RNAi.

2.4.1.1 Setting up the RNAi-induced and non-induced cell cultures

Three days before the experiment, the cultures were passaged into 30 ml of SDM-79, supplemented with tetracycline free FBS (Gibco®), at a concentration of 5.5 x 10^5 cells.ml^{-1}, with the appropriate drugs: hygromycin and neomycin for the 29-13 parental line, and hygromycin, neomycin and zeocin (Table 2-7) for the 29-13 lines containing the integrated RNAi stem loop constructs. In the case of the TbORC4 RNAi Clb cell line, which has one allele of *TbORC4* endogenously tagged with 12myc, blasticidin was also added. On the day of the experiment, the different cell cultures' concentration was assessed, and 9 x 10^7 cells were diluted in 180 ml of SDM-79 with the respective drugs in a 75 cm^3 non-vented flask, and mixed, in order to have a homogeneous culture with 5 x 10^5 cells.ml^{-1}. From this, 90 ml of culture was transferred into a new 75 cm^3 non-vented flask, to which tetracycline (Calbiochem®, Merck Millipore) was added to a final concentration of 2 µg.ml^{-1} (Tet+ culture). The two resulting flasks, each with a 90 ml culture (Tet- and Tet+), were then placed in the upright position, in the incubator set at 27°C.

2.4.1.2 Growth curves

Every 24 h, both Tet- and Tet+ cultures were shaken gently, to homogenise the culture, and 10 µl of each were taken and used for cell concentration assessment using Neubauer improved hemocytometer, as standard. Because of the cell density at time points from 48h onwards (24h intervals), 10 µl of each culture were taken and diluted 1:20, to allow cell counting. Concentrations at the different time points were recorded and plotted on a xy graph using Prism 6 (GraphPad software Inc.), y-axis as a Log_{10} scale, x-axis as days post-tetracycline induction.

2.4.1.3 DNA Replication assessment assay

DNA replication was assessed at 6 h, 12 h and every 24 h time point intervals for 5 to 7 days by quantifying the number of cells that incorporated the thymidine analogue, 5-ethynyl-2´-deoxyuridine (EdU), following the protocol described in section 2.5.3. Images were acquired on an Axioskop 2 microscope system, as described in section 2.5.7, and were later analysed using the Fiji software (http://fiji.sc/Fiji) (Schindelin *et al.*, 2012). EdU-positive and EdU-negative cells were counted using the Cell Counter plugin http://fiji.sc/wiki/index.php/Cell_Counter), and the percentage of EdU-positive cells was calculated for each sample (Tet- and Tet+) per time point. The percentage of EdU-positive cells was then calculated relative to the percentage of the Tet- sample (considered to be 100%), and the resulting values were represented on a column graph, using Prism 6 (GraphPad software Inc.).

2.4.1.4 Cell cycle analysis by DAPI staining

Microscopy slides prepared for EdU detection were also stained with 4',6-diamidino-2-phenylindole (DAPI), and therefore also allowed the quantification of the number of cells in the different cell cycle stages according to their nucleus (N) and kinetoplast (K) ratio (1N1K, 1N2K, 2N2K, 0N1K, others). Images were acquired on an Axioskop 2 microscope system (section 2.5.7, and were analysed using the Fiji (http://fiji.sc/Fiji) (Schindelin *et al.*, 2012). The different cell types found in the population were quantified using the Cell Counter plugin http://fiji.sc/wiki/index.php/Cell_Counter), and the values represented as percentage of cells in each cell cycle stage relative to the total

number of cells in the sampled population, on a stacked bars grouped graph, using Prism 6 (GraphPad software Inc.).

2.4.1.5 Cell cycle analysis by flow cytometry

At 6 h, 12 h, and every 24 h time point intervals for 5 to 7 days, 1 ml of cells was collected and centrifuged at 1620 g for 10 min. The cells were then washed in 1x PBS supplemented with 5 mM of EDTA (Gibco®), and centrifuged again. The resulting pellet was re-suspended in 300 µl of 1x PBS supplemented with 5 mM EDTA, to which 700 µl of 100% Methanol (cooled at 4°C) was added, in a drop-wise fashion, while vortexing gently, in order to have a final fixing solution of 70% (v/v) Methanol. The cells were then wrapped in aluminium foil paper and left at 4°C for at least overnight. Next, the cells were centrifuged at 1620 g for 10 min, at 4°C, and washed once in 1x PBS supplemented with 5 mM EDTA. The pellet was then re-suspended in 1 ml of 1x PBS supplemented with 5 mM EDTA, 10 $\mu g.ml^{-1}$ of Propidium Iodide (PI; Sigma Aldrich), and 10 $\mu g.ml^{-1}$ of RNaseA (Sigma Aldrich), and incubated for 45 min at 37°C, protected from light. The cells were then transferred through a 35 µm nylon mesh cell strainer cap to a 6 ml tube (BD Falcon™ tube,BD Biosciences), and kept on ice until analysed using a BD FACSCalibur™ platform (BD Biosciences). Data was acquired from the FSC, SSC and FL2-A channels, using BD CellQuest™ software (BD Biosciences). Data was analysed and graphically represented using ©FlowJo software, version X.

2.4.1.6 Sample preparation for assessment of knock-down efficiency

For cell lines in which the RNAi targeted gene has been endogenously tagged, samples were prepared and analysed by western blot as described in previous sections (2.2.5.2 and 2.2.5.3). For the other cell lines, total RNA was extracted and converted into cDNA as described earlier (sections 2.2.3 and 2.2.4), which was diluted 1:10 in RNase-free water (Qiagen) prior to analysis by quantitative real-time PCR (below).

2.4.1.7 Quantitative Real-Time PCR (qPCR) of RNAi samples

For quantitative real-time PCR (qPCR), primers (Table 2-8) were designed using the Primer Express version 3.0 software (Bio Rad), and according to the guidelines (Alvarez-Fernandez, 2013; Dymond, 2013) for primers to be used in

qPCR. Primer sequences targeting TbORC1/CDC6 have been published previously (Tiengwe *et al.*, 2012b). For normalisation, previously described primers targeting the Tb927.10.12970 gene (C1) were used (Kabani *et al.*, 2009), as the "endogenous reference" gene. Primer efficiency and specificity were assessed for all pairs of primers through, respectively, the analysis of the calibration curve and melting profile, which resulted in efficiencies of approximately 100%, all within a 15% interval. For each pair of primers, triplicates of each sample cDNA were run per plate (MicroAmp® Optical 96-well Reaction Plate, Applied Biosystems®, Life Technologies), which were sealed with MicroAmp® clear adhesive film (Applied Biosystems®, Life Technologies). Precision™ qPCR MasterMix with SYBR Green and low ROX (Primerdesign) was used as follows: 12.5 µl of 2x MasterMix, 2 µl of primers mix (forward and reverse primers, 10 pmol stock), 5 µl of cDNA diluted 1:10, and 5.5 µl of RNase-free water (Qiagen), to a total of 25 µl per well. All experiments were performed using a 7500 Real Time PCR system (Applied Biosystems®), using the following PCR cycling conditions: 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. Fluorescence intensity data was collected at the end of the extension step (60°C for 1 min). The resulting fluorescence intensity data was then analysed by relative quantification using the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001), using the non-induced sample (Tet-) as the calibrator. Analysis of the amplification and dissociation curves was performed using the 7500 software version 2.3 (Applied Biosystems®), and results were exported to Excel (Microsoft Office), with which the $\Delta\Delta C_t$ calculations were performed. The results were then represented on a column graph, using Prism 6 (GraphPad software Inc.).

Table 2-8. Primers used in qPCR - RNAi assays

Number	Gene	Gene ID	Sense	Sequence
CTOL_7	TbORC1/CDC6*	Tb927.11.7216	Fwd	TTCACCCTGTCATGCAGGTTT
CTOL_8			Rev	GGTTCACTGACGCTGTCTTTCC
C 95	C1**	Tb927.10.12970	Fwd	TTGTGACGACGAGAGCAAAC
C 96			Rev	GAAGTGGTTGAACGCCAAAT
C 87	Tb3120	Tb927.9.4530	Fwd	GCTGCTTTGCAGGAAATACC
C 88			Rev	GCAGTGAAATGCTTCTGCTG
C 91	TbORC1B	Tb927.9.2030	Fwd	ACGTCAACTGTGCGGATATG
C 92			Rev	TCCAAGCGAACCTGTGAAC

* (Tiengwe *et al.*, 2012b); ** (Kabani *et al.*, 2009)

2.4.2 RNAi assays in bloodstream forms

For this set of experiments, the following cell lines were used: 2T1 wt (parental cell line), TbORC1/CDC6 RNAi Cl9a (with one allele of TbORC1/CDC6 tagged with C-terminal 12myc), TbORC1B RNAi Cl2a (with one allele of TbORC1B tagged with C-terminal 12myc), TbORC4 RNAi Cl12a, Tb7980 Cl3a (with one allele of Tb7980 tagged with N-terminal 12myc), and Tb3120 RNAi Cl5a (with one allele of Tb3120 tagged with C-terminal 12myc).

2.4.2.1 Setting up the RNAi-induced and non-induced cell cultures

The day before the experiment, all cell lines were passaged into 10 ml of fresh HMI-11 thymidine-free media (described in section 2.3.3) to a concentration of 1 x 10^5 cell.ml^{-1}, with the appropriate drug concentrations (Table 2-7): puromycin and phleomycin for the 2T1 wt parental line, hygromycin and phleomycin for the RNAi lines, and hygromycin, phleomycin and blasticidin for the RNAi and myc-tagged cell lines. On the day of the experiment, the different cell cultures' concentration was assessed, and 1.4 x 10^6 cells were diluted in 140 ml of HMI-11 thymidine free media, with the respective drugs, in a 75 cm^3 vented flask, and mixed, in order to have an homogeneous culture with 1 x 10^4 cells.ml^{-1}. From this, 70 ml of culture was transferred into a new 75 cm^3 vented flask, to which tetracycline (Calbiochem®, Merck Millipore) was added to a final concentration of 1 µg.ml $^{-1}$(Tet+ culture). The two resulting flasks, each with a 70 ml culture (Tet- and Tet+), were then placed, in the horizontal position, inside the incubator set at 37°C, with 5% of CO_2.

2.4.2.2 Growth curves

Every 24h, both Tet- and Tet+ cultures were shaken gently, to homogenise the culture, and 10 µl of each were taken for cell concentration assessment using Neubauer improved hemocytometer, as standard. Concentrations at the different time points were recorded and plotted on a xy graph using Prism 6 (GraphPad software Inc.), y-axis as a Log_{10} scale, x-axis as days post-tetracycline induction, Tet- and Tet+ of each cell line on the same graph.

2.4.2.3 DNA Replication assessment assay

For all cell lines, DNA replication was assessed every 24h time point intervals by quantifying the number of cells that incorporated EdU, by following the protocol described in section 2.5.6. Analysis was performed as in section 2.4.1.3.

2.4.2.4 Cell cycle analysis by microscopy

Performed as in section 2.4.1.4.

2.4.2.5 Sample preparation for assessment of knock-down efficiency

Performed as in section 2.4.1.6.

2.4.2.6 Quantitative Real-Time PCR (qPCR) of RNAi samples

Performed as in section 2.4.1.7.

2.5 Microscopy Techniques

2.5.1 Immunofluorescence assay of tagged proteins in PCF cells – generic protocol

Approximately 1 x 10^7 cells were collected from an exponentially growing PCF cell culture (~1 x 10^7 cells.ml^{-1}), and centrifuged for 10 min at 1620 g. The supernatant was then discarded, the pellet washed in 1x PBS (pH 7.2), and centrifuged again for 10 min at 1620 g. Next, the pellet was re-suspended in 200 µl of 1x PBS, and 20 µl of cell suspension was loaded onto each well of a 12 multi-well glass slide (Thermo Scientific), previously coated with Poly-L-lysine (Sigma Aldrich). The cells were allowed to settle for 4 min, after which they were fixed with 20 µl (per well) of 3.7% paraformaldehyde in 1x PBS, for 15 min. Next, each well was washed three times, for 5 min each, with 20 µl of 3% Bovine Serum Albumin (BSA) (Sigma Aldrich) in 1x PBS, and then incubated for 20 min with 20 µl of 0.2% Triton X-100 (Promega), diluted in 1x PBS. Subsequently, the wells were washed twice, for 5 min each, with 20 µl of 3% BSA in 1x PBS, after which each well was incubated for 1 h with 20 µl of the primary antiserum diluted in 1% BSA, according to the values on Table 2-9. Next, each well was washed three times, for 5 min each, with 20 µl of 3% BSA in 1x PBS, and

incubated for 1 h with 20 µl of the secondary antiserum conjugated to a fluorophore diluted in 1% BSA, according to the values on Table 2-9, with the slide protected from light. The wells were then washed three times, for 5 min each, with 20 µl of 3% BSA in 1x PBS, after which 10 µl of Fluoromount G with DAPI (Cambridge Bioscience, Southern Biotech) mounting media was added to each well, covered with a coverslip, and the slide sealed with regular nail varnish. If not imaged immediately, the slide was stored at 4˚C, protected from light. Note that if the primary antiserum is already conjugated to a fluorophore, the secondary antiserum step is omitted. If more than one protein was to be detected, another antisera were incubated as described above after the washes following the secondary antiserum incubation. All antibodies used for immunofluorescence and respective dilutions are summarised in Table 2-9.

Table 2-9. Antisera used for immunofluorescence assays

Name	Host	Target	Clone	Concentration	Manufacturer
α-myc Alexa Fluor® 488 conjugated	mouse	myc	4A6	1:500	Millipore
α-HA	mouse	HA	HA-7	1:1000	Sigma Aldrich
α-EP procyclin from *T. brucei*	mouse	EP1/EP2	TBRP1/247	1:500	Cedarlane®
α-VSG 221 from *T. brucei*	rabbit	VSG 221	-	1:10000	David Horn Lab
Alexa Fluor® 594 goat α-mouse IgG (H+L)	goat	mouse	-	1:1000	Molecular Probes®
Alexa Fluor® 594 goat α-rabbit IgG (H+L)	goat	rabbit	-	1:1000	Molecular Probes®
Alexa Fluor® 488 goat α-mouse IgG (H+L)	goat	mouse	-	1:1000	Molecular Probes®

2.5.2 Immunofluorescence assay of EP-procyclin in PCF cells

Approximately 1 x 10^7 cells were collected from an exponentially growing PCF cell culture (~1 x 10^7 cells.ml^{-1}), and centrifuged for 10 min at 1620 g. The supernatant was then discarded, the pellet washed in 1x PBS (pH 7.2), and centrifuged again for 10 min at 1620 g. Next, the pellet was re-suspended in 200 µl of 1x PBS, and 20 µl of cell suspension was loaded onto each well of a 12 multi-well glass slide (Thermo Scientific), previously coated with Poly-L-lysine (Sigma Aldrich). The cells were allowed to settle for 4 min, after which they

were fixed with 20 µl (per well) of 3.7% paraformaldehyde in 1x PBS, for 5 min. The cells were then incubated twice with 35 µl of 100 mM of glycine, for 10 min each, and further washed three times with 1x PBS, 5 min each, after which they were blocked for 1 h with 35 µl of 1% BSA with 0.2% Tween-20 in 1x PBS. Next, the cells were incubated for 1 h with 35 µl of mouse IgG1 α-EP procyclin antiserum (Table 2-9) diluted 1:500 in 1% BSA with 0.2% Tween-20 in 1x PBS. The wells were then washed three times with 1x PBS, 5 min each, and further incubated for 1 h, protected from light, with Alexa Fluor® 488 conjugated goat α-mouse antiserum (Table 2-9) diluted 1:1000 in 1% BSA with 0.2% Tween-20 in 1x PBS. Again, the wells were washed three times with 1x PBS, after which 10 µl of Fluoromount G with DAPI (Cambridge Bioscience, Southern Biotech) mounting media was added to each well, covered with a coverslip, and the slide sealed with regular nail varnish. The protocol was optimised and shared by Fernando Fernandez-Cortes, from Jeremy Mottram's laboratory, University of Glasgow.

2.5.3 Immunofluorescence assay of incorporated EdU in PCF cells

From an exponentially growing PCF culture (~1 x 10^7 cells.ml^{-1}), a 1 ml aliquot was transferred to a 24-well plate, and the cells were then incubated for 3 h at 27°C with 50 µM of 5-ethynyl-2´-deoxyuridine (EdU; Life Technologies). The cells were then collected and centrifuged for 10 min at 1620 g. The supernatant was then discarded, the pellet washed in 1x PBS (pH 7.2), and centrifuged for 10 min at 1620 g. Next, the pellet was re-suspended in 200 µl of 1x PBS, and 20 µl of cell suspension was loaded onto each well of a 12 multi-well glass slide (Thermo Scientific), previously coated with Poly-L-lysine (Sigma Aldrich). The cells were allowed to settle for 4 min, after which the cell suspension was removed from the wells. The cells were then fixed for 15 min with 20 µl (per well) of 3.7% paraformaldehyde in 1x PBS. Following, each well was washed three times, for 5 min each, with 20 µl of 3% BSA in 1x PBS, and then permeabilised for 20 min with 20 µl of 0.5% Triton X-100 (Promega), diluted in 1x PBS. Afterwards, the Triton solution was removed, and the wells washed twice, for 5 min each, with 20 µl of 3% BSA in 1x PBS. To each well, 25 µl of Click-iT® EdU detection mix (Click-iT® EdU Alexa Fluor® 555 or 594 Imaging Kit – Life Technologies) was added and incubated for 1 h. The EdU detection mix (for a volume of 25 µl per well) was prepared immediately before use, as follows: 21.5µl of 1x Reaction Buffer, 1 µl

of copper sulphate ($CuSO_4$), 0.06 µl of Alexa Fluor® 555 or 0.24 µl of Alexa Fluor® 594, and 2.5 µl of 1x Reaction Additive, added by this order. Each well was then washed 4-5 times with 3% BSA in 1x PBS (after this step, in the case tagged proteins were to be detected, the cells were incubated with the appropriate antisera, following the steps described in section 2.5.1), after which 10 µl of Fluoromount G with DAPI (Cambridge Bioscience, Southern Biotech) mounting media was added to each well, covered with a coverslip, and the slide sealed with regular nail varnish.

2.5.4 Immunofluorescence assay of 12myc-tagged proteins in BSF cells

Approximately 2 x 10^6 cells were collected from an exponentially growing BSF cell culture (~1 x 10^6 cells.ml^{-1}), and centrifuged for 5 min at 1000 g. The supernatant was then discarded, the pellet washed in 1 ml of 1x PBS (pH 7.2), and centrifuged again for 5 min at 1000 g. The pellet was then re-suspended in 50 µl of 1x PBS, to which was then added 250 µl of 3.7% paraformaldehyde, allowing the cells to fix for 5 min. Next, 10 ml of 1x PBS was added, and the cells centrifuged again for 5 min at 1000 g. The supernatant was then discarded, leaving ~20 µl to re-suspend the pellet. The cell suspension was then loaded onto each well of a 12 multi-well glass slide (Thermo Scientific), previously coated with Poly-L-lysine (Sigma Aldrich), and let to settle for 5 min. The cell suspension was removed from the wells, and the cells were permeabilised with 20 µl of 0.2% Triton X-100 (Promega), diluted in 1x PBS. Afterwards, the wells were washed three times, for 5 min each, with 20 µl of 3% BSA in 1x PBS. Then, each well was incubated for 1 h with 20 µl of mouse α-myc Alexa Fluor 488 conjugated antiserum (Table 2-9), diluted 1:500 in 1% BSA in 1x PBS, and the slide protected from light. Afterwards, each well was washed twice, for 5 min each, with 20 µl of 3% BSA in 1x PBS, after which 10 µl of Fluoromount G with DAPI (Cambridge Bioscience, Southern Biotech) mounting media was added to each well, covered with a coverslip, and the slide sealed with regular nail varnish.

2.5.5 Immunofluorescence assay of VSG 221 in BSF cells

Approximately 1 x 10^6 cells were collected from an exponentially growing BSF cell culture (~1 x 10^6 cells.ml^{-1}), and centrifuged for 5 min at 1000 g. The supernatant was then discarded, and the pellet re-suspended in 500 µl of HMI-9 media, to which 500 µl of 2% formaldehyde was added. The suspension was then incubated for 1 h on ice, and then centrifuged at 1000 g for 5 min. The pellet was then re-suspended in 1 ml of ice-cold 1x PBS, and centrifuged again. The resulting pellet was then re-suspended in 1 ml of 1% BSA in 1x PBS, and centrifuged for 5 min at 1000 g. The supernatant was then mostly removed, leaving ~30 µl in which the pellet was re-suspended, and then loaded onto each well of a 12 multi-well glass slide (Thermo Scientific), previously coated with Poly-L-lysine (Sigma Aldrich), and let to settle for 10 min. The wells were then washed once in 1x PBS for 5 min, and incubated with 25 µl of 50% FBS (v/v) in 1x PBS for 15 min. Following, the cells were incubated with 20 µl of the rabbit α-VSG 221 antiserum (Prof. David Horn Lab, University of Dundee) diluted 1:10000 in 3% FBS in 1x PBS, for 1h. Next, the wells were washed twice, 5 min each, with 1x PBS, and further incubated with 20 µl of Alexa Fluor® 594 conjugated goat α-rabbit antiserum (Table 2-9) diluted 1:1000 in 3% FBS in 1x PBS, for 1 h, protected from light. The wells were then washed twice, 5 min each, with 1x PBS, after which 10 µl of Fluoromount G with DAPI (Cambridge Bioscience, Southern Biotech) mounting media was added to each well, covered with a coverslip, and the slide sealed with regular nail varnish. The protocol was adapted from the one routinely used in Professor David Horn's laboratory, University of Dundee. Kindly shared by Dr Lucy Glover.

2.5.6 Immunofluorescence assay of incorporated EdU in BSF cells

During the course of the RNAi assay, 3 ml of culture were transferred to a vented flask, to which 150 mM of EdU was added, and incubated for 4 h at 37°C with 5% CO_2. The cells were then centrifuged at 1000 g for 5 min, the pellet washed in 1x PBS, and the cells centrifuged again. The pellet was then re-suspended in 20 µl of 1x PBS, and loaded onto each well of a 12 multi-well glass slide (Thermo Scientific), previously coated with Poly-L-lysine (Sigma Aldrich), and allowed to settle for 5 min. The cells were fixed in 3.7% paraformaldehyde for 5 min, and

then washed twice, 5 min each, with 3% BSA in 1x PBS. Following, the cells were permeabilised with 0.2% Triton X-100 (Promega), diluted in 1x PBS. Afterwards, the wells were washed twice, for 5 min each, with 20 µl of 3% BSA in 1x PBS. To each well, 25 µl of Click-iT® EdU detection mix (Click-iT® EdU Alexa Fluor® 555 Imaging Kit – Life Technologies) was added and incubated for 1 h. The EdU detection mix (for a volume of 25 µl per well) was prepared just before use, and consists of: 21.5µl of 1x Reaction Buffer, 1 µl of copper sulphate ($CuSO_4$), 0.25 µl of Alexa Fluor® 555, and 2.5 µl of 1x Reaction Additive, added by this order. Each well was then washed 6 times with 3% BSA in 1x PBS, after which 10 µl of Fluoromount G with DAPI (Cambridge Bioscience, Southern Biotech) mounting media was added to each well, covered with a coverslip, and the slide sealed with regular nail varnish.

2.5.7 Microscopy imaging acquisition and image analysis

For quantification of localisation and intensity of each protein's (TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc}, Tb1120^{12myc} and TbORC1B^{12myc}) signal throughout the cell cycle, images were acquired using an Axioskop 2 fluorescence microscope (Zeiss) attached to a HBO100 lamp and a digital ORCA-ER camera and camera controller (Hamamatsu Photonics). In all cases, the 63x DIC magnification lens was used. Images were acquired using the Volocity® 6.1.1 Cellular Imaging and Analysis software (Perkin Elmer).

For the acquisition of images with a better resolution, an Olympus IX71 DeltaVision Core System (Applied Precision, GE) equipped with a CoolSNAP HQ2 CCD camera (Photometrics®) was used. In all cases, the Olympus 100x/1.40, UPLS Apo lens was used, and Z-stacks of 5 µm thick (25 sections of 0.2 µm thickness each) were acquired in images with a 512 x 512 resolution using the SoftWoRx suite 2.0 software (Applied Precision, GE). Images were deconvolved using the ratio conservative method applied by the SoftWoRx software (Applied Precision, GE).

Super resolution structure illumination microscopy was performed using an Elyra PS.1 super resolution microscope (Zeiss) equipped with sCMOS PCO camera. In all cases, the Plan-Apochromat 63x/1.4 Oil DIC M27 lens was used, and Z-stacks of 4 µm thick (45 sections of 0.09 µm thickness each) were acquired using the ZEN

Black Edition Imaging software (Zeiss). Images were then processed in ZEN Black Edition Imaging software (Zeiss), by choosing the structural illumination manual processing tool, with a noise filter of -6.0, and an output as SR-SIM (super resolution structural illumination microscopy). Microscope handling, image acquisition and processing were all done with the essential support of Prof. Markus Meissner, Dr Gurman Pall and Carmen Melatti.

All images were further analysed using Fiji (http://fiji.sc/Fiji) (Schindelin *et al.*, 2012) in order to generate the images shown in Chapters 3 and 4. In both Chapters, all microscopy images shown per figure were obtained from the same experiment (same sample and microscopy slide), using the same exposure time, and later analysed in Fiji using the same parameters. In cases where brightness and contrast were adjusted, the same values were applied to all images being processed from the same experiment. In the case of measuring the intensity of the detected fluorescent signal (DAPI, myc or EdU), a circular 21 x 21 pixel region of interest (ROI) was drawn around each individual cell nucleus. Next, the image was treated using the rolling ball background subtraction plugin (http://fiji.sc/Rolling_Ball_Background_Subtraction), set up with a radius of 50 pixels. After the background was removed, the mean pixel intensity within the ROI was measured and plotted onto a vertical scatter plot against the corresponding cell type (1N1K, 1N1eK, 1N2K, 2N2K), using Prism 6 (GraphPad software Inc.).

2.6 Protein Interactions and Dynamics Assays

2.6.1 Immunoprecipitation

2.6.1.1 Preparation of the magnetic beads

Approximately ~3 x 10^7 (50 µl) of Dynabeads® M-280 Sheep α-mouse IgG (Novex®, Life Technologies) magnetic beads were transferred into a 1.5 ml tube (DNase, RNase and protease free). The beads were then washed in 1 ml of cold blocking solution (0.5% BSA in 1x PBS, pH 7.2), vortexed gently, and placed on a DynaMag™-2 magnet (Life Technologies). The supernatant was discarded, and the beads washed again, twice. Finally, the beads were re-suspended in 125 µl of blocking solution containing 5 µg of mouse α-myc clone 4A6 antiserum (Millipore), and left overnight on mixing rotator, at 4°C.

2.6.1.2 Cell lysis

Around 1.5 x 10^9 cells were collected from an exponentially growing PCF cell culture (~1 x 10^7 cells.ml^{-1}), and centrifuged for 10 min at 1620 g. The cells were then washed once in 1x PBS, and centrifuged again. The pellet was then re-suspended in 3 ml of whole cell extract buffer (WCE; 50 mM of HEPES pH 7.55, 100 mM of NaCl, 1 mM of EDTA pH 8, 1 mM of EGTA pH 8, 10% Glycerol, 1% Triton X-100, 1 mM of DTT, and 2x complete protease inhibitor cocktail - Roche), divided into 3 individual 1.5 ml tubes, 1 ml each, and left on ice for 30 min. The lysates were then centrifuged at 15000 g for 30 min, at 4°C, and from the supernatant, 900 µl were saved for the immunoprecipitation step, while 30 µl were added to 10 µl of 4x NuPAGE® LDS Sample Buffer (Life Technologies), and stored at -20°C, as the input sample. Optionally, 2.5 x 10^6 cells were also collected form the cell culture, and the pellet re-suspended in 10 µl of 1x NuPAGE® LDS Sample Buffer (Life Technologies) diluted in 1x PBS, and stored at -20°C, to be used as a pre-lysis sample.

2.6.1.3 Immunoprecipitation and Elution

All steps were performed at 4°C, except otherwise mentioned. The magnetic beads were washed three times in blocking solution (as described in section 2.6.1.1), and re-suspended in 50 µl of blocking solution. To this, 900 µl of cell lysate was added, and then incubated for 2 h on a mixing rotator. The 1.5 ml tubes were then placed on the magnet, and 30 µl were collected from the supernatant, and added to 10 µl of 4x NuPAGE® LDS Sample Buffer (Life Technologies), and stored at -20°C, as the flowthrough sample. The remaining supernatant was discarded. To the pelleted beads, 1 ml of washing buffer (50 mM of HEPES pH 7.55, 100 mM of NaCl, 1 mM of EDTA pH 8, 1 mM of EGTA pH 8, 10% Glycerol, 0.1% Triton X-100, 1 mM of DTT, and 2x complete protease inhibitor cocktail) was added, the tube gently shaken, and placed on the mixing rotator for 5 min. These washing steps were repeated another three times. After the last wash, the supernatant was discarded, and the beads centrifuged for 3 min at 1000 g. The tubes were placed onto the magnet, and the remainder supernatant discarded. The beads were then re-suspended, by gentle vortexing and short spin, in 15 µl of 1x NuPAGE® LDS Sample Buffer (Life Technologies) diluted in 1x PBS, and incubated for 10 min at 70°C. The samples were then

centrifuged for 1 min at 15000 g, room temperature, and placed on the magnet. The supernatant was then saved as the elution sample. Because the lysate was separated onto three fractions, in the end we will have three eluates that correspond to the same original sample. Therefore, these three eluates were pooled out into a same final elution sample of 45 µl. All collected samples, pre-lysis, input, flowthrough and elution, were then incubated for 5 min at 100°C, and loaded onto 10 – 12 wells NuPAGE® Novex® 10% Bis-Tris Protein Gels, 1.0 mm, which were run as described in section 2.2.5.2. For western blot analysis (myc-tagged proteins detection), 10 µl of pre-lysis, input, flowthrough samples were loaded, while only 1 µl of elution sample was used. The remainder 44 µl of elution sample were loaded onto a separate gel, and stained with SYPRO® Ruby, as described in section 2.2.5.4. Sections of the gel were then excised using a sterile scalpel, stored in a 1.5 ml tube at -20°C, and further sent for mass spectrometry analysis at Glasgow Polyomics (University of Glasgow; see below).

2.6.1.4 Nanoflow HPLC Electrospray Tandem Mass Spectrometry (nLC-ESI-MS/MS

The following protocol and analysis were performed at Glasgow Polyomics (University of Glasgow) by Dr Christina Naula and Dr Stefan Weidt, under the supervision of Dr Richard Burchmore. Protocol provided by Dr Richard Burchmore.

From the received samples, peptides were solubilised in 2% acetonitrile with 0.1% trifluroacetic acid and fractionated on a nanoflow uHPLC system (Thermo RSLCnano). Next, online analysis by electrospray ionisation (ESI) mass spectrometry on an Amazon Speed ion trap MS/MS (Bruker Daltonics) was performed. Peptide separation was then carried out on a Pepmap C18 reversed phase column (LC Packings), and peptides were then desalted and concentrated for 4 min on a C18 trap column, followed by an acetonitrile gradient (in 0.1% (v/v) formic acid) (3.2 - 32% (v/v) 4 - 27 min, 32% to 80% (v/v) 27 - 36 min, held at 80% (v/v) 36- 41 min and re-equilibrium at 3.2%), for a total time of 45 min. A solvent flow rate of 0.3 µl.min^{-1} was used for the analytical column, while the trap column solvent flow rate was of 25 µl.min^{-1}, using 2% acetonitrile with 0.1% v/v trifluroacetic acid. Mass spectrometric (MS) analysis was performed using a continuous duty cycle of survey MS scan, followed by up to ten MS/MS analyses

of the most abundant peptides, choosing the most intense multiply charged ions with dynamic exclusion for 120 sec. MS data was then processed using Data Analysis software (Bruker) and the automated Matrix Science Mascot Daemon server (v2.4.1). Peptides were next identified using the Mascot search engine to interrogate protein sequences in the annotated proteins database (TriTrypDB), allowing a mass tolerance of 0.4 Da for both MS and MS/MS analyses.

2.6.2 Co-immunoprecipitation

Identical to the protocol described in section 2.6.1, but with slight differences. Only 5 x 10^8 cells were collected from an exponentially growing PCF cell culture (~1 x 10^7 cells.ml^{-1}) for lysis, in 1 ml of whole cell extract buffer. All co-immunoprecipitations presented in this work were performed using the HA-tag as "bait", and therefore the beads were incubated overnight in blocking solution with 5 µg of mouse α-HA clone HA-7 antiserum (Sigma Aldrich), as described in section 2.6.1.1. All immunoprecipitation, washing, and elution steps were performed as described in section 2.6.1.3. All samples were then analysed by western blot as described in sections 2.2.5.2 and 2.2.5.3, with the membrane being probed for both myc- and HA-tagged proteins detection.

2.6.3 Cell Fractionation

Protocol adapted from (Zeiner *et al.*, 2003). Around 5 x 10^8 PCF cells were collected from an exponentially growing PCF cell culture (~1 x 10^7 cells.ml^{-1}), and centrifuged at 1620 g, for 10 min at 4°C. The pellet was then washed twice in 5 ml of fractionation buffer (FB; 150 mM of Sucrose, 20 mM of KCl, 3 mM of $MgCl_2$, 20 mM of HEPES-KOH, pH 7.9, 1 mM of DTT, and 1x complete protease inhibitor cocktail – Roche), and centrifuged at 1620 g, for 10 min at 4°C. The pellet was then re-suspended in 1 ml of FB buffer, supplemented with 0.2% (v/v) of NP-40. The suspension was then passed through a 26G syringe needle, to generate the cell lysate, that was then centrifuged at 20000 g for 10 min at 4°C. The supernatant (cytoplasmic fraction) and pellet (nuclear fraction) were then separated and treated separately, as follows. The supernatant was then re-centrifuged twice, at 20,000 g for 10 min at 4°C, each time, transferred to a new 1.5 ml tube, and saved at -20°C until analysis by western blot. The pellet was then re-suspended in 500 µl of FB buffer, and passed 15x through a 26G

syringe needle. The solution was then centrifuged at 20000 g for 10 min at 4°C, and the supernatant discarded. The pellet was then rinsed in 500 µl of FB buffer, without being re-suspended. Finally, the pellet was re-suspended in 500 µl of FB buffer, and stored at -20°C until analysis by western blot. Both cytoplasmic and nuclear fractions were then analysed by western blot as described in section 2.2.5.3, and probed with α-myc antiserum, to detect the myc-tagged protein; α-NOG1 to detect TbNOG1, or α-H2A to detect histone H2A, and test the presence of an exclusively nuclear protein in both fractions; and α-OPB to detect the presence of TbOPB, and thus test the presence of a cytoplasmic protein in both fractions (antisera information in Table 2-6).

2.6.4 Gel Filtration

All steps involving the column and the ÄKTApurifier system were performed by Mr Alan Scott (Institute of Infection, Immunity and Inflammation, University of Glasgow).

2.6.4.1 System preparation and column equilibration

Approximately 24 h before the run, 1 L of running buffer (50 mM of HEPES pH 7.55, 100 mM of NaCl, 1 mM of EDTA pH 8, 1 mM of EGTA pH 8, 10% Glycerol, 1% Triton X-100, 1 mM of DTT, and 0.25x complete protease inhibitor cocktail) was prepared, filter sterilised (0.22 µm bottle top filter, Millipore), degasified by filtration and kept at 4°C. The column used, HiLoad 16/60 Superdex 200 Prep Grade (GE Healthcare Life Sciences), was set up in a ÄKTApurifier system (GE Healthcare Life Sciences), and controlled with Unicorn 5.31 software (GE Healthcare Life Sciences). By default, the column was kept in 20% Ethanol until use. The column was first equilibrated with two column volumes of degasified MilliQ water, using a flow rate of 750 µl per min, followed by two column volumes of the running buffer, using the same flow rate, before sample injection.

2.6.4.2 Cell lysis

Around 7.5 x 10^8 cells were collected from an exponentially growing PCF cell culture (~1 x 10^7 cells.ml^{-1}), and centrifuged at 1620 g for 10 min. The cells were washed in 10 ml of 1x PBS, and centrifuged again. The pellet was then re-

suspended in 2 ml of lysis solution (50 mM of HEPES pH 7.55, 100 mM of NaCl, 1 mM of EDTA pH 8, 1 mM of EGTA pH 8, 10% Glycerol, 1% Triton X-100, 1 mM of DTT, and 2x complete protease inhibitor cocktail), and incubated on ice for 30 min. The 2 ml of lysate were then transferred into a 2.2 ml thin wall propylene centrifuge tubes (11 x 35 mm) (Beckman Coulter), and centrifuged at 100000 g for 1 h, at 4°C, using an Optima™ TL Ultracentrifuge (Beckman Coulter) equipped with a TLS 55 rotor (Beckman Coulter). The lysate was then filtered (0.2 µm Ministart® Syringe Filter, Sartorius), and 30 µl were added to 10 µl of 4x NuPAGE® LDS Sample Buffer (Life Technologies), and stored at -20°C, as the lysis sample, while the remainder volume was used for the gel filtration.

2.6.4.3 Gel filtration run and fraction collection

The 2 ml of lysate were injected into the ÄKTApurifier system, and run in a total of 210.8 ml of running buffer at a flow rate of 500 µl.min^{-1}. Fractions of 1 ml were collected into ABgene 2.2 ml 96-well storage plates (Thermo Scientific) for further analysis. Fractions collected immediately before the void volume peak, from 43 ml (corresponding to around 2116 kDa) down to 84 ml (corresponding to 48 kDa, according to the standards used) of eluted volume, were used (40 µl from each fraction) for further western blot analysis, as described in sections 2.2.5.2 and 2.2.5.3.

2.7 Fluorescence-Activated Cell Sorting (FACS)

In all cases, the BD FACSAria I™ Cell Sorter system set up and calibration were performed by Ms Diane Vaughan and Mr Craig Lapsley (Flow Cytometry Laboratory, Institute of Infection, Immunity and Inflammation, University of Glasgow).

2.7.1 FACS of *T. brucei* PCFs

Approximately 1 x 10^9 cells were collected from an exponentially growing PCF culture (~1 x 10^7 cells.ml^{-1}), and centrifuged for 10 min at 1620 g. The pellet was then washed in 10 ml of 1x PBS supplemented with 5 mM of EDTA (Gibco®), and centrifuged for 10 min at 1620 g. Next, the cells were re-suspended in 12 ml of 1x PBS supplemented with 5 mM of EDTA, to which to which 28 ml of 100% ice cold-Methanol was added, in a drop-wise fashion while vortexing gently, so that

the final fixing solution was of 70% (v/v) Methanol, and the cell concentration of 2.5 x 10^7 cells.ml^{-1}. The tube was wrapped in aluminium foil paper and kept at 4°C from overnight up to three weeks. For every FACS sorting run, four FACS tubes (Becton Dickinson) were prepared, each starting with 4 ml of fixed cells (~1 x 10^8 cells). The cells were collected and centrifuged for 10 min at 1000 g, at 4°C, washed in 1 ml of 1x PBS supplemented with 5 mM of EDTA, and centrifuged again for 10 min at 1000 g, at 4°C. The pellet was then re-suspended in 4 ml of 1x PBS supplemented with 5 mM of EDTA, 10 µg.ml^{-1} of PI (Sigma Aldrich) and 10 µg.ml^{-1} of RNase A (Sigma Aldrich), and incubated for 45 min at 37°C, in the dark. The cells were then transferred to a FACS tube through a cell strainer cap (BD Biosciences), and sorted into G1, early S, late S and G2 phases using a BD FACSAria I™ Cell Sorter (BD Biosciences). An average of 1 x 10^7 cells were recovered for the G1 subset, 1 x 10^6 cells for both early and late S phases, and 3 x 10^6 cells for G2 phase, per FACS sorting session (three sessions were done in total for each *T. brucei* strains). The sorted cells were collected at 4°C into new FACS tubes containing 200 µl of lysis buffer (1 M NaCl, 10 mM EDTA, 50 mM Tris-HCL pH 8.0, 0.5% SDS, 0.4 mg.ml^{-1} Proteinase K, and 0.8 µg.ml^{-1} of Glycogen) (Azuara, 2006). The collected cells were then incubated for 2 h at 55°C, and the lysate was stored at -20°C until gDNA extraction (section 2.7.4).

2.7.2 FACS of *T. brucei* BSFs

For each sorting session, approximately 3 x 10^8 cells were collected from an exponentially growing BSF cell culture (~1 x 10^6 cells.ml^{-1}), and centrifuged for 10 min at 1000 g. For each pellet containing around 1 x 10^8 cells (resulting from each 100 ml cultures), cells were then re-suspended in 25 ml of 1x PBS and centrifuged for 10 min at 1000 g. The pellet was then re-suspended in 500 µl of 1x PBS, and 9.5 ml of 1% Formaldehyde (methanol-free, Thermo Scientific) diluted in 1x PBS was added. The cells were fixed for 10 min at room temperature, and then centrifuged for 10 min at 1000 g, washed once in 10 ml 1x PBS, and centrifuged again for 10 min at 1000 g. The pellet was next re-suspended in a final volume of 1x PBS so to have a concentration of 2.5 x 10^7 cells.ml^{-1}, and was stored protected from light at 4°C overnight. The fixed cells were centrifuged for 10 min at 1000 g, and then re-suspended and incubated in 20 ml of 0.01% Triton X-100 (Promega) in 1x PBS for 30 min at room temperature. The cells were then centrifuged for 10 min at 700 g, washed in 20

ml of 1x PBS, and centrifuged again. The resulting pellet was then re-suspended in a final volume of 1x PBS with 10 µg.ml^{-1} of PI (Sigma Aldrich) and 100 µg.ml^{-1} of RNase A (Sigma Aldrich) so to have a concentration of ~2.5 x 10^7 cells.ml^{-1}, and were incubated for 1 h at 37°C, protected from light. The cells were then transferred to a FACS tube through a cell strainer cap (BD Biosciences), and sorted into G1, early S, late S and G2 phases using a BD FACSAria I™ Cell Sorter (BD Biosciences). An average of 5 x 10^6 cells were recovered for the G1 subset, 1.5 x 10^6 cells for both early and late S phases, and 4 x 10^6 cells for G2 phase, per FACS sorting session (two sessions were done in total). The sorted cells were collected at 4°C into new FACS tubes containing 200 µl of lysis buffer (1 M NaCl, 10 mM EDTA, 50 mM Tris-HCL pH 8.0, 0.5% SDS, 0.4 mg.ml^{-1} Proteinase K, and 0.8 µg.ml^{-1} of Glycogen) (Azuara, 2006). After the sorting has been completed, the collected cells were then incubated for 2 h at 55°C, and the lysate was stored at -20°C until gDNA extraction (section 2.7.4).

2.7.3 FACS of *L. major* and *L. mexicana* promastigotes

For these assays, *Leishmania major* strain Friedlin and *Leishmania mexicana* strain U1103 promastigote cell cultures were kindly set up by Dr Amy Goundry (Jeremy Mottram's laboratory), in 150 ml of modified Eagle's medium (designated HOMEM medium, Gibco®), supplemented with 10% (v/v) FCS and 1% (v/v) penicillin-streptomycin solution (Sigma Aldrich), and incubated at 25°C.

Approximately 1 x 10^9 cells were collected from an exponentially growing promastigote cell culture (~5 x 10^6 cells.ml^{-1}), and centrifuged for 10 min at 1000 g. The pellet was then washed in 10 ml of 1x PBS supplemented with 5 mM of EDTA (Life Technologies), and centrifuged for 10 min at 1000 g. Next, the cells were re-suspended in 12 ml of 1x PBS supplemented with 5 mM of EDTA, to which to which 28 ml of 100% ice cold-Methanol was added, in a drop-wise fashion while vortexing gently, so that the final fixing solution was of 70% (v/v) Methanol, and the cell concentration of 2.5 x 10^7 cells.ml^{-1}. The tube was wrapped in aluminium foil paper and kept at 4°C from overnight up to three weeks. For every FACS sorting run, four FACS tubes were prepared, each starting with 4 ml of fixed cells (1 x 10^8 cells). The cells were collected and centrifuged for 10 min at 1000 g, at 4°C, washed in 1 ml of 1x PBS supplemented with 5 mM of EDTA, and centrifuged again for 10 min at 1000 g, at 4°C. The pellet was then

re-suspended in 4 ml of 1x PBS supplemented with 5 mM of EDTA, 10 µg.ml^{-1} of Propidium Iodide (Sigma Aldrich) and 10 µg.ml^{-1} of RNase A (Sigma Aldrich), and incubated for 45 min at 37°C, in the dark. The cells were then transferred to a 6 ml BD Falcon tube through a cell strainer cap, 35 µm nylon mesh, (BD Biosciences), and sorted into G1, early S, late S and G2 phases using a BD FACSAria I™ Cell Sorter (BD Biosciences). An average of 1 x 10^7 cells were recovered for the G1 subset, 1 x 10^6 cells for both early and late S phases, and 3 x 10^6 cells for G2 phase, per FACS sorting session (three sessions were done in total for each *Leishmania* species). The sorted cells were collected at 4°C into new FACS tubes containing 200 µl of lysis buffer (1 M NaCl, 10 mM EDTA, 50 mM Tris-HCL pH 8.0, 0.5% SDS, 0.4 mg.ml^{-1} Proteinase K, and 0.8 µg.ml^{-1} of Glycogen) (Azuara, 2006). After the sorting has been completed, the collected cells were then incubated for 2 h at 55°C, and the lysate was stored at -20°C until gDNA extraction (section 2.7.4).

2.7.4 gDNA extraction of post-FACS samples

gDNA was extracted using the Blood and Tissue DNA extraction kit, from Qiagen, with some adaptations to the manufacturer's protocol. The lysates from the different FACS sorting sessions were thawed at 37°C, and pooled out per cell cycle stage. One-third volume of 100% ethanol was then added, the resulting solution thoroughly vortexed, and then transferred into a DNeasy Mini spin column (Qiagen). The column was then centrifuged for 1 min at 6000 g, and the flow through discarded; this step was repeated until all volume had been passed through the column. The column was then washed in 500 µl of Buffer AW1, and again centrifuged for 1 min at 6000 g, discarding the flowthrough. Next, 500 µl of Buffer AW2 were added, and the column was centrifuged for 3 min at 20000 g, after which the flowthrough was discarded. The gDNA was then eluted in 50 µl of Buffer AE, and stored at -20°C until further use.

2.7.5 Whole cell extract preparation of post-FACS PCF samples

In the case of analysis by western blot, FACS of PCF cells was performed as described in section 2.7.1, but instead of collecting the cells in lysis buffer, the cells were collected into 1x PBS supplemented with 5 mM of EDTA. The collected cells were then transferred from the collection tubes into 1.5 ml tubes and

centrifuged at 2000 g for 10 min. The supernatant was discarded and the pellets pooled out per sorting group, repeating the centrifugation step several times. Finally, for each sorting group, the pellet was then re-suspended in 10 µl 1x NuPAGE® LDS Sample Buffer (Life Technologies), and used for western blot analysis as mentioned in sections 2.2.5.2 and 2.2.5.3.

2.8 Marker Frequency analysis (MFA)

2.8.1 MFA by deep sequencing (MFA-seq)

2.8.1.1 DNA Library preparation and sequencing

Using the gDNA extracted from sorted samples of *T. brucei* PCFs strain TREU 927, and *L. major* and *L. mexicana* promastigotes, DNA libraries were prepared using the Nextera® XT DNA Sample Preparation kit (Illumina), by staff at Glasgow Polyomics (University of Glasgow). The libraries were the sequenced using the Illumina MiSeq paired-end 250bp sequencing system (Illumina), at Glasgow Polyomics (University of Glasgow). In all cases, the samples were multiplexed with each of the early S, late S, and G2 phases per species/strain in the same run for ease of comparison. In contrast, gDNA from *T. brucei* PCF and BSF Lister 427 cells was sent for sequencing at Eurofins Genomics (Germany). The DNA library was prepared using the TruSeq® DNA Sample Preparation kit (Illumina), and sequenced using Illumina HiSeq paired- end 100 bp sequencing system (Illumina). The samples were multiplexed, with each of the early S, late S, and G2 phase samples library DNA, both from BSF and PCF, being processed in the same run, for ease of comparison.

2.8.1.2 Whole genome deep sequencing results analysis

All program coding and pipeline design was conceived and developed by Dr Nicholas J. Dickens (Wellcome Trust Centre for Molecular Parasitology Bioinformatics team, University of Glasgow). The generic programming code is shown in the appendices section 7.8, with notes.

Data resultant from the sequencing was firstly analysed for quality control using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and then trimmed using fastq-mcf (Aronesty, 2011) (http://code.google.com/p/ea-

utils), to exclude the adapter sequences used during the library preparation and sequencing. The reads were then aligned to the respective reference genomes (*T. brucei* TREU 927, *L. major* Friedlin, and *L. mexicana* U1103 reference genomes were retrieved from TriTrypDB version 6.0; *T. brucei* Lister 427 was retrieved from TriTrypDB version 8.0) using Bowtie2 (version 2.2.0 --very-sensitive-local -k1) (Langmead and Salzberg, 2012). The aligned reads were then compared using a method adapted from the one described previously (Tiengwe *et al.*, 2012a), but simplified to facilitate inter-species comparisons. Briefly, the reads were binned in 2.5 Kbp sections along each chromosome, and the number of reads in each bin was then used to calculate the ratios between early S and G2, as well as between late S vs G2 samples, scaled for the total size of the read library (reads per 2.5 Kbp per million reads mapped). These data were then represented in a graphical form using ggplot2 and the R package (version 3.0.2) (R Development Core Team, 2010), or Prism 6 (GraphPad software Inc.)

2.8.2 MFA by Quantitative Real-Time PCR (MFA-qPCR)

A strategy employed previously (Tiengwe *et al.*, 2012a) was used, following the MIQE guidelines (Bustin *et al.*, 2009). Primers (Table 2-10) were designed for several regions across *L. major* chromosomes 8, 20, 29, 34 and 36, as well as *L. mexicana* chromosomes 8 and 20, using Primer Express version 3.0 (Bio Rad), and according to the guidelines (Alvarez-Fernandez, 2013; Dymond, 2013) for primers to be used in qPCR. Primer sizes ranged from 17-24 bp, with melting temperatures from 58-60°C, resulting in amplicons of 55-113 bp with melting temperatures from 79-85°C. Primer efficiency and specificity was assessed for all pairs of primers by the analysis of calibration curves and melting profiles, respectively, which resulted in efficiencies of approximately 100%, all within a 12% interval. For normalization, two different genes were used independently as reference: LmjF.34.3440 (I2 pair) was used for chromosome 34 mapping, and LmjF.36.1980 (equivalent to LmxM.36.1980 in *L. mexicana*; R3 pair) was used for the fusion chromosomes mapping. For each pair of primers, triplicates of each sample (early S and G2 phases) were run per plate (MicroAmp® Optical 96-well Reaction Plate, Life Technologies), which were sealed with MicroAmp® clear adhesive film (Life Technologies). SYBR Select Master Mix (Life Technologies) was used, together with 400 nM of primers (Eurofins MWG Operon, Ebersberg, Germany) and 0.01 ng of sample gDNA, to a total of 20 µl per reaction. All

experiments were run in a 7500 Real Time PCR system (Applied Biosystems), using the following PCR cycling conditions: 50 °C for 2 min and 95 °C for 2 min, followed by 40 cycles of 95 °C for 15 sec, 59 °C for 15 sec, and 72 °C for 1 min. Fluorescence intensity data was collected at the end of the extension step (72 °C for 1 min). The resulting fluorescence intensity data were then exported to Excel, and analysed by relative quantification using the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001), using the G2 phase sample as the calibrator. Graphs were generated using Prism 6 (GraphPad software Inc.).

Table 2-10. Primers used for MFA-qPCR analysis.

Number	Primer Name	Gene ID	Sense	Sequence	Chr. region
C 222	LmjF_A1_fwd	LmjF.34.0520	Fwd	CAACCCTCTCAGATGGGCATT	A
C 223	LmjF_A1_rev		Rev	GTGACCCGTCTTCGTCATCTC	
C 228	LmjF_B1_fwd	LmjF.34.0670	Fwd	TGGCTGGGTGTGTACATTTCC	B
C 229	LmjF_B1_rev		Rev	CTTGGTACTGCGCAAAATGGT	
C 230	LmjF_B2_fwd	LmjF.34.0710	Fwd	CAGGCGAAGTTAGTGCGAAAG	B
C 231	LmjF_B2_rev		Rev	GTTCTCCGGTGACAAACTGATG	
C 232	LmjF_C1_fwd	LmjF.34.1000	Fwd	AAGCAGTGCACACCCCACTAC	C
C 233	LmjF_C1_rev		Rev	AAGAAGTTGGCCCGCTTCA	
C 266	LmjF_L2_fwd	LmjF.34.1500	Fwd	CGCTCCGTAGATCTGAGCATT	L
C 267	LmjF_L2_rev		Rev	CTGTGGGTTGCTCGTAATAGCA	
C 234	LmjF_D1_fwd	LmjF.34.2020	Fwd	CGCCGTTCCCATTTACGT	D
C 235	LmjF_D1_rev		Rev	TGTTCCAGCTCATGGACATAGC	
C 238	LmjF_E1_fwd	LmjF.34.2190	Fwd	GGGCATGTTTGGCGACAT	E
C 239	LmjF_E1_rev		Rev	GAAATGCTCTGCTCGCTGATG	
C 240	LmjF_F1_fwd	LmjF.34.2450	Fwd	TGACCGAGTGATGGAGTCCTT	F
C 241	LmjF_F1_rev		Rev	CCTTGATCACTGTGCCATCCT	
C 268	LmjF_G3_fwd	LmjF.34.2520	Fwd	CCATCTGCGAGGGTGTGATACT	G
C 269	LmjF_G3_rev		Rev	GATCAAGTCGCTAAAGCAGTCCTT	
C 252	LmjF_H3_fwd	LmjF.34.2670	Fwd	AGGGTCTGCAGTACGCTGTCTT	H
C 253	LmjF_H3_rev		Rev	GCCATGTGGTTGAACTTGAGGTA	
C 260	LmjF_J2_fwd	LmjF.34.3750	Fwd	TTTGTGTGCTGGCAGCTACAC	J
C 261	LmjF_J2_rev		Rev	GCAACTGACGCCTTCCACAT	
C 262	LmjF_K1_fwd	LmjF.34.3930	Fwd	CCGCGAACTGCAAAGTACGT	K
C 263	LmjF_K1_rev		Rev	TTGTCGACCGTCGTGTGAAT	
C 256	LmjF_I2_fwd	LmjF.34.3440	Fwd	GGCTGGCAACATGAAGTACGT	I
C 257	LmjF_I2_rev		Rev	CCATGTCAGACTGTCCCTTGAC	
C 276	LPhu_N1_Fwd	LmjF.29.0030	Fwd	CGATGTCGGGACTTACGTAAAGT	N
C 277	LPhu_N1_Rev		Rev	TCCACAGCGTGTATCCTTTCG	
C 280	LPhu_N3_Fwd	LmjF.08.0090	Fwd	CAGCCTCTACCGCGTCTTTC	N
C 281	LPhu_N3_Rev		Rev	TCTCCTTCAGTCGGACGTATGTC	
C 270	LPhu_M2_Fwd	LmjF.29.0810	Fwd	CATCATGATCAAGACCCTCGAGTA	M
C 271	LPhu_M2_Rev		Rev	GGCGACTTCGCAGCTTCTC	
C 274	LPhu_M4_Fwd	LmjF.29.0930	Fwd	ACTCGACTGCGCCTCATTG	M
C 275	LPhu_M4_Rev		Rev	TGACAGGAGAGGGACGAAGAG	
C 302	LPhu_O2_Fwd	LmjF.29.2060	Fwd	AGCCACCTTTAACGCCATTGT	O
C 303	LPhu_O2_Rev		Rev	GGAACAGGAGGCCATCGAA	
C 306	LPhu_P2_Fwd	LmjF.08.0260	Fwd	CAACAAGTCGGCCACTTACAAG	P
C 307	LPhu_P2_Rev		Rev	CGCCACATCTGCCATGAG	
C 308	LPhu_P3_Fwd	LmjF.08.0360	Fwd	CCCTCCGCCACAATGAG	P
C 309	LPhu_P3_Rev		Rev	TTCGCCCACGCTAGTATCG	
C 282	LPhu_Q1_Fwd	LmjF.08.1000	Fwd	GGAACCTGACCTACCCCTTCTC	Q

C 283	LPhu_Q1_Rev		Rev	GTCGAAGTTGAAGACGTTGTTGA	
C 288	LPhu_R2_Fwd	LmjF.36.1900	Fwd	CCACACACTCGCCTCTTACTACA	R
C 289	LPhu_R2_Rev		Rev	AGCTCAGGGTCACGCAAAAG	
C 310	LPhu_S1_Fwd	LmjF.36.2830	Fwd	TGCGGAGCGCAAGAATG	S
C 311	LPhu_S1_Rev		Rev	GGCGAGGCGGAACATCT	
C 314	LPhu_S3_Fwd	LmjF.36.3000	Fwd	TGTGGGAGGAAACAATCAGCTT	S
C 315	LPhu_S3_Rev		Rev	GTGGCGGAGAGGAAAACGTA	
C 292	LPhu_T1_Fwd	LmjF.36.3790	Fwd	GCACACACGGTACTGCTTCAA	T
C 293	LPhu_T1_Rev		Rev	CACGGGCTAAGCGCACTAG	
C 294	LPhu_U1_Fwd	LmjF.20.0705	Fwd	TGGGCTAGCTCCTTCTTTCACT	U
C 295	LPhu_U1_Rev		Rev	TTCGTCCTTGAGCTTGTACTTGAC	
C 296	LPhu_V1_Fwd	LmjF.20.1210	Fwd	GTCGCCGCAACCAGTACAT	V
C 297	LPhu_V1_Rev		Rev	CCGGAGAAGTGCTGGTACA	
C 298	LPhu_W1_Fwd	LmjF.20.1530	Fwd	TCCGCTGTTTGACGTGTATAGC	W
C 299	LPhu_W1_Rev		Rev	TCAACTCCTCCACCTTGCATATC	
C 290	LPhu_R3_Fwd	LmjF.36.1980	Fwd	GAGGTTCATGAGCTTGGGTTTAA	R3
C 291	LPhu_R3_Rev		Rev	TGCAAGGGAACAGGTGGTTT	

2.9 Analysis of Origins of DNA Replication Features in *Leishmania* and *Trypanosoma brucei*

2.9.1 Strand switch region size analysis

Strand switch regions (SSRs) containing origins were identified, and viewed on 'genome browser' using TriTrypDB version 8.0 (http://tritrypdb.org/tritrypdb/). The distance between the two most proximal genes within the SSR (divergent, convergent, or head-to-tail) was measured by subtracting the coordinates of the stop or start codon of the gene to the left of the SSR from the coordinates of the stop or start codon of gene on the right. The same was performed for other SSRs, where origins were not detected. The size of the distance between genes at the SSRs was then plotted onto a vertical scatter plot using Prism 6 (GraphPad software Inc.).

2.10 Statistical Analysis

All statistical analyses were performed using Prism 6 (GraphPad software Inc.). The statistical tests used (Cann, 2003) are described in the figure legend of the corresponding graph, and the reasons for their specific application are explained in the main text of Chapters 3, 4 and 5.

3 Analysis of putative Origin Recognition Complex factors in *T. brucei* procyclic form cells

3.1 Introduction

Until recently, very little was known about how nuclear DNA replication is initiated in *T. brucei*. Sequencing of the *T. brucei* genome (El-Sayed *et al.*, 2005a) has been invaluable in understanding how this process might be carried out and, together with various studies in the last few years, has started to suggest a surprising scenario: it appears that *T. brucei* possesses a conserved molecular machinery acting in replication fork establishment and synthesis (El-Sayed *et al.*, 2005a; Tiengwe *et al.*, 2012b; Kaufmann *et al.*, 2012; Liu *et al.*, 2009; Kim *et al.*, 2013; Tiengwe *et al.*, 2013; Li, 2012), while components of the pre-replication complex (pre-RC) appear to be considerably diverged (Tiengwe *et al.*, 2012b; Tiengwe *et al.*, 2013; Dang and Li, 2011; Godoy *et al.*, 2009; El-Sayed *et al.*, 2005a).

Initial analysis of the kinetoplastid genomes (El-Sayed *et al.*, 2005a) was only able to identify a single putative ORC subunit, an orthologue of both the Orc1 subunit and Cdc6, as these two proteins are believed to be paralogues (Duncker *et al.*, 2009; Giraldo, 2003). The inability to identify the remaining subunits of ORC (Orc2-6), or Cdc6 and Cdt1, led to the suggestion that the *T. brucei* and related kinetoplastid initiation machineries could resemble that found in the majority of archaeal organisms (El-Sayed *et al.*, 2005a; Godoy *et al.*, 2009), where a single factor, orthologous to both eukaryotic Orc1 and Cdc6, fulfils the role of the 6-subunit eukaryotic ORC (reviewed in Lindas and Bernander, 2013). Consequently, the only identified *T. brucei* Orc-like factor was named TbORC1/CDC6 (Godoy *et al.*, 2009). This hypothesis was somewhat undermined by later studies that identified four further potential Orc-like factors, all of which have been shown to interact with TbORC1/CDC6 in procyclic form (PCF) cells (Dang and Li, 2011; Tiengwe *et al.*, 2012b). These factors include another putative Orc1-like factor, TbORC1B, which was identified through searching the parasite's genome using Orc1 protein sequences from various organisms as queries (Dang and Li, 2011), and a highly divergent putative orthologue of the Orc4 subunit, TbORC4, identified via immunoprecipitation (IP) of TbORC1/CDC6 (Tiengwe *et al.*, 2012b). The two other factors, Tb7980 and Tb3120, were also identified as a result of TbORC1/CDC6 IP, although no orthology could be found between these and eukaryotic ORC factors, perhaps suggesting kinetoplastid specificity (Tiengwe *et al.*, 2012b).

Sequence analysis of the above factors revealed great divergence from ORC proteins in other eukaryotes. DNA replication initiator factors, including the Orc1-5 subunits of ORC and Cdc6, are members of the DnaA/CDC6/ORC clade of the AAA+ superfamily of ATPases (Iyer *et al.*, 2004), although the Orc2 and Orc3 subunits appear to be more distantly related (Duncker *et al.*, 2009). Sequence analyses of TbORC1/CDC6 (Godoy *et al.*, 2009) and TbORC1B (Dang and Li, 2011) identified AAA+ domains, while examination of TbORC4 (Tiengwe *et al.*, 2012b) and Tb7980 (Tiengwe *et al.*, 2013) suggested potentially degenerate AAA+ domains. Scrutiny of the Tb3120 sequence however, did not reveal any AAA+ domain homology. Detailed and updated analysis of signature motifs in the predicted AAA+ domains of the *T. brucei* proteins is described below (section 3.2.3). Another common feature of some ORC subunits and Cdc6 is the presence of a C-terminal winged-helix domain (WHD), which is responsible for DNA binding (reviewed in Duncker *et al.*, 2009). This domain however, has only been predicted in TbORC1/CDC6, though with poor conservation (Godoy *et al.*, 2009). An N-terminal bromo-adjacent homology (BAH) domain is a signature domain of eukaryotic Orc1 subunits that mediates interaction with transcriptional silencing factors and chromatin (reviewed in Duncker *et al.*, 2009; Costa *et al.*, 2013). No such domain has been described in TbORC1/CDC6 or any of the other *T. brucei* proteins, including TbORC1B (Dang and Li, 2011).

Several studies have investigated the functions of TbORC1/CDC6 and it's interacting partners (Godoy *et al.*, 2009; Benmerzouga *et al.*, 2013; Tiengwe *et al.*, 2012b) by individual expression downregulation through inducible RNA interference (RNAi). In PCF cells, RNAi of each factors resulted in a very similar phenotype, in which cellular growth was impaired only after prolonged RNAi, and enucleated cells (0N1K, termed zoids) (Robinson *et al.*, 1995) accumulated in the cell population (Godoy *et al.*, 2009; Benmerzouga *et al.*, 2013; Tiengwe *et al.*, 2012b). To date, no RNAi of TbORC1B has been described. In addition, the role of only TbORC1/CDC6 in nuclear DNA replication has been assessed (Benmerzouga *et al.*, 2013): incorporation of bromo-5'-deoxyuridine (BrdU), a thymidine analogue, was shown to be reduced following RNAi. Like other eukaryotic Orc1 factors (Lygerou and Nurse, 1999; Tatsumi *et al.*, 2000; Liang and Stillman, 1997; Asano and Wharton, 1999), TbORC1/CDC6 was shown to localise, potentially as puncta, to the nucleus of PCF cells throughout the cell

cycle, during which it was confirmed to bind to chromatin (Godoy *et al.*, 2009). No such localisation has been described for any of the TbORC1/CDC6 interacting factors.

The above summary demonstrates that, although some light has been shed on DNA replication in *T. brucei*, many questions remain. Although TbORC1B, TbORC4, Tb7980 and Tb3120 have been shown to interact with TbORC1/CDC6 (Tiengwe *et al.*, 2012b; Dang and Li, 2011), clear evidence that these have replication-associated functions is lacking. In addition, though it has been suggested that all these factors may interact in an ORC-like complex (Li, 2012), presently, the available data is insufficient to allow such a conclusion to be reached (Tiengwe *et al.*, 2013). Moreover, it is also not known if further factors interact with TbORC1/CDC6, including in such a complex. For instance, although not published (Tiengwe *et al.*, 2012b), other hits were retrieved from the TbORC1/CDC6 IP assay that identified TbORC4, Tb7980 and Tb3120 as TbORC1/CDC6-interacting partners (Tiengwe, 2010). One of these was Tb927.6.1120 (here referred to as Tb1120), which was actually the 'top' hit obtained (14 peptides, comparing to the 10 peptides that identified TbORC4) in the assay but, due to the lack of discernible homology with any replication-related factor, it was not analysed further (Tiengwe, 2010).

In this chapter, several techniques are employed to try to answer the above questions and, ultimately, better understand how initiation of DNA replication is controlled in *T. brucei*. New analysis of the protein sequences of TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120 are described (section 3.2), the role of some of these factors in DNA replication is investigated and compared with TbORC1/CDC6 (section 3.3), as is localisation of these proteins throughout the cell cycle (section 3.5). Finally, the question of whether there is an ORC-like complex or in *T. brucei* is examined (section 3.6).

3.2 TbORC1/CDC6-binding partners are highly divergent from characterised eukaryotic ORC-subunits

Previous work has described domains and motifs present in the protein sequences of TbORC1/CDC6 and its interacting partners (Godoy *et al.*, 2009; Dang and Li, 2011; Tiengwe *et al.*, 2012b; Tiengwe *et al.*, 2013). Nevertheless, because most of the protein domain, motif and solved structure databases are updated frequently (Finn *et al.*, 2014; Roy *et al.*, 2010; Källberg *et al.*, 2012; Hunter *et al.*, 2012), we sought to re-examine the proteins' sequences, and combine information obtained through new online tools (described throughout this section) to ask if more insight into putative orthology and functions is discernible. The protein sequences of TbORC1/CDC6 (Tb927.11.7216), TbORC1B (Tb927.9.2030), TbORC4 (Tb927.10.13380), Tb7980 (Tb927.10.7980), Tb3120 (Tb927.9.4530) and Tb1120 (Tb927.6.1120) were obtained from TriTrypDB database (http://tritrypdb.org/tritrypdb/) and used independently as queries in the different search engines used.

3.2.1 Basic Local Alignment Search Tool (BLAST) search

An initial analysis was conducted using the standard protein-protein Basic Local Alignment Search Tool (BLAST) (blastp) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul *et al.*, 1990), in which each putative *T. brucei* Orc-like factor protein sequence was used as query against the non-redundant protein sequences database (nr) (Li *et al.*, 2013). The main purpose of this analysis was to identify similar protein sequences available in the database and, if possible, infer function or homology.

As expected, analysis of TbORC1/CDC6 and all its interacting factors retrieved orthologues in most other kinetoplastid organisms, including *Trypanosoma cruzi* and most of the *Leishmania* species (only *L. major* and *L. mexicana* represented in Table 3-1, together with two strains of *T. cruzi*). This suggests that these factors are conserved throughout the kinetoplastid grouping. Indeed, analysis of genomic localisation of the orthologous genes in the different organisms (*T. brucei*, *T. cruzi*, *L. major* and *L. mexicana*) reveals that these are, as expected, syntenic in each genome (shown in the appendices, Figure 7.1).

Table 3-1.Orthologues of *T. brucei* TbORC1/CDC6 and its interacting factors in other kinetoplastids.
Orthologues of TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120, in *L. major*, *L. mexicana*, and *T. cruzi* (both CL Brener Esmeraldo-like, CLB EL, and Non-Esmeraldo-like, CLB NEL, strains). Gene identification numbers as in TriTrypDB, version 8.1. database. E-values and percentage sequence identity as retrieved by blastp searches performed in November 2014.

blastp query	*L. major*		*L. mexicana*		*T. cruzi* CLB EL		*T. cruzi* CLB NEL	
TbORC1/CDC6	LmjF.28.0030		LmxM.28.0030		TcCLB.508239.10		TcCLB.511159.20	
	e-value	Identity	e-value	Identity	e-value	Identity	e-value	Identity
	4e-168	55%	2e-168	55%	0.0	79%	0.0	78%
TbORC1B	LmjF.26.2210		LmxM.26.2210		TcCLB.507939.14		TcCLB.508119.150	
	e-value	Identity	e-value	Identity	e-value	Identity	e-value	Identity
	3e-35	42%	1e-33	42%	3e-152	52%	0.0	55%
TbORC4	LmjF.18.0720		LmxM.18.0720		TcCLB.506357.20		TcCLB.511277.92	
	e-value	Identity	e-value	Identity	e-value	Identity	e-value	Identity
	1e-61	42%	8e-65	45%	0.0	54%	0.0	54%
Tb7980	LmjF.36.6700		LmxM.36.6700		-		TcCLB.506247.280	
	e-value	Identity	e-value	Identity	e-value	Identity	e-value	Identity
	3e-103	41%	5e-106	43%	-	-	0.0	66%
Tb3120	LmjF.01.0660		LmxM.01.0660		TcCLB.511585.90		TcCLB.510155.90	
	e-value	Identity	e-value	Identity	e-value	Identity	e-value	Identity
	6e-59	31%	2e-61	32%	0.0	48%	0.0	48%
Tb1120	LmjF.12.0180		Lmxm.12.0180		TcCLB.507603.170		TcCLB.509429.240	
	e-value	Identity	e-value	Identity	e-value	Identity	e-value	Identity
	3e-36	31%	4e-39	33%	2e-175	42%	2e-170	41%

TbORC1/CDC6 was the first initiator factor to be identified and characterised in trypanosomatids due to its similarity to other eukaryotes' Orc1 and Cdc6 protein sequences (El-Sayed *et al.*, 2005a; Godoy *et al.*, 2009). Therefore, it was not surprising that the TbORC1/CDC6 search retrieved a large number of Orc1 and Orc1-like proteins as main hits, with high confidence (the e-values of the first 1000 hits ranged from 3e-154 to 3e-07). Due to the high number of hits (>2000), only a small sample of these is represented in Table 3-2. Amongst these were several Orc1 factors that have been experimentally characterised, including the protein from *Saccharomyces cerevisiae* (Liang and Stillman, 1997), *Schizosaccharomyces pombe* (Lygerou and Nurse, 1999), the domestic mouse (Zisimopoulou *et al.*, 1998), humans (Tatsumi *et al.*, 2000) and *Caenorhabditis elegans* (Sonneville *et al.*, 2012) (Table 3-2). As expected, since Orc1 and Cdc6 proteins have long been suggested to be paralogues (Giraldo, 2003), several Cdc6 and predicted Cdc6-like proteins were also identified in the search, although not as strongly as their Orc1 counterparts, but still with considerable confidence (e.g. Cdc6 factor of *Arabidopsis thaliana*, e-value of 2e-11). Together, these observations corroborate previous analysis (El-Sayed *et al.*, 2005a; Godoy *et al.*, 2009) and strengthen the assumption that TbORC1/CDC6 is most likely an Orc1-like factor. Some archaeal Orc1/Cdc6 proteins (Cdc6 from

Sulfolobales archeon; Cdc6 from *Sulfolobus acidocaldarius*; Cdc6 from *Sulfolobus solfataricus*; and Orc1 from *Aeropyrum pernix*) were also recovered (not shown), though with poorer e-values (for e.g. from 7e-07 to 0.39) when compared with most the eukaryote Orc1 and Cdc6 'hits'. Whether kinetoplastid ORC1/CDC6 provides both Orc1 and Cdc6 functions therefore remains unclear, despite experimental evidence showing that it can complement yeast Cdc6 mutants in culture (Godoy *et al.*, 2009).

Analysis of TbORC1B resulted in a smaller number of hits, and all with less confidence (the lowest e-value being of only 8e-04) than the results obtained for TbORC1/CDC6 (Table 3-2). Although TbORC1B was identified as being an Orc1-related protein (Dang and Li, 2011), and blastp analysis using human and yeast Orc1 proteins to interrogate the TriTrypDB database identify TbORC1B as a hit (data not shown), the TbORC1B search retrieved mainly Cdc6 or Cdc6-like hits (Table 3-2), including the characterised mouse and human proteins, with only a single Orc1 candidate recovered. Whether this suggests that TbORC1B is more related to Cdc6 than Orc1 remains unclear, however, in particular because the high e-values of nearly all hits provide low confidence regarding homology with the retrieved sequences. Thus, experimental evidence will be crucial to evaluate TbORC1B's function. This is reinforced by the observation that other replication-associated proteins were seen as hits (albeit with high e-values of between 0.058 and 2.1), including a 'MCM loader', a replication factor C subunit, and an Orc4 subunit protein (Table 3-2). All these proteins are predicted to possess an AAA+ ATPase domain (identified through InterPro, not shown), which might account for their similarity with TbORC1B, which has been suggested to have such domain (Dang and Li, 2011; Tiengwe *et al.*, 2013).

TbORC4 analysis retrieved very few hits, all of which with very low confidence (e-values ranging from 0.31 to 4.6) (Table 3-2). However, with one exception, all these hits are annotated as Orc4 subunits, thus corroborating previous observations (Tiengwe *et al.*, 2012b), and suggesting that it is most likely a highly divergent Orc4-like protein. Indeed, sequence alignments of TbORC4 with other eukaryotic Orc4 subunits (Tiengwe *et al.*, 2012b) support this likely ORC subunit homology (alignments shown in the appendices, Figure 7.4).

Table 3-2. List of sequences identified from blastp analysis.
Results from *Trypanosoma* and *Leishmania* species are not listed (shown in Table 3-1 instead), as well as unidentified proteins, or those labelled as predicted. Not all hits are represented for TbORC1/CDC6 and TbORC1B, as there were too many (>2000 and >350, respectively) and therefore only some relevant examples were selected. Accession numbers were retrieved from the NCBI database as of November 2014.

Query	blastp hits	Accession number	Species	e-value	Identity
TbORC1/CDC6	Orc1 subunit	EPY20124.1	*Strigomonas culicis*	3e-154	52%
	Orc1 subunit	EPY39172.1	*Angomonas deanei*	1e-67	47%
	Orc1-like subunit	XP_003079059.1	*Ostreococcus tauri*	4e-18	29%
	Orc1 subunit	NP_567440.1	*Arabidopsis thaliana*	9e-18	27%
	Orc1 subunit	EPY51628.1	*Schizosaccharomyces cryophilus*	1e-17	27%
	Orc1 subunit	NP_477303.1	*Drosophila melanogaster*	2e-17	29%
	Orc1 subunit	XP_004224191.1	*Plasmodium cynomolgi*	4e-17	26%
	Orc1 subunit	XP_001616506.1	*Plasmodium vivax*	1e-16	26%
	Orc1 subunit	NP_001014918.1	*Bos taurus*	6e-15	28%
	Orc1 subunit	XP_678269.1	*Plasmodium berghei* ANKA	2e-14	27%
	Orc1 subunit	XP_001350439.1	*Plasmodium falciparum* 3D7	4e-14	28%
	Orc1 subunit	NP_499347.1	*Caenorhabditis elegans*	5e-13	27%
	Orc1 subunit	NP_596060.1	*Schizosaccharomyces pombe*	6e-17	27%
	HsORC1 subunit	AAC50325.1	*Homo sapiens*	9e-12	27%
	Cdc6	NP_172207.2	*Arabidopsis thaliana*	2e-11	24%
	Orc1 subunit	NP_035145.2	*Mus musculus*	1e-10	26%
	Orc1 subunits	NP_013646.1	*Saccharomyces cerevisiae*	6e-07	24%
TbORC1B	Cdc protein	EFQ26461.1	*Colletotrichum graminicola*	1e-01	29%
	Cdc6	GAA29103.2	*Chonorchis sinensis*	8e-04	25%
	Cdc protein	XP_003005252.1	*Verticillium alfalfae*	0.006	27%
	Cdc6 homolog isoform a	NP_035929.1	*Mus musculus*	0.088	29%
	Cdc6 homolog isoform b	NP_001020950.1		0.089	
	DNA Replication factor C, large subunit	KEQ79100.1	*Aureobasidium pullulans*	0.058	29%
	Cdc6 homolog	NP_001179336.1	*Bos taurus*	0.14	30%
	Cdc6-related protein	NP_001081844	*Xenopus laevis*	0.69	23%
	Cdc6 homolog	NP_001245.1	*Homo sapiens*	0.75	29%
	Orc1 subunit	XP_003868271.1	*Candida orthopsilosis*	1.1	22%
	MCM loader	EEB07000.2	*Schizosaccharomyces japonicus*	2.0	29%
	Orc4 subunit	EIT78085.1	*Aspergillus oryzae*	2.1	39%
TbORC4	Orc4 subunit	AAD39473.1	*Drosophila melanogaster*	0.31	34%
	Predicted: Orc4-like	XP_005179312.1	*Musca domestica*	0.50	30%
	Predicted: Orc4-like	XP_004535926.1	*Ceratitis capitata*	0.68	33%
	ORC subunit	XP_001653985.1	*Aedes aegypti*	4.6	31%

Table 3-2. *(continued).*

Query	blastp hits	Accession number	Species	e-value	Identity
Tb7980	DNA helicase	WP_006043754.1	*Prevotella pallens*	0.16	35%
	ABC transporter	WP_030305899.1	*Streptomyces* sp.	0.95	43%
	ABC transporter	WP_012854200.1	*Thermomonospora curvata*	1.9	42%
	DNA polymerase III	WP_020966355.1	*Treponema pedis*	3.5	22%
	Cell division protein	WP_033044043.1	*Pseudomonas putida*	5.1	28%

All hits that resulted from the Tb7980 BLAST analysis presented very high e-values (0.16-5.1), and thus low confidence (Table 3-2). Nevertheless, it is intriguing that all these hits were from bacteria, and included a DNA helicase, a DNA polymerase III and a 'cell division protein', besides two ABC transporter hits (Table 3-2). All of these factors possess AAA+ ATPase domains, similar to the Orc1-5 proteins (Iyer *et al.*, 2004), supporting the suggestion that Tb7980 might possess such a domain (Tiengwe *et al.*, 2013).

Analysis of both Tb3120 and Tb1120 retrieved only kinetoplastid hits (Table 3-1), all of which corresponded to hypothetical and unidentified proteins, possibly suggesting that these might be kinetoplastid-specific factors, as it has been proposed in the case of Tb3120 (Tiengwe *et al.*, 2012b; Tiengwe *et al.*, 2013).

3.2.2 Structure Prediction analysis

In order to complement the information from the blastp analysis, TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120 protein sequence were further analysed through the protein structure and function prediction online platforms RaptorX (Källberg *et al.*, 2012) and I-TASSER (Zhang, 2008a; Roy *et al.*, 2010). Both platforms rely on a composite approach, in which a combination of various prediction techniques is used. For example, analysis using I-TASSER results from the combination of various prediction techniques such as comparative modelling (e.g. blastp, as discussed above) and threading methods, both template-based modelling systems, as well as *ab initio* modelling, a template-free strategy where no structure homologous information is used (Källberg *et al.*, 2012; Roy *et al.*, 2010; Zhang, 2008b). For the present analysis, both RaptorX and I-Tasser default parameters were used. In both online platforms, experimentally resolved

protein structures archived in the Protein Data Bank (PDB) (Berman *et al.*, 2000) library were used for template search and modelling.

Three solved archaeal protein structures were mainly selected by both RaptorX and I-TASSER platforms to model the structures of TbORC1/CDC6, TbORC1B, TbORC4, and Tb7980: Orc1 from *Aeropyrum pernix* (PDB identification number 2V1U) (Gaudier *et al.*, 2007), the Orc1 initiator factor of *Sulfolobus solfataricus* (2QBY) (Dueber *et al.*, 2007), and the Orc1 protein of *Pyrobaculum aerophilum* (1FNN) (Liu *et al.*, 2000) (Table 3-3). Very recently, the crystal structure of the *D. melanogaster* ORC complex has been solved (Bleichert *et al.*, 2015). It will be necessary to repeat this analysis in the future once the structures are incorporated into the database, as it is possible that these would be more appropriate for modelling the *T. brucei* proteins. Nevertheless, the use of archaeal Orc1 and Orc1-like initiator factors strengthens the idea that these proteins might indeed be Orc-like proteins. Other protein structures were also selected as models for the different *T. brucei* proteins, although these were not common between the four factors (not shown). Nevertheless, these hits were generally involved in ATP-binding related functions, and most likely possess AAA+ ATPase domains, consistent with the potential presence of these domains in TbORC1/CDC6, TbORC1B, TbORC4 and Tb7980.

In contrast to the abovementioned proteins, and reflecting the results from the BLAST analysis, little information was retrieved from modelling both Tb3120 and Tb1120. The confidence in the models was poor, and both RaptorX and I-TASSER platforms did not use replication-related proteins for the modelling of either factor. In the case of Tb3120, nucleoporins, hydrolases and virus capsid proteins were mainly used for modelling (Table 3-3), while for Tb1120, transportins were predominantly used (Table 3-3). These outputs reinforce the lack of obvious Orc-like homology in these proteins. Nonetheless, potentially interesting hits in the Tb1120 search included a cohesin subunit (4PK7) (Hara *et al.*, 2014), which is involved in sister chromatid cohesion after DNA replication is complete, and an SYS-1 protein (3C2G) (Liu *et al.*, 2008), which is involved in the positive regulation of transcription from RNA polymerase II promoters. Whether these hits provide functional clues for Tb1120 is not known, but might imply cell cycle- or DNA replication-related roles.

Table 3-3. Top hits of solved protein structures used by RaptorX and I-TASSER search engines to model TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120.
Main hits obtained from structure prediction modelling platforms RaptorX and I-TASSER, using each of *T. brucei* Orc-like proteins as queries. Not all hits are shown. Analysis performed in November 2014. ID number from both PDB and UniProt (http://www.uniprot.org/) were retrieved in November 2014.

Query	PDB	UniProt	Protein name	Species
TbORC1/CDC6, TbORC1B, TbORC4, Tb7980	2V1U	Q9YEV6	Orc1-type protein	*Aeropyrum pernix*
	2QBY	Q980N4	Orc1-type protein	*Sulfolobus solfataricus*
	1FNN	Q8ZYK1	Orc1-type protein	*Pyrobaculum aerophilum*
Tb3120	1IUN	P96965	Hydrolase	*Pseudomonas flurescens*
	4KF7	G2QD05	Nup188 nucleoporin	*Myceliophthora thermophila*
	2CSE	P11077	Outer capsid protein mu-1	*Mammalian orthoreovirus* 1
Tb1120	4PK7	Q8N3U4	Cohesin subunit SA-2	*Homo sapiens*
	4C0O	Q9YAL0	Transportin-3	*Homo sapiens*
	3C2G	Q9XVI2	Protein SYS-1	*Caenorhabditis elegans*
	1QBK	Q92973	Transportin-1	*Homo sapiens*

3.2.3 Alignments, and domain and motif searches

To build further on the analyses performed in the previous sections, the protein sequences of TbORC1/CDC6 and its putative interacting partners were further analysed through domain and motif online search tools, in order to build up schematic representations of the functional domains of the proteins. For initial domain searches, all sequences were submitted for analysis through both the Pfam (http://pfam.xfam.org/) (Finn *et al.*, 2014) and InterPro (http://www.ebi.ac.uk/interpro/) (Hunter *et al.*, 2012) databases. While Pfam is mainly a database of curated protein families and clans, defined by probabilistic models used to infer homology between protein sequences (Finn *et al.*, 2014), InterPro retrieves and combines information from various such databases (including Pfam and PHANTER), and therefore may provide a wider set of information on protein families and domains within the protein sequences (Hunter *et al.*, 2012). To try and identify more specific sequence characteristics, and complement the domain search analysis, the sequences were also analysed through a series of motif and protein signature databases: the GenomeNet Motif search tool (http://www.genome.jp/tools/motif/) (Kyoto University Bioinformatics Centre), the MyHits© Motif Scan tool (http://myhits.isb-sib.ch/) (Pagni *et al.*, 2007), and the ScanProsite online tool (http://prosite.expasy.org/scanprosite/) (Sigrist *et al.*, 2013). Finally, the *T. brucei* protein sequences were aligned with protein sequences from Orc factors of other eukaryotes, in order to identify further signatures, domains and motifs that might have not been identified through the search engines described above

(all alignments are shown in the appendices, section 7.2). The combined output of these analyses is summarised in Figure 3.1.

Analysis of TbORC1/CDC6 protein sequence with Pfam and InterPro retrieved a clear AAA+ ATPase domain in the protein's N-terminus (57-202 aa) (Figure 3.1), as well as a Cdc6-related family domain that encompassed most of the protein's length (20-383 aa, not represented in Figure 3.1). However, and as previously described, TbORC1/CDC6 lacks a detectable BAH domain (Tiengwe *et al.*, 2013; Godoy *et al.*, 2009), which appears to be found in all Orc1 proteins from other eukaryotic organisms (reviewed in Duncker *et al.*, 2009; Costa *et al.*, 2013). Motif searches combined with sequence alignment of TbORC1/CDC6 with Orc1 subunits from eukaryotes (detailed in the appendices, Figure 7.2), allowed the identification of the Walker A, Walker B and arginine finger motifs characteristic of the AAA+ ATPase domain, as well as a putative nuclear localisation signal (NLS), as described previously (Godoy *et al.*, 2009; Dang and Li, 2011) (Figure 3.1). None of the online tools were able to identify the WHD, which is characteristic of Orc1-5 proteins and mediates DNA binding (Duncker *et al.*, 2009). Nonetheless, the region of TbORC1/CDC6 that aligned with the WHD regions annotated in the other Orc1 proteins (303-433 aa region, alignment shown in detail in Figure 7.2) localised in the C-terminus, which is a region previously suggested to harbour the WHD of TbORC1/CDC6 (Godoy *et al.*, 2009)(Figure 3.1). Alignment of TbORC1/CDC6 with other Orc1 subunits revealed further short sequences that are highly conserved between the aligned proteins (Figure 3.1), but to which specific activities have not been detailed, and whose relevance to protein structure or function is, therefore, unknown.

TbORC1B domain analysis resembled that of TbORC1/CDC6: a clear AAA+ ATPase domain was easily identified in the N-terminal region of the protein (7-270 aa, Figure 3.1), as well as a Cdc6-related family domain throughout the protein length (2-539 aa, not depicted in Figure 3.1). Motif search tools retrieved a potential bipartite NLS (RPTKR), and sequence alignment of TbORC1B with Cdc6 proteins (from model eukaryotes and *Clonorchis sinenis*, the first hit from the BLAST search, shown in the appendices, Figure 7.3) allowed the identification of likely Walker A and Walker B motifs, consistent with those suggested previously (Dang and Li, 2011), as well as the identification of a series of potential

conserved motifs to which, like for TbORC1/CDC6, it was not possible to allocate a specific function or motif name (Figure 3.1). Amongst these motifs, it is worth noting that the FE(A/P)Y and the GDVR motifs are common between TbORC1B and TbORC1/CDC6 (Figure 3.1). Again, like for TbORC1/CDC6, the search tools were not able to identify a WHD, but such a motif could be accommodated in the C-terminal region of the protein (419-601 aa, Figure 3.1), based on the alignment of TbORC1B with Cdc6 protein sequences in which the WHD has been annotated (shown in Figure 7.3).

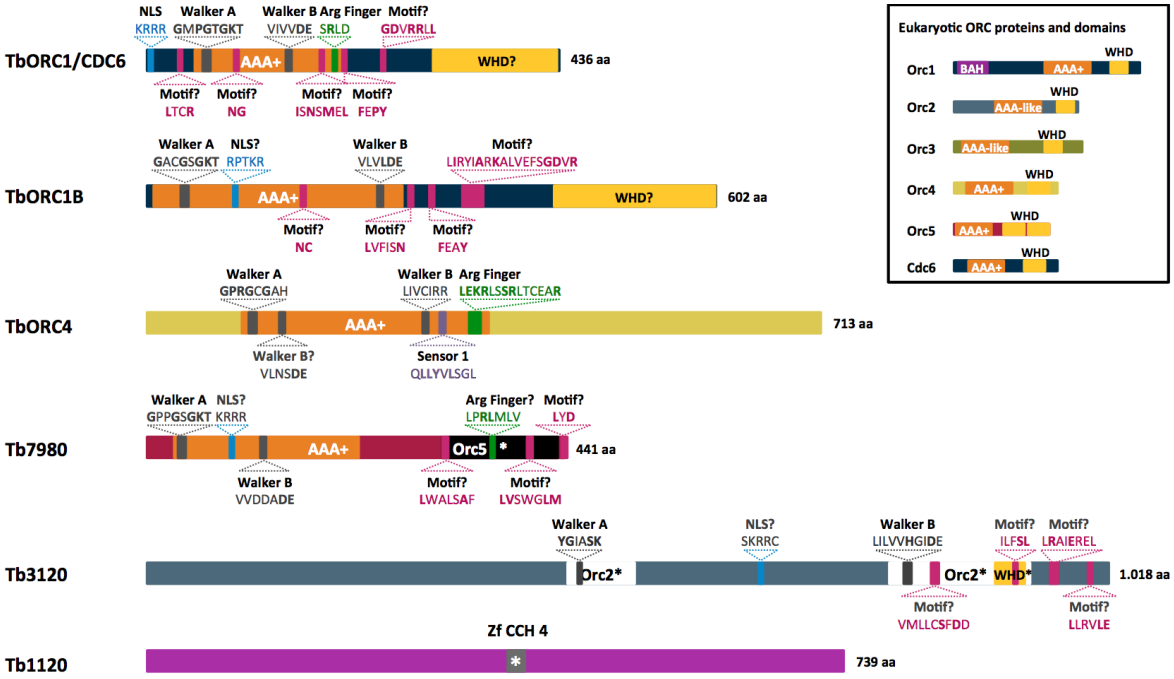


Figure 3.1. Schematic representation of TbORC1/CDC6 and other putative ORC factors in *T. brucei*.
TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120 proteins of *T. brucei* are schematically represented, with the identified domains highlighted in a colour-coded fashion: AAA+ ATPase domains in orange, winged-helix domains (WHD) in yellow, nuclear localisation signal (NLS) in light blue, arginine finger (Arg Finger) in green, Orc5 domain in black, and Orc2 domain in white; further potential motifs, and their sequences, are depicted in pink. Each factor is represented in a colour corresponding to the potential orthologue in other model eukaryotes, which are represented in the insert box together with their characteristic domains (adapted from (Duncker *et al.*, 2009). (*) marks domains that have been rated as non-significant; (?) represents that the confidence in the motif is not high. Predicted lengths of the proteins (in amino acids, aa) are shown. The overall figure has been adapted from (Tiengwe *et al.*, 2013), © 2013 Elsevier Ltd. All rights reserved, to which new information (described in the main text) has been added.

Searches through both Pfam and InterPro, as well as through the motif search tools, were unable to identify any domains or motifs in TbORC4. The AAA+ ATPase domain was thus mapped through the alignment of TbORC4 with Orc4

subunits of other eukaryotes (Figure 3.1, alignments shown in the appendices, Figure 7.4), and appears to include both the Walker A and Walker B motifs, as well as the sensor 1 and arginine finger motifs (Tiengwe *et al.*, 2012b). Nevertheless, in the TbORC4 protein sequence, both the Walker A and Walker B motifs fail to possess the characteristic amino acids that usually define these motifs (the $GX_4GK[S/T]$[d] of the Walker A motif and the hhhh[D/E][e] of the Walker B); however, it is now recognised, as more protein structures are solved and made available in databases, that perhaps these motifs are not as conserved as previously thought (Matte and Delbaere, 2010). Still, the TbORC4 Walker B motif identified from the alignment appears to be highly divergent and possibly degenerate (Tiengwe *et al.*, 2012b), since the sequence LIVCIRR contrasts with the apparently highly conserved VIFILDE sequence found in all other analysed Orc4 subunits (Figure 7.4). However, re-analysis of the TbORC4 protein sequence revealed a VLNSDE sequence, which possesses the DE amino acid combination, localises within the putative AAA+ ATPase domain and is positioned downstream of the Walker A motif, but whether this is the true Walker B motif is not known. In contrast with TbORC1/CDC6 and TbORC1B, it was not possible to infer the presence of a WHD in the TbORC4 protein sequence, solely because domain searches of the model eukaryotes Orc4 subunits used in the alignment failed to identify such a domain (not shown).

Analysis of Tb7980 through the InterPro database identified an AAA+ ATPase domain in the protein's N-terminus (14-128 aa), which contains a putative Walker A motif (GPPGSGKT), identifiable through the ScanProsite online tool. In addition, a bipartite NLS profile (77-91 aa) that includes a KRRR motif (like in TbORC1/CDC6) was also identified by searches both on ScanProsite and MotifScan online tools (Figure 3.1). Analysis of Tb7980 using Pfam version 27.0 retrieved no results, but searches performed with previous versions of the Pfam database retrieved both the AAA+ ATPase domain (similar to the one identified by InterPro) and an Orc5 domain, although the latter was categorised as non-significant (Figure 3.1) (Tiengwe *et al.*, 2013). This difference may result from the fact that in previous versions of Pfam it was possible to obtain results comprising contextual domain-hits, meaning that it was possible to identify weak

[d] X refers to any amino acid.

[e] h refers to any hydrophobic amino acid.

domain hits that fell below the gathering threshold but were supported by surrounding domains or sequence features (Finn *et al.*, 2014). Unfortunately, this feature was abandoned in the newest version of Pfam (version 27.0). Nevertheless, the potential alignment of Tb7980 with eukaryotic Orc5 subunits was used to attempt to identify further motifs (shown in the appendices, Figure 7.5). This provided confirmation of the Walker A motif, and further identification of a putative Walker B motif (VVDDADE), as well as a potential arginine finger motif (LPRLMLV) (Figure 3.1). Moreover, three other sequence patterns, present within the putative Orc5 domain, and to which it was not possible to allocate a motif name or function, were also identified (Figure 3.1). Similar to TbORC4, it was not possible to infer the presence of a WHD because analysis of the eukaryotic Orc5 subunits used for the multi-sequence alignment failed to identify this domain, both on Pfam and InterPro databases (not shown). Despite this possible homology with Orc5 factors, it is not clear whether or not Tb7980 is indeed a *T. brucei* Orc5 subunit orthologue, and therefore will still be referred to in this work as Tb7980.

Similar to Tb7980, Pfam analysis of Tb3120 retrieved no results when using the newest version, although searches using previous versions of the database retrieved two regions of the protein as potential Orc2 domains, as well as a small region as a WHD, although both were classified as non-significant (Figure 3.1) (Tiengwe *et al.*, 2013). Further analysis through InterPro and the various motif search engines retrieved no results besides a potential bipartite NLS profile, SKRRC (Figure 3.1). Strikingly, there were no results suggesting the presence of an AAA+ ATPase domain or of the Walker A and Walker B motifs. Alignment of Tb3120 with Orc2 subunits from other eukaryotes (shown in the appendices, Figure 7.6) also failed to identify both the AAA+ ATPase and WHD domains. These observations may be consistent with previous work that showed that Orc2 and Orc3 subunits have highly divergent AAA+ ATPase domains and are therefore not easily mapped (reviewed in Costa *et al.*, 2013; Duncker *et al.*, 2009). Alignments of Orc2 subunits from various organisms revealed that though the Walker A and Walker B motifs in Orc2 do not follow the canonical sequences (mentioned above) present in other Orc proteins such as Orc1 and Orc5, these sequences are conserved among the different Orc2 subunits (Speck *et al.*, 2005): the Walker A motif sequence is X(L/F)(Y/F)GXGSKXX(L/F), while the Walker B

motif is defined by X(L/F)X(I/V)HN(L/I)D(G/S)X(M/S). The alignment of Tb3120 with the Orc2 subunits of model eukaryotes allowed the identification of such motifs in Tb3120 (Figure 3.1) (Figure 7.6), within the predicted (though non-significant) Orc2 domains. This strengthens the idea that Tb3120 might be an Orc-like factor, more precisely an Orc2-like protein, though more evidence, especially experimental, will be needed before labelling Tb3120 as being an Orc2 subunit. From the alignment it was also possible to identify various potential conserved motifs, the majority of which are present within the C-terminus region of Tb3120 (Figure 3.1), but their functional importance, if any, is unknown. It is worth noting that Tb3120 and orthologues in *T. cruzi* and *L. major* (referred here as Tc3120 and Lm3120, respectively, Figure 7.6) are considerably larger than any of the characterised eukaryotic Orc2 subunits (ranging from 576-620 aa) used in the alignment, with *T. brucei* and *T. cruzi* proteins being approximately double the size (1018 aa and 1049 aa, respectively), and the *L. major* almost three times the size (1487 aa). From the alignment it appears that, if Tb3120 and its homologues in *T. cruzi* and *L. major* are indeed Orc2-like factors, the kinetoplastid proteins have undergone an expansion in their N-terminus, in which it was not possible to identify any characteristic signatures or domains (although it appears to be conserved between the three organisms). This interpretation is nevertheless consistent with the observation that characterised Orc2 proteins (Figure 7.6) possess their Orc2 domains in the C-terminus region of the protein.

Domain analysis of Tb1120, which is confirmed to interact with TbORC1/CDC6 later in section 3.6.1, retrieved only a putative zinc knuckle (Pfam family zf-CCHC4, which was considered non-significant), using the GenomeNet Motif search engine and a previous version of Pfam (Figure 3.1). Therefore, Tb1120 remains the only potential TbORC1/CDC6-interacting protein identified so far that has no discernible homology, however weak, with a eukaryotic ORC subunit.

3.3 TbORC1/CDC6, TbORC4, Tb3120 and TbORC1B are involved in DNA replication

To facilitate gene function analysis through RNA interference (RNAi), both the procyclic and bloodstream form life cycle stages of *T. brucei* have been genetically modified to allow the conditional expression of double-stranded RNA (dsRNA), thereby permitting the inducible activation of the RNAi machinery against any gene of interest (reviewed in Kolev *et al.*, 2011). The PCF 29-13 cell line (Wirtz *et al.*, 1999) is derived from wild type *T. brucei* PCF cells of strain Lister 427 and constitutively expresses the bacteriophage T7 RNA polymerase (T7RNAP) as well as the *E. coli* tetracycline repressor (TetR), the expression of the latter being under the control of a T7 promoter (Wirtz *et al.*, 1999). The expression of the TetR allows the repression of promoters that are adjacent to a tetracycline operator (TetO). Upon addition of tetracycline to the cell, TetR is bound by the drug and removed from the TetO, allowing transcription of the DNA under the influence of the promoter. The 29-13 cell line has been widely used in conjunction with two different strategies to express gene-specific dsRNA. One involves the use of opposing (or 'head-to-head') promoters' constructs, in which a fragment of the target gene is inserted between two opposing T7 promoters that, in turn, are flanked by Tet operators (LaCount *et al.*, 2000; Wang *et al.*, 2000). The second entails the use of stem-loop vectors, in which two identical fragments of a gene of interest are cloned into the vector in opposite directions, one on each side of a "stuffer" fragment, thus allowing the expression of the dsRNA from a single promoter that is located upstream of a Tet operator (Shi *et al.*, 2000; Wang and Englund, 2001; Durand-Dubief *et al.*, 2003). In both cases, the constructs are integrated in the ribosomal RNA locus (*RRNA*) of the parasite's genome. Although both approaches have been shown to be efficient in downregulating the expression of targeted genes, the stem-loop approach seems to be more efficient (Durand-Dubief *et al.*, 2003), apparently because one molecule generates the dsRNA by self-folding (intramolecular dsRNA), while in the opposing promoters strategy the two RNA molecules need to anneal with each other intermolecularly (intermolecular dsRNA).

Studies investigating the effects of TbORC1/CDC6 downregulation by RNAi have used both the above strategies (Tiengwe *et al.*, 2012b; Godoy *et al.*, 2009; Benmerzouga *et al.*, 2013), while the effects of TbORC4, Tb3120 or Tb7980

silencing have been analysed using the opposing promoters approach alone (Tiengwe *et al.*, 2012b). It was notable that RNAi-induced silencing of the three TbORC1/CDC6 interacting factors in PCF cells resulted in just mild effects on cell growth and cell cycle progression, which were detected only around four days post-RNAi induction; in addition, attempts to evaluate if levels of BrdU incorporation changed after RNAi induction were inconclusive (Tiengwe *et al.*, 2012b). Given the apparently better performance of the stem-loop approach, together with the lack of any data on the effects of TbORC1B or Tb1120 silencing by inducible RNAi, we sought to generate RNAi cell lines, in the 29-13 cell line, for each of the putative *T. brucei* Orc-like factors using stem-loop constructs (below). The RNAi cell lines thus generated were analysed for cell growth, cell cycle, and DNA replication defects, as shown in the next sections. A cell line targeting TbORC1/CDC6 was also examined, allowing comparison with published data (Benmerzouga *et al.*, 2013), while the parental 29-13 cell line was used as a negative control.

3.3.1 RNAi system used and generation of the RNAi PCF cell lines

The stem-loop vector chosen was a variation of the pLew111 construct (Hoek *et al.*, 2000), containing the bleomycin resistance gene marker (*BLE*, which confers resistance to both the zeocin and phleomycin antibiotics). In this version, the pLew111 plasmid has been modified to include part of the human polo-like kinase 1 (PLK1) gene as a "stuffer" within the HindIII-BamHI linker (Figure 3.2; gift from Dr Calvin Tiengwe, Prof. Paul Englund's laboratory). This created two cloning sites, HindIII/XhoI and AflII/BamHI, which allow the sequential insertion, in a head-to-head configuration, of two identical PCR products, each flanked with one of the two different restriction site combinations. The region of each gene to be PCR-amplified (sizes between 400-600 bp), as well as the best pair of primers to use, was chosen using RNAit (http://trypanofan.bioc.cam.ac.uk/software/RNAit.html), an automated web-based tool for the selection of gene regions more suitable for RNAi targeting in *T. brucei* (Redmond *et al.*, 2003). Each gene sequence (retrieved from TriTrypDB) was used as query. In the case of *Tb7980*, the region suggested by RNAit contained a BamHI site and, therefore, a region downstream of this site was used as query. Sequences of restriction sites to allow cloning of the PCR products into the vector were added as follows (Figure 3.3): PCR product A (to be inserted in the sense orientation) was

amplified with a forward primer containing the HindIII restriction site, while the reverse primer possessed a XhoI site; PCR product B (to be cloned in the antisense orientation) was amplified with a forward primer with a BamHI site, while the reverse primer had a AflII restriction site. All gene-specific primers used and PCR conditions are described and shown in the materials and methods Chapter 2, Table 2-1. After PCR amplification, each PCR product was digested with the appropriate restriction enzyme combinations, purified, and ligated into the plasmid. For each gene, PCR product A was first cloned into pLew111, and PCR product B then cloned after validation, by restriction enzyme digestion, of the prior insertion. The final plasmids (Figure 3.4, A-F) were confirmed by restriction enzyme digestion using the two enzyme combinations used for cloning (Figure 3.4, G), as well as by sequencing. Plasmids targeting TbORC1/CDC6, TbORC4, Tb3120 and TbORC1B were produced by Dr Calvin Tiengwe (in Prof. Paul Englund's laboratory in Baltimore, U.S.A.), while the constructs targeting Tb7980 and Tb1120 were generated by me in Glasgow. Each plasmid was digested with NotI prior to transfection into the 29-13 cells, allowing homology-directed integration into the genome of the parasite at the *RRNA* spacer region. Transformant clones were selected with 10 $\mu g.ml^{-1}$ of zeocin, to which the *BLE* marker confers resistance. Growth of individual clones was examined in the presence of 2 $\mu g.ml^{-1}$ of tetracycline to look for impairment over a five-day time course (data not shown). Clones presenting a mild-severe growth phenotype were selected for further analysis. Overall, these included two clones, independently obtained, for TbORC1/CDC6, TbORC4, and Tb3120 (clones Cla were generated by Dr Calvin Tiengwe in Prof. Paul Englund's laboratory in Baltimore, U.S.A., while clones Clb were generated by me in Glasgow), and one clone for TbORC1B. Unfortunately, despite several attempts, no viable clones were retrieved for Tb7980 or Tb1120. In the case of Tb7980, a few clones were recovered after transfection, but could not be successfully maintained thereafter (data not shown). Several attempts were also made to generate RNAi cell lines that had one allele of the gene of interest endogenously 'tagged' with 12myc (construct cloning and cell line generation are detailed in sections 3.4.1 and 3.4.5), so that efficiency of protein loss after RNAi induction could be assessed by western blot, but this was only successful for TbORC4 (clone Clb, see below).

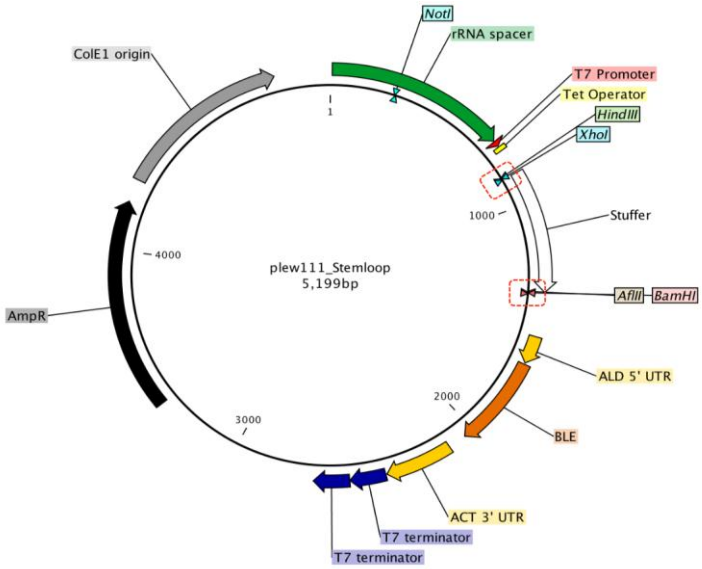


Figure 3.2. pLew111 construct.
Version of the pLew111 plasmid used as a backbone to generate the stem-loop RNAi constructs for TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120. In this version, the original pLew111 plasmid (Hoek *et al.*, 2000) has been modified to include a "stuffer" region, made up of a fragment of the human *PLK1* gene, between the HindIII and BamHI restriction sites (white arrow). The red dashed boxes highlight the cloning sites used to introduce the two PCR fragments of each gene, in order to generate a stem-loop dsRNA molecule. The Tet operator (yellow) is located downstream of the T7 promoter (red) to allow tetracycline-inducible expression of the fragment that will result in the dsRNA stem-loop. The bleomycin resistance marker (*BLE*, shown in orange) is flanked by *T. brucei* aldolase mRNA processing 5' and the actin mRNA processing 3' regions (represented in dark yellow). The bacterial drug resistance gene, against ampicillin, is shown as *amp*R (in black), and the bacterial origin in depicted as ColE1 origin (in grey). The *RRNA* spacer fragment (green) is shown with the NotI restriction site used to linearize the plasmid for transfection into the parasite.

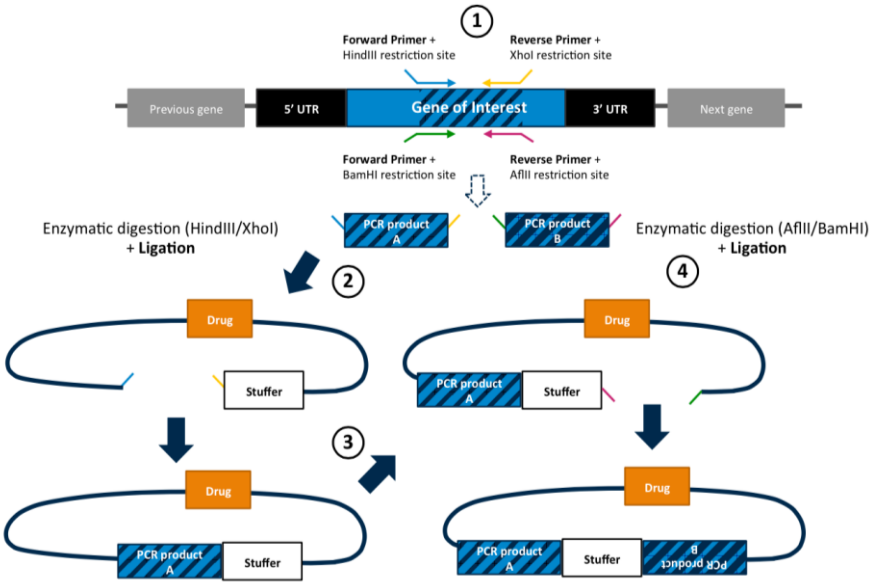


Figure 3.3. Stem-loop construct cloning strategy.
Graphical representation of the cloning process to generate the RNAi stem-loop constructs from the pLew111 vector. (1) Both PCR products A and B are amplified using the appropriate pair of primers. (2) The pLew111 vector and the PCR product A are digested with HindIII and XhoI, and ligated. (3) The resulting plasmid and the PCR product B are then digested with AflII and BamHI, and ligated (4), generating the final plasmid containing the two PCR products inserted in opposing orientations. Not all features of the plasmid are shown. Diagrams are not to scale.

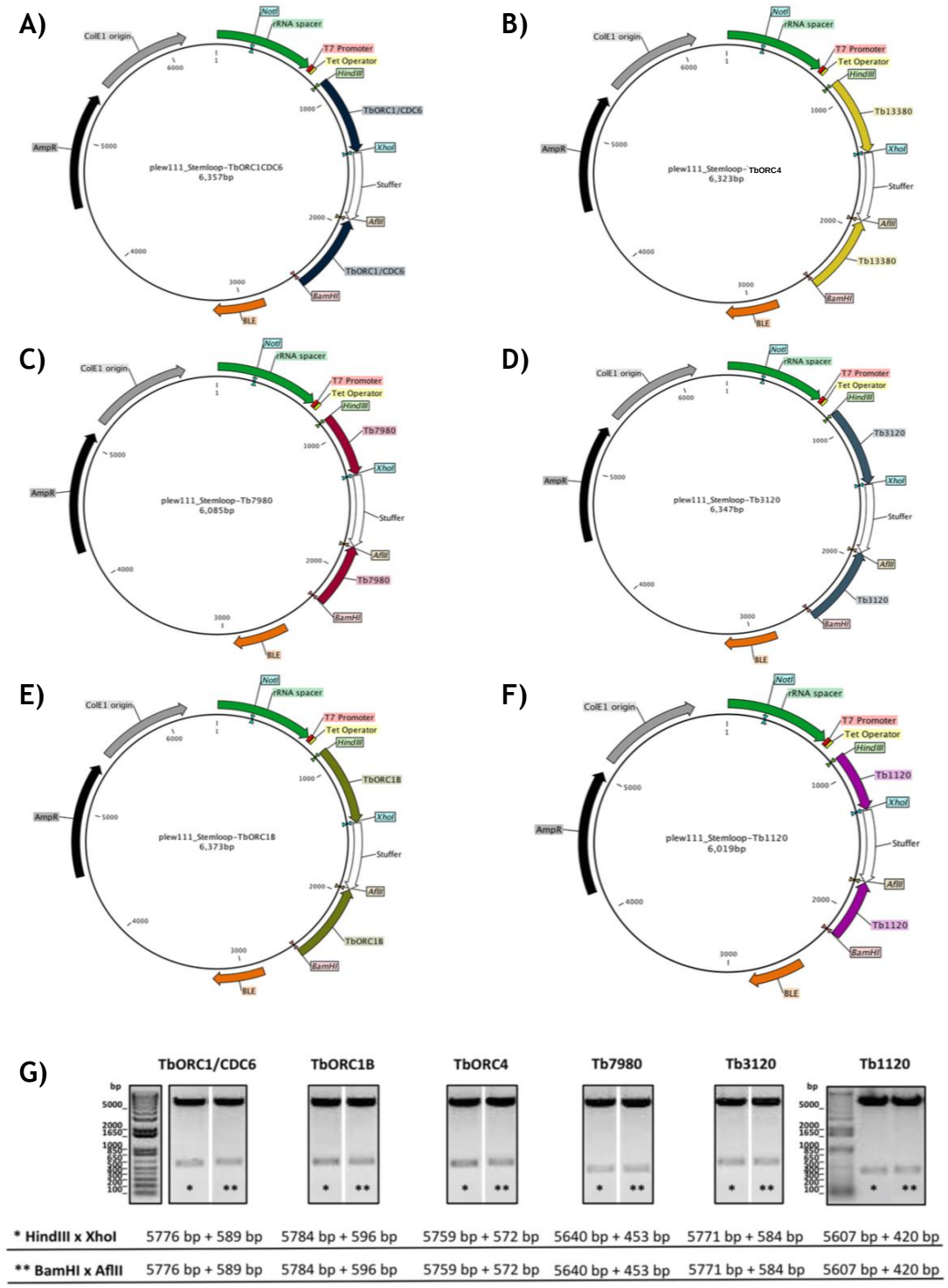


Figure 3.4. Plasmid maps and confirmation restriction digestions of the constructs used for RNAi of the different factors.

Plasmid maps of the constructs used for RNAi of TbORC1/CDC6 (A), TbORC4 (B), Tb7980 (C), Tb3120 (D), TbORC1B (E), Tb1120 (F), all derived from pLew111 (Figure 3.2). The gene-specific products are shown flanking the "stuffer" region (white), and so are the restriction sites used for their cloning into the plasmid. The Tet operator (yellow) is located downstream of the T7 promoter (red), and the bleomycin resistance marker (*BLE*; orange) is shown. The ampicilin resistance gene is shown as *amp*R (in black), and the bacterial origin in depicted as ColE1 origin (grey). The *RRNA* spacer fragment (green) is shown with the NotI restriction site used to linearize the plasmid for transfection into the parasite. All constructs were confirmed through restriction enzyme digestions (G), detailed description in the main text. The 1 Kb plus DNA ladder (Invitrogen™) is shown as a size reference, and the expected fragment sizes are shown below each gel image. (*) digestion with HindIII and XhoI results in the excision of PCR product A; (**) digestion with AflII and BamHI results in the excision PCR product B. Note that in some cases the gel image has been cropped to display the results from the same plasmid side-by-side, and are separated by a white space.

3.3.2 Effect of TbORC1/CDC6, TBORC4, Tb3120 or TbORC1B expression silencing by RNAi on cell growth, cell cycle progression, and DNA replication,

The effect of depletion of TbORC1/CDC6 by RNAi has been reported in several published works (Tiengwe *et al.*, 2012b; Godoy *et al.*, 2009; Benmerzouga *et al.*, 2013). In each case, depletion of TbORC1/CDC6 in PCF cells caused slowing of growth after ~3-4 days (but growth arrest in only one study), accumulation of 0N1K (zoid) cells and, in one study, reduced BrdU incorporation (Benmerzouga *et al.*, 2013). As work from this lab has not previously targeted TbORC1/CDC6 by stem-loop RNAi, the new cell line targeting TbORC1/CDC6 in this way was used as a positive control and to allow comparison with each of the putative ORC-like factors. In order to assess any influence of tetracycline in the assays, the 29-13 cell line was used as a negative control, as no effect is expected.

The aim of these assays was to elucidate whether each of the putative ORC-like factors are involved in DNA replication. For this purpose, a mid-log phase (~1 x 10^7 cells.ml^{-1}) culture of each RNAi cell line was diluted to a concentration of 5.5 x 10^5 cells.ml^{-1}, and divided into two different cell culture flasks. To one flask, tetracycline (2 µg.ml^{-1}) was added (Tet +) to induce expression of the stem-loop dsRNA, while no tetracycline was added to the other, thus serving as a non-induced control (Tet -). Cell density was assessed every 24 h in order to evaluate growth. Cell cycle progression was also examined at the same time points, using two complementary strategies. First, the cells were fixed and stained with DAPI, allowing the quantification of the different cell cycle stages in the asynchronous cell cultures. This analysis is possible in *T. brucei* because the nuclear and kinetoplastid genomes possess distinct S phase timing, with the DNA replication and segregation of the kinetoplast occurring before the nuclear S phase (McKean, 2003), as shown in Figure 3.49. Therefore, the analysis of the ratio, morphology, and configuration of the nucleus and kinetoplast DNA in a cell can be used to identify the cell cycle stage of an individual cell from an asynchronous culture (Woodward and Gull, 1990; Siegel *et al.*, 2008). Cells can be classified as follows (N – nucleus; K – kinetoplast): most 1N1K cells are assumed to be in G1 phase; 1N1eK (eK – elongated kinetoplast; replicating but not yet divided kinetoplast) are in S phase; 1N2K cells are in G2 phase; and 2N2K cells are post-mitotic cells that are about to, or are already, undergoing

cytokinesis (McKean, 2003; Wheeler *et al.*, 2013b). Secondly, the cells were fixed and permeabilised with methanol, stained with propidium iodide (PI), and analysed by flow cytometry (fluorescence-activated cell sorting, FACS) (Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013; Siegel *et al.*, 2008). Propidium iodide is a fluorescent DNA intercalating agent, widely used in flow cytometry analysis to evaluate cell viability[f] or DNA content in cell cycle studies (Moore *et al.*, 1998). Briefly, a minimum of 5 x 10^5 cells from both the Tet - and Tet + cultures were stained with PI and analysed using a FACSCalibur™ (BD Biosciences) flow cytometer system, as described in the materials and methods Chapter 2, section 2.4.1.5. Because PI fluorescence excitation and emission maximums are 535 nm and 617 nm, respectively, stained cells were detected in the FL2-A channel of the flow cytometer, and plotted onto histograms representing the amount of cells according to their DNA content (for example, as shown in Figure 3.11, box). A non-perturbed asynchronous cell culture should result in two predominant peaks, one at an intensity of around 200 in the FL2-A channel (in this case) corresponding to cells in G1 phase (with 2N DNA content), and a smaller peak at around 400, representing cells in G2 phase or post-mitotic but pre-cytokinesis (G2/M population; 4N content). Cells with fluorescence intensity between these two peaks have a variable DNA content between 2N and 4N and correspond to cells that are in different stages of S phase. All other populations detected, with either more or less DNA content, are considered abnormal. To complement the above growth and cell cycle assays, the effect of the RNAi on DNA replication was more directly assayed. For this purpose, cells were incubated for three hours with 5-ethynyl-2'-deoxyuridine (EdU), a synthetic nucleoside analogue of thymidine (Salic and Mitchison, 2008), after which the number of EdU-positive cells was assessed. The incubation time of three hours was chosen as it is somewhat longer than the predicted nuclear S phase of *T. brucei* PCF cells (estimated to be 1.51 hours), but shorter than S and G2 (G2 phase estimated to take around 1.87 hours) phases combined (Woodward and Gull, 1990). Therefore, it is short enough so that most cells that are labelled should correspond with those undergoing or having just completed DNA synthesis (i.e. primarily 1N1eK and 1N2K cells), but not sufficiently long to label 2N2K cells that have progressed through the whole of S and G2 phases and have undergone

[f] Propidium iodide is a membrane impermeant molecule and is therefore only incorporated by cells whose cellular membrane has been compromised (dying or permeabilised cells) and is therefore excluded from viable cells.

mitosis. In practice, this pattern of labelling was consistently observed in unperturbed cells, where ~25% of the cells were labelled with EdU, the vast majority of which were 1N1eK or 1N2K cells (data not shown). Like 5-bromo-2'-deoxyuridine (BrdU), another synthetic thymidine analogue widely used to detect DNA synthesis in proliferating cells (Benmerzouga *et al.*, 2013; Gassen *et al.*, 2012; Woodward and Gull, 1990; Aparicio *et al.*, 2009; Semple *et al.*, 2006), EdU is incorporated into the cell's DNA and can be detected and quantified after cell fixation and staining. The major difference between the two analogues resides in the method of detection. BrdU is detected using anti-BrdU antibodies, thus requiring denaturation of the DNA helix to access the incorporated molecules. In contrast, EdU detection relies on a copper-catalysed (Cu(I)-catalysed [3+2] cycloaddition, or "click" chemistry) reaction between an alkyne group in the EdU molecule and an azide (Salic and Mitchison, 2008; Rostovtsev *et al.*, 2002), usually coupled to an AlexaFluor® fluorophore. Because the AlexaFluor®-conjugated azide is small enough to access the double-stranded DNA (dsDNA), EdU detection does not require treatment of the cells with either acid or DNase, thus allowing simultaneous detection of EdU and nuclear proteins by immunofluorescence (Salic and Mitchison, 2008). For this study the cells stained for EdU incorporation were also counterstained for DAPI, and the same samples were used to generate the data on cell cycle, based on DAPI counting, as well as the percentage of EdU positive cells. Cells were first categorised into the different N/K configurations, and then the same cells were analysed for EdU incorporation. EdU positive cells were counted and the percentage of such positive cells then calculated for both the Tet – and Tet + samples at every time point. To allow comparisons between time points, the number of Tet – EdU positive cells was considered to be 100% at each time point, and the percentage of Tet + EdU positive cells was then calculated relative to that. To test whether the RNAi being induced is targeting the mRNA of the expected gene, it was hoped that we could examine protein levels. Because antibodies against the factors under study are not available, it was sought to endogenously tag one of the alleles of each gene with 12 copies of the myc epitope (12myc; construct transfection strategy are detailed in sections 3.4.1, 3.4.4) in the corresponding RNAi cell line, as has been previously done (Rudd *et al.*, 2013). This would allow protein levels to be assessed throughout the course of the RNAi experiment by western blot using anti-myc antiserum (as described in section 3.4.9).

Unfortunately, despite several attempts, this was only accomplished for one RNAi cell line, the one targeting TbORC4 (here referred to as TbORC4 Clb). For the remaining RNAi cell lines, the efficiency of silencing was assessed by reverse transcription quantitative real-time PCR (RT-qPCR), similar to other studies (Tiengwe *et al.*, 2012b; Jones *et al.*, 2014). For this purpose, primers targeting *TbORC4*, *Tb3120* and *TbORC1B* were designed according to the guidelines for primers to be used in quantitative real-time PCR (qPCR) (Alvarez-Fernandez, 2013; Dymond, 2013), while the primer sequences targeting *TbORC1/CDC6* have been published previously (Tiengwe *et al.*, 2012b). For normalisation, previously described primer sequences targeting the Tb927.10.12970 gene (referred to as *C1*) were used (Kabani *et al.*, 2009). All primer sequences are shown in the materials and methods Chapter 2, Table 2-8. Primer efficiency and specificity were assessed for all pairs of primers through, respectively, the analysis of calibration curves and melting profiles, which resulted in efficiencies of approximately 100%, all within a 15% interval (data not shown). For each time point, cells from both Tet – and Tet + cultures were collected for RNA extraction and subsequent conversion into complementary DNA (cDNA), as described in the materials and methods Chapter 2, section 2.4.1.6. The cDNA was then used for the qPCR reaction, and the resulting fluorescence intensity data was analysed by relative quantification using the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001), using *C1* as the endogenous control and the respective Tet – sample as calibrator (in which the expression of the analysed gene should be 100%), as detailed in the materials and methods Chapter 2, section 2.4.1.7. The results below, with the exception of the flow cytometry, result from two experiments performed independently.

3.3.2.1 Tetracycline *per se* has no effect on cell growth, cell cycle progression, or on DNA replication

As expected, no clear differences were observed between the Tet – and Tet + cultures of the parental cell line 29-13 (Figure 3.5) in both cell growth (Figure 3.5, A), cell cycle progression (Figure 3.5, B and C; and Figure 3.11), or EdU uptake by the cells (Figure 3.5, D), thus showing that tetracycline causes no significant effect in these conditions. The smaller proportions of the 1N2K and 2N2K cell populations at 72 h, 120 h and 168 h most likely reflect the fact that these samples were collected from cultures at high cell densities, where growth

rate slows (Figure 3.5, A, days 3, 5 and 7, respectively), rather than an effect of tetracycline, as the same percentages were seen for the Tet – sample (data not shown individually; in Figure 3.5, B the Tet – values represent the mean percentage of each cell type population assessed for the 8 time points). Measurement of EdU positive cells showed some variability between experiments and time points (Figure 3.5, D), most likely due to either variability in uptake or detection.

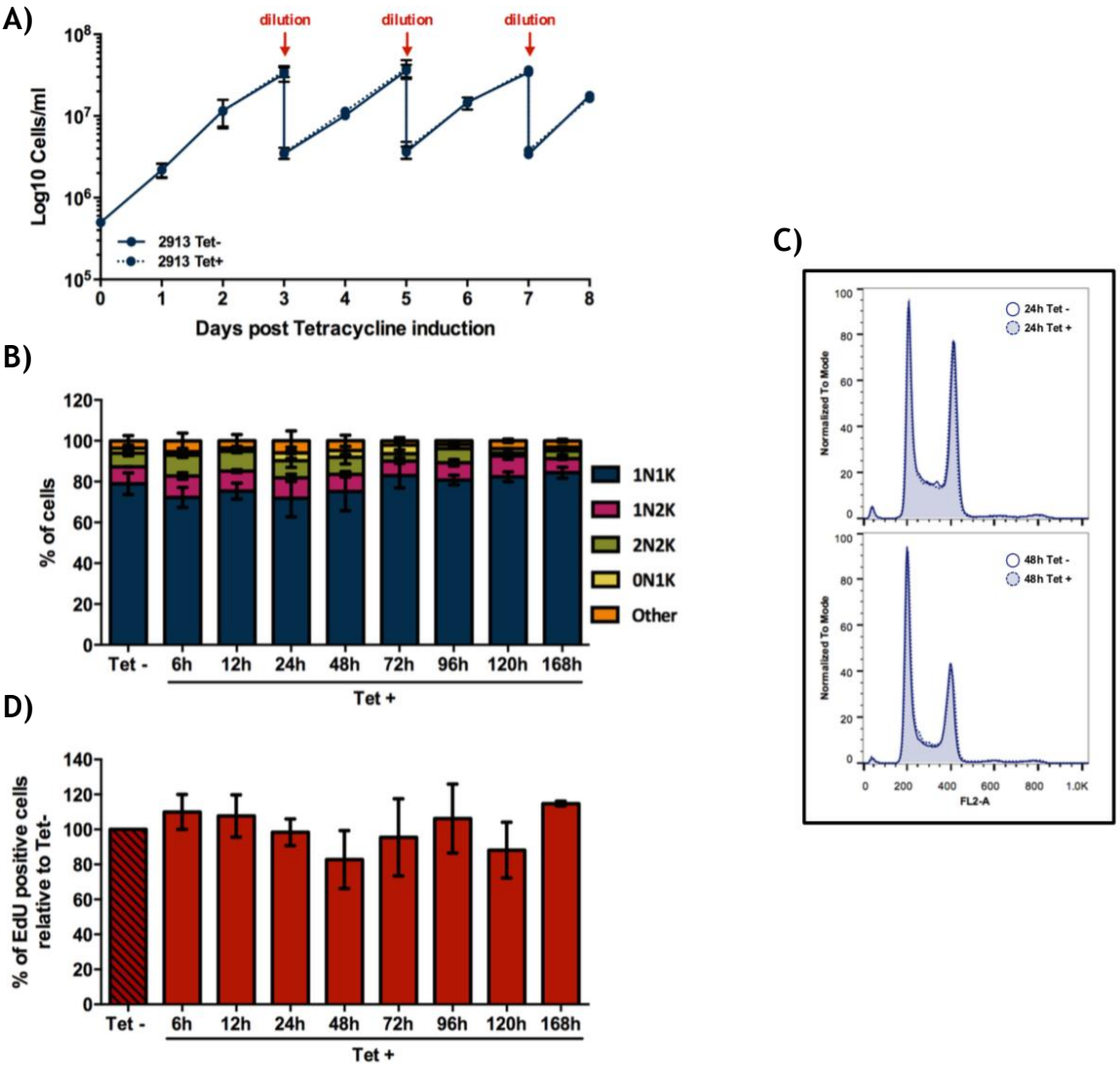


Figure 3.5. Effect of tetracycline on the parental RNAi 29-13 cell line.
A) Growth curves depicting the cell growth of the un-induced (Tet -) and tetracycline-induced (Tet +) cell cultures over six days. Cell concentration was assessed every 24 h, and plotted on a Log_{10} Y-axis graph. The individual points represent the mean from two independent experiments (n = 2), while the error bars depict the standard error of the mean (SEM). The red arrow pinpoints a 1:10 dilution of both Tet - and Tet + cultures. B) Quantification of cells in the different cell cycle stages throughout the course of six days under tetracycline induction, based on the nuclear and kinetoplast configuration of the cells stained with DAPI. A minimum of 150 cells was counted per time point and experimental group (Tet - and Tet +), and percentages of each cell type (1N1K, 1N2K, 2N2K, 0N1K, and others) were calculated relative to the total amount of cells analysed. The graph represents the mean of each cellular population obtained in two independent experiments (n = 2), while the error bars refer to SEM. The Tet - bar represents the mean of the percentage of each cell type throughout the course of each of the two experiments. C) Histograms representing

the distribution of the cell population according to DNA content, stained with PI (FL2-A channel), and assessed by flow cytometry. Approximately 30,000 cells were analysed per sample. Histograms from Tet - (full line) and Tet + (dashed line) cell cultures are shown for 24 h and 48 h. The histograms of other time points are shown in the appendices, section 7.3.2. Results from only one experiment are shown. D) Percentage of EdU positive cells in the Tet + samples relative to the percentage of EdU positive cells in the Tet - culture from the same time point. A minimum of 150 cells were analysed per time point and group (Tet - and Tet +). The mean from two independent experiments (n = 2) is shown; error bars represent the SEM.

3.3.2.2 RNAi of TbORC1/CDC6

Induction of TbORC1/CDC6 RNAi silencing resulted in a pronounced growth defect: 48 h post-tetracycline addition, growth slowed and there was evidence of cell death thereafter (results from clone b (Clb) are shown in Figure 3.6, A). This growth defect was accompanied by a dramatic increase of enucleated cells (0N1K, or zoids; an example is shown in Figure 3.6 F): only 1-3% of zoid cells were seen prior to RNAi induction, but this increased to 50-60% of the population by 48 h post-induction, and reached a maximum of ~70% of the population at ~72 h (Figure 3.6, C). Concurrently, the numbers of 1N1K cells (an example is shown in Figure 3.6 F) decreased almost proportionally, from being approximately 80% of the un-induced population to about 30% at 48 h post-induction, and only 15% of the cells from 72 h onwards (Figure 3.6, C). As an indication of TbORC1/CDC6's involvement in DNA replication, 2N2K cells (post-mitotic or undergoing cytokinesis, an example is shown in Figure 3.6 F) were virtually abolished from the population by 72 h post-induction (Figure 3.6, C). Though increased levels of 1N2K cells (Figure 3.6, F) might be expected if replication is affected, there was little evidence for this (Figure 3.6, C). If anything, levels of 1N2K cells decreased throughout the experimental time course (Figure 3.6, C). In addition, some cells, with no clear classifiable morphology or nucleus to kinetoplast ratio, were also observed in late time points after induction, and were categorised as "other" (an example is shown in Figure 3.6, F). The amount of these 'other' cells increased mainly from 72 h after induction, although zoids remained the dominant cell type in the population. These results were supported by the data obtained by FACS, where the increase in zoids corresponded with a decrease in the number of cells in both G1 and G2 phases, and a clear increase in the levels of sub-G1 cells (peak below 200 in the FL2-A axis in Figure 3.6, D) from 24 h onwards (Figure 3.6,D, shows 24 h, 48 h and 96 h time points, and further time points are summarised in Figure 3.11).

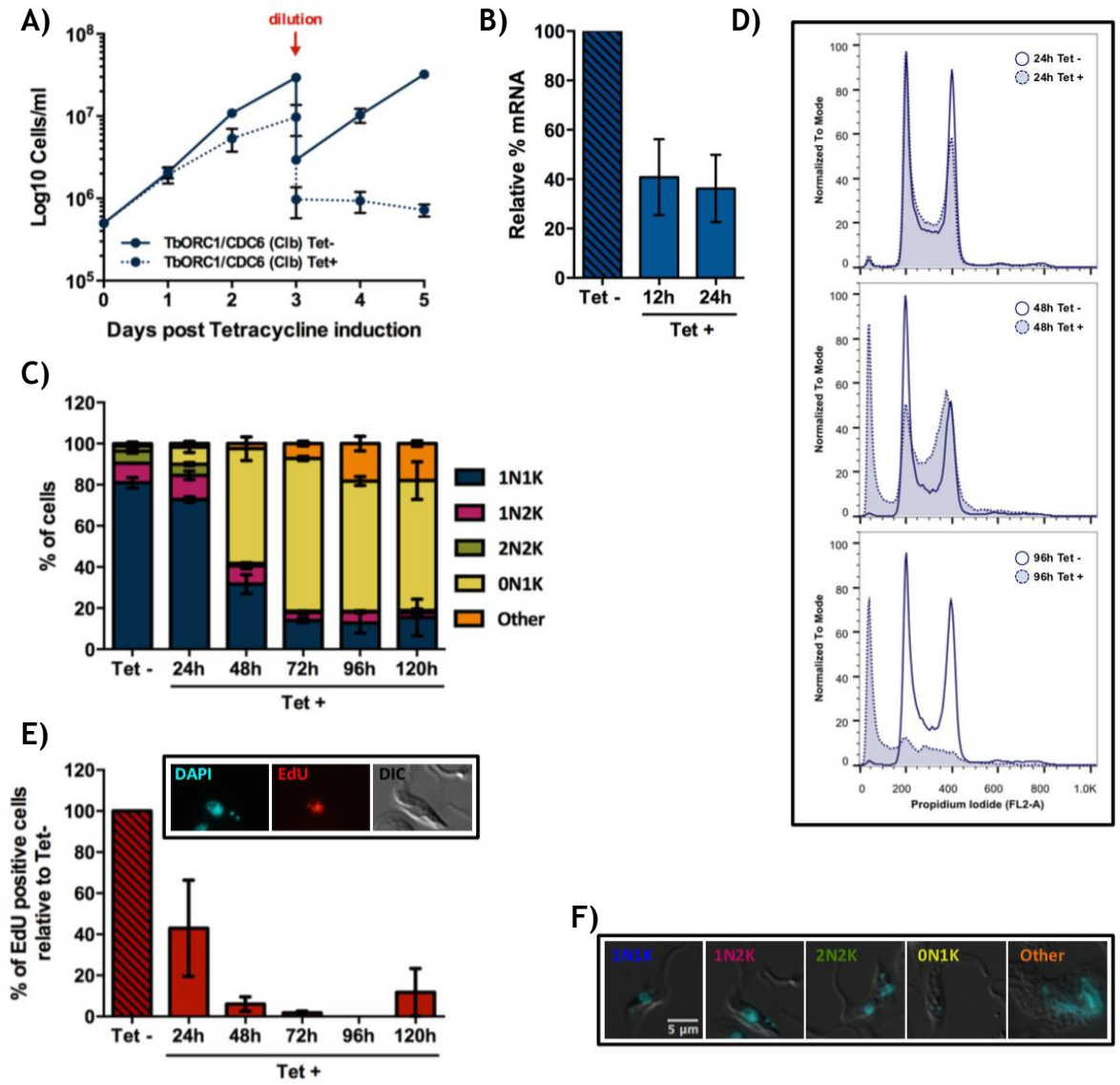


Figure 3.6. Effect of TbORC1/CDC6 depletion by induction of specific gene targeted RNAi over time.

A) Growth curves of un-induced (Tet -) and tetracycline-RNAi induced (Tet +) cell cultures over five days. Cell concentration was assessed every 24 h, and plotted on a Log_{10} Y-axis graph. The individual points represent the mean from two independent experiments, while the error bars depict the SEM. The red arrow pinpoints a 1:10 dilution of both Tet - and Tet + cultures. B) Efficiency of RNAi knockdown of TbORC1/CDC6 mRNA levels assessed by RT-qPCR. The results represent the amount of mRNA at 12 h and 24 h after RNAi induction relative to the non-induced sample (Tet -), calculated using the $\Delta\Delta C_t$ method. The mean of two independent experiments (n = 2) is shown, and the error bars represent SEM. C) Quantification of cells in the different cell cycle stages throughout the course of five days of RNAi induction, based on the nuclear and kinetoplast configuration of the cells stained with DAPI. A minimum of 150 cells was counted per time point and experimental group (Tet - and Tet +), and percentages of each cell type (1N1K, 1N2K, 2N2K, 0N1K, and others) were calculated relative to the total amount of cells analysed. The graph represents the mean of each cellular population obtained in two independent experiments (n = 2), while the error bars refer to SEM. D) Histograms representing the distribution of the cell population according to DNA content, stained with PI (FL2-A channel), assessed by flow cytometry. Approximately 30,000 cells were analysed per sample. Histograms from Tet - (full line) and Tet + (dashed line) cell cultures are shown for the 24 h, 48 h and 96 h time points. The histograms of other time points are shown in the appendices, section 7.3.2. Results from only one experiment are shown. E) Percentage of EdU positive cells in the Tet + samples relative to the percentage of EdU positive cells in the Tet - culture from the same time point. A minimum of 150 cells were analysed per time point and group (Tet - and Tet +). The mean from two independent experiments is shown (n = 2); error bars represent the SEM. An example of an EdU positive cell is shown. F) Examples of cells categorised as 1N1K, 1N2K, 2N2K, 0N1K, or other; DAPI (cyan) and DIC (cell outline) are shown overlaid.

Overall, the above observations are consistent with the results published for TbORC1/CDC6 depletion by RNAi in PCF cells (Tiengwe *et al.*, 2012b; Godoy *et al.*, 2009; Benmerzouga *et al.*, 2013), where the predominant phenotype is the generation of zoids. The notable difference is that the cell growth impairment and accumulation of zoids appear to arise more quickly and are more severe in the cell line used here. Whether this is because the stem-loop RNAi system used here was more efficient is not clear, though plausible, since (Benmerzouga *et al.*, 2013) also used stem-loop RNAi, in which the rapidity of cell growth impairment and levels of zoid accumulation were closer to the results here shown than those of (Tiengwe *et al.*, 2012b) and (Godoy *et al.*, 2009). Moreover, (Tiengwe *et al.*, 2012b) reported ~70% reduction in TbORC1/CDC6 mRNA after 4 days of RNAi induction, and RT-qPCR here suggests 60-70% loss after 12 h and 24 h. No later time points were examined by RT-qPCR, as most of the cell population was composed of zoids and the lack of a nucleus in such a high proportion of cells resulted in aberrant mRNA levels and distorted the RT-qPCR analysis (data not shown).

Recently, loss of TbORC1/CDC6 has been shown to impair DNA replication through lowered levels of BrdU incorporation after RNAi (Benmerzouga *et al.*, 2013), and results herein corroborate these observations. As shown in Figure 3.6, E, the number of EdU-positive cells (an example of a positive cell is shown inside the box) decreased with time after RNAi induction, until essentially no cells displayed EdU labelling after ~72 h. Indeed, reduced EdU staining appeared to occur before any growth or cell cycle defect were evident. For instance, at 24 h, a 40-50% decrease in EdU positive cells was already observed in the Tet + culture compared with Tet – (Figure 3.6, E), although growth was not affected and only a small proportion of zoids were detected (Figure 3.6, C). At 120 h after RNAi induction, a small increase in EdU-positive cells was seen in one experiment (Figure 3.6, E), though not in the other (Figure 3.6, E). It is possible that at this stage some 'revertants' arose in which RNAi is less effective.

As mentioned earlier, two clones targeting TbORC1/CDC6 by RNAi have been analysed, and all above data refers to one clone (Clb). In the other clone (Cla), reversion to growth was seen earlier (from 72 h onwards) in both experimental repeats, illustrating that non RNAi-reactive cells arise after that time point

(results from one experiment are shown in the appendices, Figure 7.7). Nonetheless, until the 48 h time point, the phenotypes observed for clone Cla were the same in every aspect to the ones observed for clone Clb.

Taken together, the growth, cell cycle, and EdU incorporation data strongly support TbORC1/CDC6 involvement in DNA replication, and suggest that it is the impairment of nuclear DNA replication that results in the formation of zoid cells, as has been proposed before (Ploubidou *et al.*, 1999).

3.3.2.3 RNAi of TbORC4

In the case of TbORC4, two non-reverting RNAi clones were analysed (clone Cla is shown in Figure 3.7, and clone Clb in Figure 3.8). In one of the clones, Clb, one allele of *TbORC4* was endogenously tagged with 12myc (as described in section 3.4.4). Previously, TbORC4 RNAi in PCF cells has been shown to result in very similar growth and cell cycle phenotypes to those seen after TbORC1/CDC6 RNAi, but no measurement of EdU incorporation was attempted (Tiengwe *et al.*, 2012b). Induction of RNAi resulted in a strikingly similar profile in both TbORC4 clones. In fact, all phenotypes observed after TbORC4 RNAi closely matched those obtained for TbORC1/CDC6 (Figure 3.6): perturbation in cell growth was noticeable 48 h post-induction, with growth arresting after 72 h (Figure 3.7 and Figure 3.8, A); at the same time points the proportion of the zoid population increased, while 1N1K, 1N2K and 2N2K cell types decreased, with both 1N2K and 2N2K being virtually absent at 120 h (Figure 3.7 and Figure 3.8, C); ~50% reduction in EdU incorporation was detected 24 h post-induction, and was almost abolished from 72 h onwards (Figure 3.7 and Figure 3.8, E). The RNAi perhaps slightly differed from TbORC1/CDC6 in that the amount of cells categorised as “other” was higher in both the TbORC4 RNAi cell lines 120 h post-induction, alongside a more pronounced reduction in 1N1K cells. Most likely, this is explained by the suggestion that at this stage in the TbORC1/CDC6 RNAi time course there was evidence for reversion (growth recovery and increased EdU incorporation relative to 96 h), which was not seen for TbORC4 RNAi. FACS analysis (Figure 3.7 and Figure 3.8, D; FACS data is summarised in Figure 3.11) confirms the parallels between TbORC4 and TbORC1/CDC6 RNAi, with accumulation of sub-G1 phase cells seen from 48 h onwards.

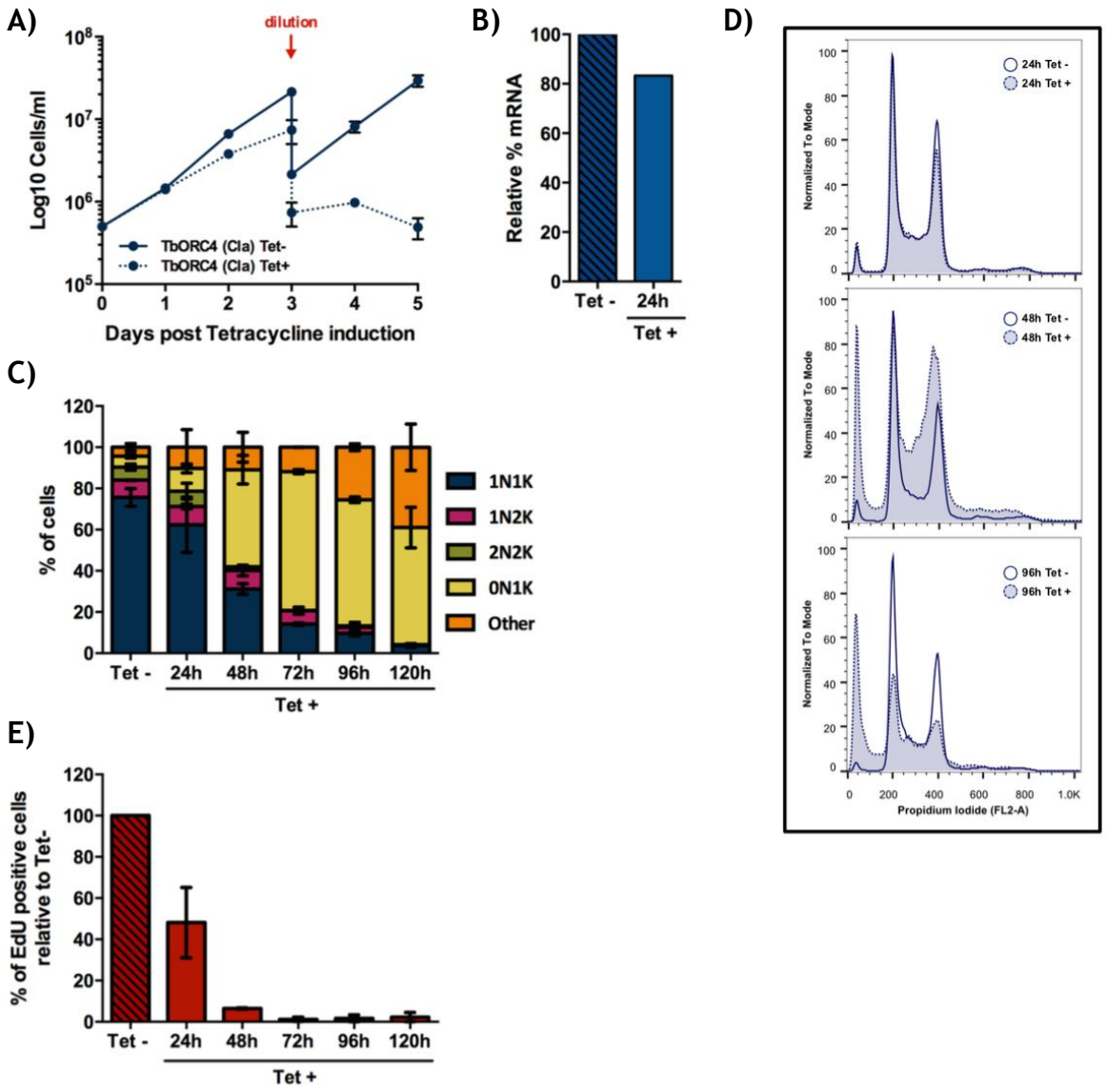


Figure 3.7. Effect of TbORC4 depletion by induction of specific gene targeted RNAi over time. Clone a.

A) Growth curves depicting the cell growth of the un-induced (Tet -) and tetracycline-induced (Tet +) cell cultures throughout the course of five days. Cell concentration was assessed every 24 h, and plotted on a Log_{10} Y-axis graph. The individual points represent the mean concentration calculated from two independent experiments (n = 2), while the error bars depict the SEM. The red arrow pinpoints a 1:10 dilution of both Tet – and Tet + cultures. B) Efficiency of the knockdown of TbORC4 mRNA levels assessed by RT-qPCR. The results represented refer to the relative amount of mRNA levels to the non-induced sample (Tet -), calculated using the $\Delta\Delta C_t$ method. The mean of two independent experiments (n = 2) is shown, and the error bars represent SEM. C) Quantification of cells in the different cell cycle stages throughout the course of five days under tetracycline induction, based on the nuclear and kinetoplast configuration of the cells stained with DAPI. A minimum of 150 cells was counted per time point and experimental group (Tet - and Tet +), and percentages of each cell type (1N1K, 1N2K, 2N2K, 0N1K, and others) were calculated relative to the total amount of cells analysed. The graph represents the mean of each cellular population obtained in two independent experiments (n = 2), while the error bars refer to SEM. The Tet - bar represents the mean of the percentage of each cell type throughout the course of each of the two experiments (five days). D) Histograms representing the distribution of the cell population according to DNA content, stained with PI (FL2-A channel), assessed by flow cytometry. Approximately 30,000 cells were analysed per sample. Histograms from Tet - (full line) and Tet + (dashed line) cell cultures are shown overlapped for the 24 h, 48 h and 96 h time points. The histograms of other time points are shown in the appendices, section 7.3.2. Results from only one experiment are shown. E) Percentage of EdU positive cells in the Tet + sample relative to the percentage of EdU positive cells in the Tet - culture from the respective time point. A minimum of 150 cells were analysed per time point and group (Tet - and Tet +). The mean from two independent experiments (n = 2) is shown; error bars represent SEM.

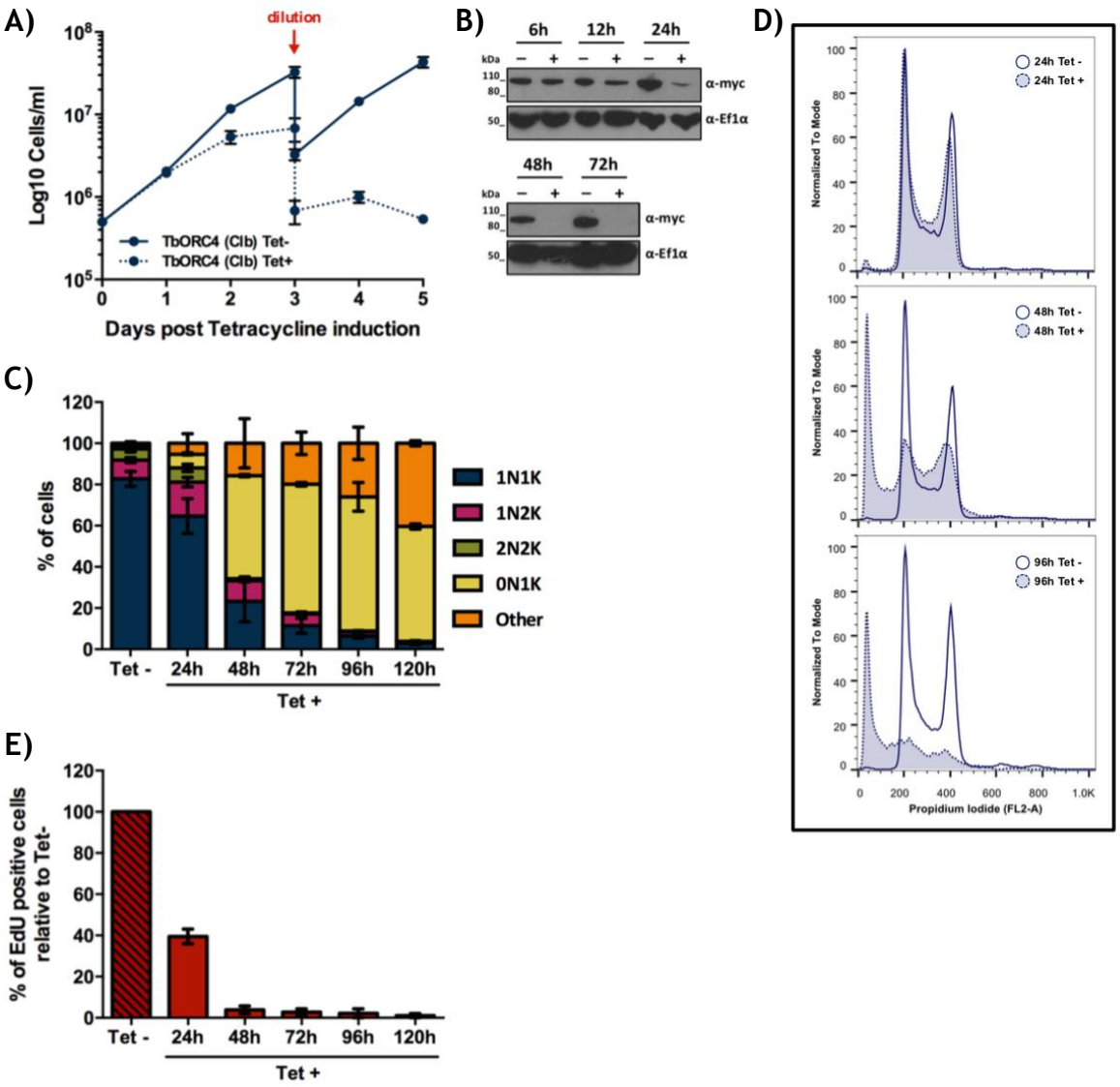


Figure 3.8. Effect of TbORC4 depletion by induction of specific gene targeted RNAi over time. Clone b.

A) Growth curves depicting the cell growth of the un-induced (Tet -) and tetracycline-induced (Tet +) cell cultures throughout the course of five days. Cell concentration was assessed every 24 h, and plotted on a Log_{10} Y-axis graph. The individual points represent the mean concentration calculated from two independent experiments (n = 2), while the error bars depict the SEM. The red arrow pinpoints a 1:10 dilution of both Tet – and Tet + cultures. B) The cell line has one allele endogenously tagged with 12myc, therefore, efficiency of the RNAi system induction was assessed by western blot, through the detection of TbORC4^{12myc} with the anti-myc antiserum (α-myc) in total protein extracts from both Tet - and Tet + cultures. The transcription elongation factor Ef1α (α-Ef1α) was used as a loading control. C) Quantification of cells in the different cell cycle stages throughout the course of five days under tetracycline induction, based on the nuclear and kinetoplast configuration of the cells stained with DAPI. A minimum of 150 cells was counted per time point and experimental group (Tet - and Tet +), and percentages of each cell type (1N1K, 1N2K, 2N2K, 0N1K, and others) were calculated relative to the total amount of cells analysed. The graph represents the mean of each cellular population obtained in two independent experiments (n = 2), while the error bars refer to SEM. The Tet - bar represents the mean of the percentage of each cell type throughout the course of each of the two experiments (five days). D) Histograms representing the distribution of the cell population according to DNA content, stained with PI (FL2-A channel), assessed by flow cytometry. Approximately 30,000 cells were analysed per sample. Histograms from Tet - (full line) and Tet + (dashed line) cell cultures are shown overlapped for the 24 h, 48 h and 96 h time points. The histograms of other time points are shown in the appendices, section 7.3.2. Results from only one experiment are shown. E) Percentage of EdU positive cells in the Tet + sample relative to the percentage of EdU positive cells in the Tet - culture from the respective time point. A minimum of 150 cells were analysed per time point and group (Tet - and Tet +). The mean from two independent experiments (n = 2) is shown; error bars represent SEM.

TbORC4 expression after RNAi levels were assessed in clone Cla by RT-qPCR, as previously described for TbORC1/CDC6 RNAi. Surprisingly, in contrast to the TbORC1/CDC6 RNAi cell line, which shows an identical phenotype and a ~60% decrease in mRNA levels after 24 h, TbORC4 mRNA was only reduced by ~20% relative to the non-induced sample (Figure 3.7). How such a small knockdown in mRNA levels can lead to such striking and rapid phenotypes, which mimic TbORC1/CDC6 (with ~3 times more mRNA loss), is not known. In the endogenously tagged cell line, clone Clb, detection of the TbORC4^{12myc} protein revealed that protein levels decreased progressively with time, with very low levels detected 24 h post-induction (when a decrease in 50% of EdU-positive cells was already observed), and no detectable signal in cells after 48 h and 72 h (Figure 3.8, B). The relative mRNA levels of TbORC4 were not assessed for clone Clb, and therefore it is not known whether the mRNA relative levels in this cell line are comparable to the ones in clone Cla. Nevertheless, the indistinguishable phenotypes observed in the two cell lines would suggest that knockdown of TbORC4 may be equivalent. Why, then, the RT-qPCR and western blot analyses do not match is unclear. Ultimately, these results show that TbORC4 has an important role in DNA replication and, given the strong phenotypic overlap, in every aspect studied, with RNAi of TbORC1/CDC6, might support the hypothesis that these two factors act on DNA replication in the same complex or, at least, play similar roles in the DNA replication process in *T. brucei*.

3.3.2.4 RNAi of Tb3120

Unlike RNAi silencing of TbORC1/CDC6 and TbORC4, RNAi targeting Tb3120 resulted in a late and mild growth defect, only noticeable four days after induction (Figure 3.9, A). Indeed, despite the use of stem-loop RNAi, this response was similar to what has been reported by (Tiengwe *et al.*, 2012b). This delayed effect on cell growth correlated with a later timing of increase in the number of zoid cells, from ~3% in the non-induced culture to ~30% of the cell population 96 h after RNAi induction. Unlike in TbORC1/CDC6 or TbORC4 RNAi, zoids only reached a maximum of ~43% of the cells at 120 h (Figure 3.9, B), and rapidly thereafter (168 h post-induction), the “other” cells became the dominant cell type in the population. The “other” did not display consistent cell morphology or ratio between nucleus and kinetoplast, as was seen following TbORC1/CDC6 and TbORC4 RNAi, although not in such high numbers. Altogether,

the cell cycle defects observed by DAPI counting analysis were confirmed by FACS (Figure 3.9, C), which showed an increase in zoid cells and a progressive reduction of both G1 and G2 phases (results are summarised in Figure 3.11). Although some of the cells categorised as “others” appeared to have an enlarged nucleus or multiple nuclei, FACS data revealed no detectable cells with a higher DNA content than 4N, suggesting that these represented only a minor component of the population, not discernible by FACS, or that these cells had not replicated their nuclear DNA.

Despite the slow accumulation of growth and cell cycle phenotypes, analysis of EdU incorporation revealed that Tb3120 silencing also had a marked effect on DNA replication. Indeed, the timing of the reduction in EdU incorporation correlated with the appearance of the growth defects and the increase in aberrant cells (Figure 3.9, D). A very small decrease in EdU incorporation (~15 – 20%) was observed at 24 and 48 h after RNAi induction, suggesting that, like for TbORC1/CDC6 and TbORC4, impaired replication precedes the other cell cycle phenotypes. Though it is possible that this small decrease might result from variability intrinsic to the EdU assay itself, at 72 h a >50% decrease in EdU incorporation was seen, immediately preceding the growth and cell cycle defects observed at 96 h. Thereafter, the steady increase in the number of aberrant cells was accompanied by a progressive decrease in EdU positive cells, culminating in a reduction of ~95% of by 168 h (Figure 3.9, D). To check that the RNAi correctly targeted Tb3120, RT-qPCR was performed at 24 h, 72 h and at 96 h, matching when defects were first detected and when the population of aberrant cells was less than 50% of the total cell population. An average decrease in 50% of mRNA levels relative to the non-induced samples suggests that RNAi induction was successful. Indeed, as no difference was seen in relative mRNA levels at the three time points examined, it is likely that this corresponds with maximum loss of mRNA (Figure 3.9, E). These results raise the question of why the phenotypes seen upon Tb3120 RNAi are so late, especially as the levels of RNAi knockdown at 24 h are comparable with those seen for TbORC1/CDC6 RNAi, are stable over time and are higher than the effects measured by RT-qPCR for TbORC4 RNAi. One possible scenario is that Tb3120 does not play an as central, or crucial role in DNA replication as both TbORC1/CDC6 and TbORC4. However, the overall similarity of the phenotypes that ultimately resulted from

Tb3120 RNAi relative to both TbORC1/CDC6 and TbORC4 suggests that all these three factors act in the same biological context. Therefore another explanation is that Tb3120 protein is more stable, perhaps due to a longer half-life or slower turnover rate, than TbORC1/CDC6 and TbORC4 and, consequently, requires longer RNAi induction to reach lowered protein levels that match the other two factors. Unfortunately, it was not possible to tag Tb3120 with 12myc in the RNAi cell line and therefore, testing these hypothesis in the PCF cells was not possible.

As for TbORC1/CDC6 and TbORC4, two independently obtained Tb3120 RNAi clones, Cla and Clb, were analysed. Clone Cla, which presented an even less pronounced phenotype, is shown in the appendices, Figure 7.8. Nonetheless, the same broad responses were seen, validating the findings that were here analysed more thoroughly for clone Clb (Figure 3.9).

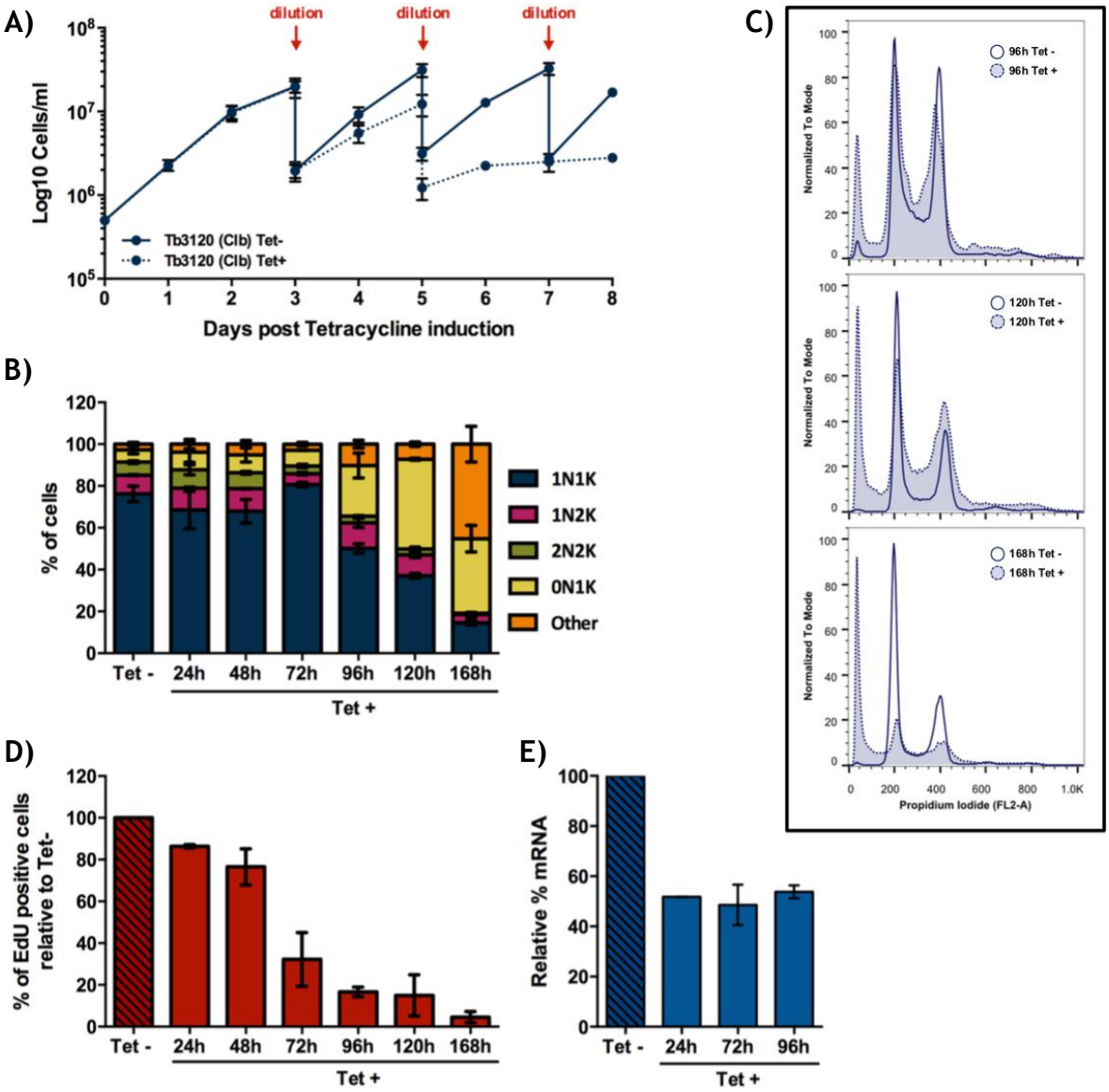


Figure 3.9. Effect of Tb3120 depletion by induction of specific gene targeted RNAi over time. A) Growth curves depicting the cell growth of the un-induced (Tet -) and tetracycline-induced (Tet +) cell cultures throughout the course of eight days. Cell concentration was assessed every 24 h, and plotted on a Log_{10} Y-axis graph. The individual points represent the mean concentration calculated from two independent experiments (n = 2), while the error bars depict the SEM. The red arrows pinpoint a 1:10 dilution of Tet - and Tet + cultures. B) Quantification of cells in the different cell cycle stages throughout the course of five days under tetracycline induction, based on the nuclear and kinetoplast configuration of the cells stained with DAPI. A minimum of 150 cells was counted per time point and experimental group (Tet - and Tet +), and percentages of each cell type (1N1K, 1N2K, 2N2K, 0N1K, and others) were calculated relative to the total amount of cells analysed. The graph represents the mean of each cellular population obtained in two independent experiments (n = 2), while the error bars refer to SEM. The Tet - bar represents the mean of the percentage of each cell type throughout the course of each of the two experiments (six days). C) Histograms representing the distribution of the cell population according to DNA content, stained with PI (FL2-A channel), assessed by flow cytometry. Approximately 30,000 cells were analysed per sample. Histograms from Tet - (full line) and Tet + (dashed line) cell cultures are shown overlapped for the 96 h, 120 h and 168 h time points. The histograms of other time points are shown in the appendices, section 7.3.2. Results from only one experiment are shown. D) Percentage of EdU positive cells in the Tet + sample relative to the percentage of EdU positive cells in the Tet - culture from the respective time point. A minimum of 150 cells were analysed per time point and group (Tet - and Tet +). The mean from two independent experiments (n = 2) is shown; error bars represent SEM. E) Efficiency of the knockdown of Tb3120 mRNA levels assessed by RT-qPCR. The results represented refer to the relative amount of mRNA levels to the non-induced sample (Tet -), calculated using the $\Delta\Delta C_t$ method. The mean of two independent experiments (n = 2) is shown, and the error bars represent SEM.

3.3.2.5 RNAi of TbORC1B

In contrast with the factors described above, the effects of RNAi on TbORC1B have never been reported before, and its proposed involvement in DNA replication is so far based only on its interaction with TbORC1/CDC6 (Dang and Li, 2011). Surprisingly, TbORC1B silencing resulted in a severe growth defect, evident as early as 24 h post-induction, one day before any growth effect was noticeable following TbORC1/CDC6 or TbORC4 RNAi (Figure 3.10, A). Due to this rapid response, putative cell cycle anomalies and defects on DNA replication were investigated at earlier time points than the ones analysed for the other putative Orc-like factors: 6 h and 12 h post-induction. At 6 h post RNAi induction, when there was no detectable emergence of zoids or other aberrant cells, a small increase in the proportion of 1N2K and 2N2K cells was seen, together with a small decrease in the number of 1N1K cells (Figure 3.10, C). This would suggest an increase in cells in G2/M phase and a decrease in G1 phase cells, respectively. However, this was not supported by flow cytometry: at the 6 h time point, an increase in the G1 phase (2N) population was observed together with a slight decrease in the G2/M (4N) population (Figure 3.11 and Figure 3.10, D). Therefore, it appears that some of the cells morphologically classified as 1N2K might have not replicated their nuclear DNA after 6 h post-induction. Quantification of EdU labelling revealed that EdU positive cells were reduced by ~60% at 6 h post-induction, suggesting that the number of actively replicating cells had already been reduced by more than 50% (Figure 3.10, E). These observations are reinforced by data obtained at 12 h post-RNAi. Here, the number of EdU positive cells was reduced by ~90% (Figure 3.10, E). This decrease in replicating cells was accompanied by a further increase in the number of 1N2K cells (~23.5% compared with ~12% in the Tet – sample), a very small increase in the amount of detectable zoid cells, further reduction of 1N1K cells, and near loss of 2N2K cells (Figure 3.10, C). FACS data at this time point suggest a decrease in G2/M (4N) cells, consistent with the observed reduction in 2N2K cells. Together with an increase in G1 (2N) cells, despite the observed reduction in numbers of 1N1K cells and the striking reduction in EdU incorporation, the decrease in G2/M (4N) cells strongly suggests that the 1N2K cells have not replicated their nuclear DNA, although the kinetoplast has replicated and segregated as normal. At 24 h post-induction, when EdU incorporation was essentially undetectable, >60% of the population was

composed of zoid cells, while both 1N1K and 1N2K populations were reduced, and virtually no 2N2K cells were observed (Figure 3.10, C), mirroring the cell cycle phenotypes seen at ~48 h post RNAi for TbORC1/CDC6 or TbORC4. This was reflected by the FACS data at 24 h post RNAi, which showed hugely reduced G1 (2N) and G2/M (4N) cell populations, and increased levels of cells with a fluorescent signal smaller than G1, supporting the increase in the percentage of zoid cells in the population (Figure 3.10, D, and Figure 3.11). Between 48 h and 72 h post-induction, cell growth stalled (Figure 3.10, A), and thereafter cell numbers decreased, which was accompanied by the number of zoids continuing to rise to ~70-80% of the cell population, in parallel with a further decrease in 1N1K and 1N2K cells, complete abolishment of 2N2K cells, and a small increase in 'others' cells (reaching ~15% at the latest time point, 120 h; Figure 3.10, C).

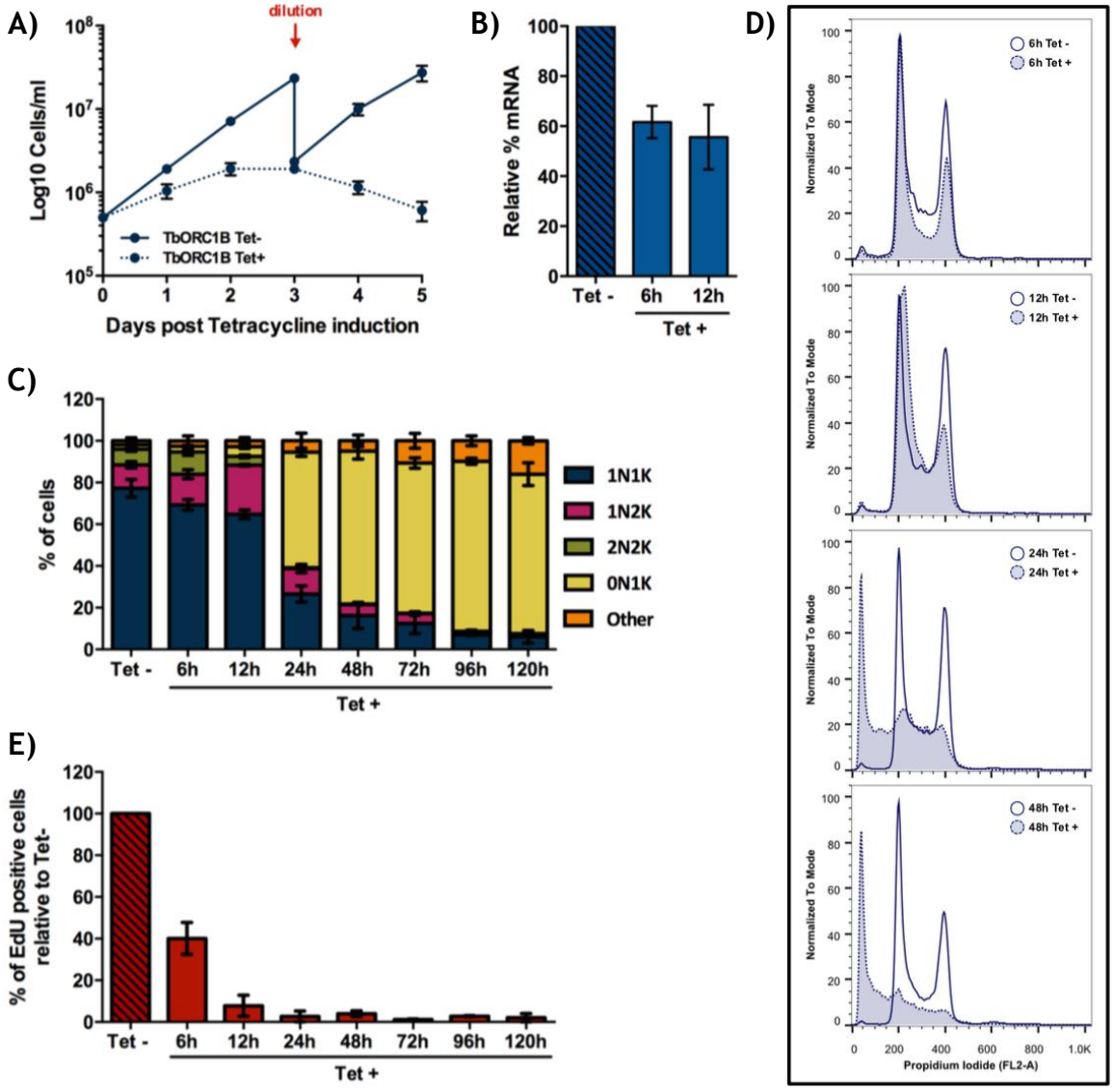


Figure 3.10. Effect of TbORC1B depletion by induction of specific gene targeted RNAi over time.

A) Growth curves depicting the cell growth of the un-induced (Tet -) and tetracycline-induced (Tet +) cell cultures throughout the course of five days. Cell concentration was assessed every 24 h, and plotted on a Log_{10} Y-axis graph. The individual points represent the mean concentration calculated from two independent experiments (n = 2), while the error bars depict the SEM. The red arrow pinpoints a 1:10 dilution of the Tet - culture. B) Efficiency of the knockdown of TbORC1B mRNA levels assessed by RT-qPCR. The results represented refer to the relative amount of mRNA levels to the non-induced sample (Tet -), calculated using the $\Delta\Delta C_t$ method. The mean of two independent experiments (n = 2) is shown, and the error bars represent SEM. C) Quantification of cells in the different cell cycle stages throughout the course of five days under tetracycline induction, based on the nuclear and kinetoplast configuration of the cells stained with DAPI. A minimum of 150 cells was counted per time point and experimental group (Tet - and Tet +), and percentages of each cell type (1N1K, 1N2K, 2N2K, 0N1K, and others) were calculated relative to the total amount of cells analysed. The graph represents the mean of each cellular population obtained in two independent experiments (n = 2), while the error bars refer to SEM. The Tet - bar represents the mean of the percentage of each cell type throughout the course of each of the two experiments (five days). D) Histograms representing the distribution of the cell population according to DNA content, stained with PI (FL2-A channel), assessed by flow cytometry. Approximately 30,000 cells were analysed per sample. Histograms from Tet - (full line) and Tet + (dashed line) cell cultures are shown overlapped for the 6 h, 12 h, 24 h and 48 h time points. The histograms of other time points are shown in the appendices, section 7.3.2. Results from only one experiment are shown. E) Percentage of EdU positive cells in the Tet + sample relative to the percentage of EdU positive cells in the Tet - culture from the respective time point. A minimum of 150 cells were analysed per time point and group (Tet - and Tet +). The mean from two independent experiments (n = 2) is shown; error bars represent SEM.

Silencing of TbORC1B was assessed by RT-qPCR at 6 h and 12 h post-RNAi (Figure 3.10, B); as before, later time points were not considered due to the potentially confounding influence of >50% of the population being zoids (Figure 3.10, C). A decrease of ~40-45% in the mRNA levels of TbORC1B relative to the non-induced samples (Figure 3.10, B) were seen, which is broadly comparable with TbORC1/CDC6 knockdown at 12 h and confirms the RNAi is on-target.

Together, the above results strongly suggest TbORC1B's involvement in DNA replication, with its silencing resulting in similar phenotypes to those obtained upon the RNAi of TbORC1/CDC6 or TbORC4. The major difference, however, lies in the time frame for the emergence of such DNA replication-related defects: while growth and cell cycle defects were only detected 48 h after RNAi induction targeting either TbORC1/CDC6 or TbORC4, the same anomalies were clearly noticeable in the TbORC1B RNAi cell line 24 h earlier, with DNA replication being severely affected as early as at 6 h post-tetracycline addition to the media. This faster appearance of the same phenotypes may suggest that TbORC1B plays a crucial role in DNA replication, even more necessary than the one played by TbORC1/CDC6 or TbORC4. *T. brucei* PCF cells are estimated to divide every 8.5 hours (the time to go through the cell cycle) (Woodward and Gull, 1990). Calculations (mathematical formula is shown in the materials and methods Chapter 2, Equation 2) suggest a doubling time of the non-induced TbORC1B

RNAi cell line to be 12-13 h (compared to 11-12 h for the 2913 parental cell line). The detection of such a severe impact on DNA replication 6 h after RNAi induction, which is ~half the time needed for the cell to complete a cell cycle, reinforces the rapidity of the effect, which data below suggests can be explained by differences in expression of TbORC1B relative to other *T. brucei* Orc-like factors during the cell cycle.

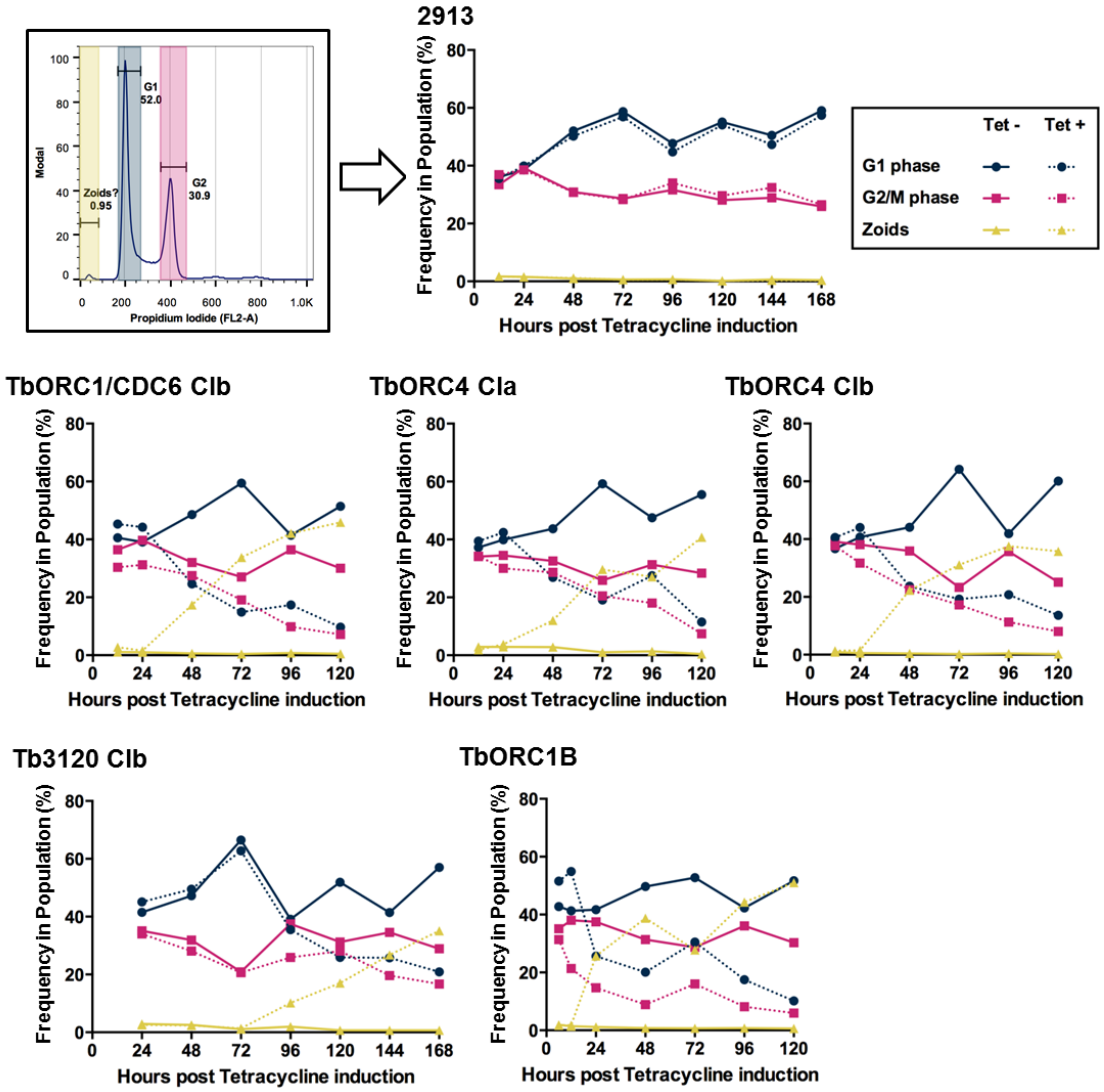


Figure 3.11. Summary of the flow cytometry results obtained for each cell line.
Each individual graph represents the summary of the data obtained by flow cytometry on DNA content, through the staining of cells with PI. The histogram within the box represents an example of a normal cell cycle profile (in this case, the 29-13 cell line, Tet - culture, time point of 48 h). Gates were drawn to allow the quantification of the percentage of the population in G1 or G2/M phases (blue and pink gates, respectively), and cells that contained sub-G1 DNA content signal suggesting the lack of a nucleus. These were thus labelled as zoids (yellow gate). The percentages given in each gate were plotted on a XY graph, as frequency in population (%) per hours post tetracycline (Tet) induction. Blue points represent cells in G1 phase, pink points G2 phase and yellow represents the population of zoids. In all cases, full line represents Tet -, and dashed line Tet +. The 29-13 graph represents the parental cell line Tet - and Tet + cultures throughout the experimental period, as a control. The remaining graphs, TbORC1/CDC6 Clb, TbORC4 Cla, TbORC4 Clb, Tb3120 Clb, and TbORC1B, represent the results obtained for those specific cell lines over time. The histograms and data used to generate these graphs are shown in the appendices, section 7.3.2.

3.4 Generating TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120 *in situ* – tagged cell lines

In order to try and better understand the different timing of the phenotypes resultant from the RNAi induction of the different factors described above, and to gain further insight into the roles of the putative Orc-like factors in PCF cells, the remaining sections of this chapter describe various sets of experiments exploring the different proteins' localisation, expression and interaction with TbORC1/CDC6, besides trying to understand whether there is an ORC-like complex in *T. brucei*.

In the absence of antibodies recognising any of the putative Orc-like proteins, epitope tagging of each protein was necessary, as it allows the detection of the epitope-tagged proteins using highly sensitive and specific antibodies that are commercially available against most used tags (Brizzard, 2008). In *T. brucei*, a wide range of molecular tools have been developed to allow the expression of proteins as translational fusions with tags such as the c-myc and the human influenza virus hemagglutinin (HA) (Oberholzer *et al.*, 2006; Kelly *et al.*, 2007; Alsford and Horn, 2008). One strategy encompasses the inducible over-expression of the tag-fused protein (either tagged at the N- or C-terminus) (Alsford and Horn, 2008; Kelly *et al.*, 2007), for instance using the 29-13 cell line described in section 3.3.1 (Wirtz *et al.*, 1999; Alsford *et al.*, 2005). Nevertheless, this strategy carries the risk that, in some cases, over-expression of the tagged protein may result in its mis-localisation and possible assembly into non-physiological complexes (Alsford and Horn, 2008; Shen *et al.*, 2001). Another strategy relies upon adding the epitope tag, either to the N- or C-terminus of the protein, *in situ* (Oberholzer *et al.*, 2006; Kelly *et al.*, 2007). In this approach one of the chromosomal copies of the gene is modified, leading to the expression of the tagged protein from the gene's endogenous locus, and, in principal, at normal physiological levels (Shen *et al.*, 2001; Oberholzer *et al.*, 2006; Kelly *et al.*, 2007; Alsford and Horn, 2008). In this strategy, endogenous tagging is achieved through homologous recombination between specifically generated cassettes or plasmids (with part of the gene of interest fused to the tag) and the endogenous locus after transformation. Because the tag is inserted fused to the gene at its endogenous locus, this strategy can be applied in different cell line backgrounds, including the ones used for inducible RNAi

(Oberholzer *et al.*, 2006). In this way, it is possible to assess the knockdown efficiency of the inducible RNAi experiment through the detection of the corresponding tagged protein throughout the assay (strategy employed in section 3.3.2.3). Nonetheless, it is important to keep in mind that there is always the risk that the fusion of the protein with a tag might interfere with the protein's structure and thus potentially alter its function and/or localisation (Brizzard, 2008). It is therefore important to ensure that the tagged version of the protein is indeed functional by, for example, deleting the remaining endogenous allele of the gene (Kelly *et al.*, 2007).

3.4.1 Cloning constructs for endogenous tagging with c-myc

In order to better understand the TbORC1/CDC6 interacting factors, each was tagged at its endogenous locus, either with c-myc or HA, essentially as described previously in (Tiengwe *et al.*, 2012b). In order to insert a c-myc epitope tag in the C-terminus of TbORC1B, TbORC4, Tb7980, Tb3120, and Tb1120, the construct pNAT^{x12M} (Alsford and Horn, 2008), containing 12 copies of the c-myc tag (12myc), and the blasticidin (*BSD*) resistance gene (the latter flanked by *T. brucei* tubulin and actin processing sequences), was used. This construct has been reported to generate fusion proteins that show very little cross-reactivity with native proteins when detected with commercially available anti-c-myc antibodies, and the 12myc tag has never been reported to display toxicity or to have cryptic localisation signals that might influence the tagged protein localisation (Alsford and Horn, 2008). Besides, it had been previously successfully used for the generation of the TbORC1/CDC6 C-terminal 12myc-tagged cell line (Tiengwe *et al.*, 2012b). A schematic representation depicting the strategy employed to generate the tagging constructs is shown in Figure 3.12, A. Briefly, the 3' end of the open reading frame (ORF) of the gene of interest (GOI) was PCR-amplified using primers containing specific restriction sites that allow the insertion of this fragment into the pNAT^{x12M} plasmid backbone. In all cases, when designing the reverse primer, the gene's stop codon was excluded, allowing the translation of the fusion of the upstream ORF fragment and the epitope. In addition, the amplified PCR product possessed a unique restriction site that is not present in the plasmid backbone, allowing linearization by enzymatic digestion prior to transformation into the parasite, and driving homologous integration into the gene.

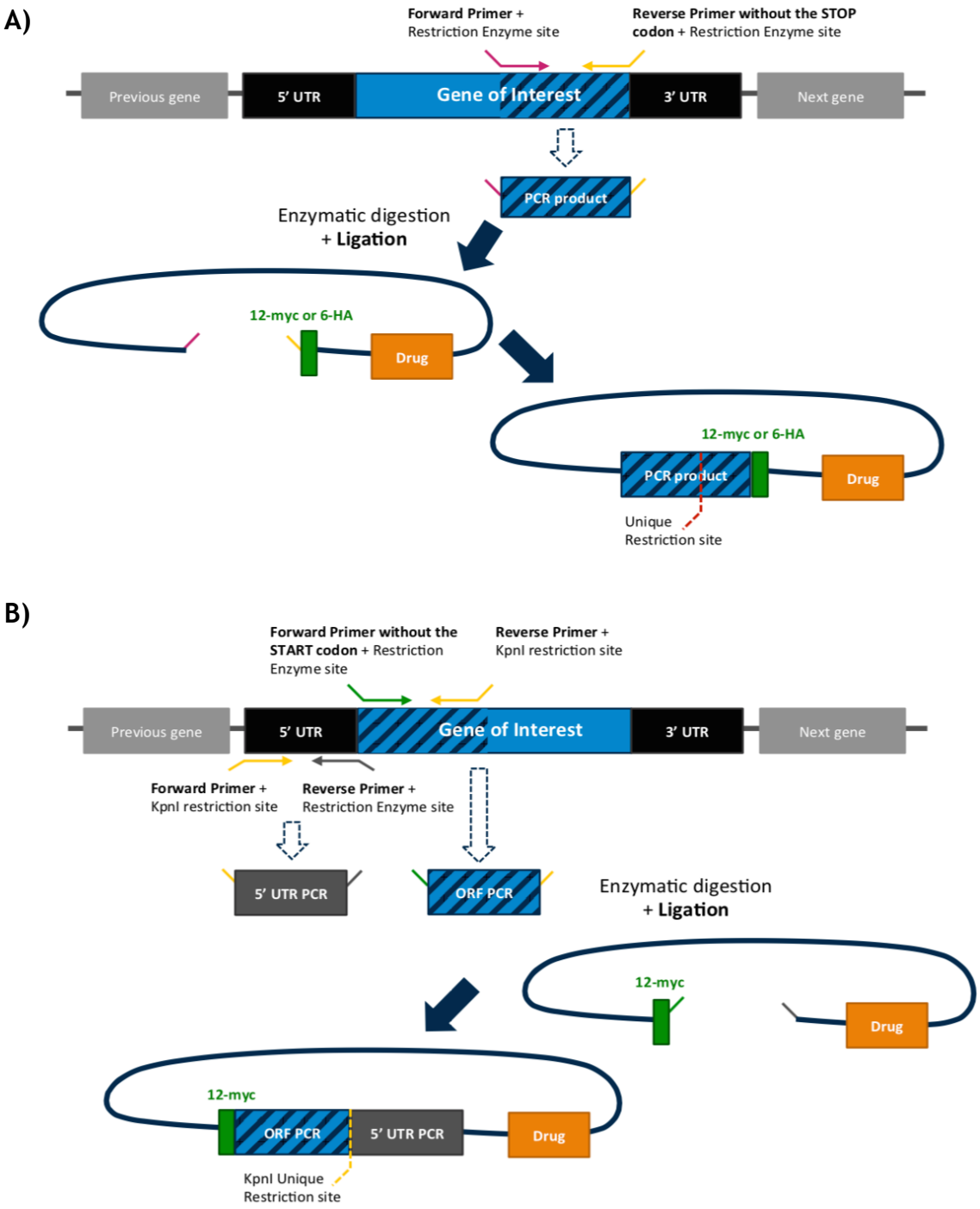


Figure 3.12. Cloning strategy used to generate the plasmids for endogenous tag with N- or C-terminal 12myc, and C-terminal 6HA.
Schematic representation of the cloning strategy used to generate the C- terminal 12myc or 6HA tag (A), or N-terminal 12myc tag (B). 5' UTR (untranslated region) refers to the 5' intergenic region of the gene of interest, as the 5' UTR regions of the genes had not been annotated in TriTrypDB at the time these constructs were designed and conceived. ORF refers to open reading frame. Diagrams are not to scale.

At the start of this project, Dr Calvin Tiengwe had already generated constructs for the 12myc C-terminal tagging of TbORC1/CDC6, TbORC4, Tb7980 and Tb3120 during the course of his PhD (plasmid repository numbers pBM3, pBM35, pBM37, and pBM36, respectively). However, with the exception of pBM3, none of the plasmids had been confirmed by either enzymatic digestion or sequencing.

Because these constructs had to be re-made for transfection of Lister 427 strain cell lines (explained in Chapter 4, section 4.3.1), the cloning strategies employed for the tagging of TbORC4, Tb7980 and Tb3120 are also described below, as well as the strategies for tagging TbORC1B and Tb1120. Full details are in the materials and methods, Chapter 2, section 2.2.1.

A fragment of 455 base pairs (bp) from the 3' region of the *TbORC1B* ORF was PCR-amplified from TREU 927 genomic DNA (gDNA) using the forward primer C5, containing the HindIII restriction site, and the reverse primer C6, containing the XbaI restriction sequence. Similarly, a fragment of 638 bp was amplified from *TbORC4*, using primers CTOL_65[g] (HindIII site) and CTOL_66 (XbaI site); a region of 639 bp of *Tb7980* was amplified using primers CTOL_69 (HindIII site) and CTOL_70 (XbaI site); a region of 659 bp of *Tb3120* was amplified using the C120 (HindIII site) and C121 (XbaI site); and a section of 1037 bp from *Tb1120*, was amplified using the primers C3 (HindIII site) and C4 (XbaI site). After purification, the PCR products and the pNAT^{x12M} plasmid were digested with both HindIII and XbaI, further purified, ligated and transformed into competent DH5α *E. coli* bacteria (details in materials and methods, Chapter 2, section 2.2.1). The resulting plasmids, TbORC1B-12mycBSD, TbORC4-12mycBSD, Tb3120-12mycBSD and Tb1120-12mycBSD (maps shown in Figure 3.13, B, D, E and F), were then purified from the bacteria, and confirmed by restriction enzyme digestion, using two combinations of enzymes: HindIII and XbaI, for confirmation of the gene fragment, and XbaI with BamHI, to confirm the presence of the 12myc tag, which is 445 bp in size (Figure 3.13, G). The TbORC1/CDC6-12mycBSD plasmid (Tiengwe *et al.*, 2012b) map and confirmation digestion are also shown in Figure 3.13 (A and G). All plasmids were further sequenced using the HA/myc primers pair (MP 15[h] and MP 16, sequence shown in materials and methods, Chapter 2, Table 2-2) and, with the exception of the Tb7980 C-terminal tag construct (not shown, but discussed further below), all were confirmed to have the correct sequence of the gene insert.

[g] Primers labelled as CTOL were designed by Dr Calvin Tiengwe during the course of his PhD, and can be found in Tiengwe *et al.*, 2012, as well as in his thesis, Tiengwe, 2010.

[h] Dr Marko Prorocic designed primers labelled as MP during the course of his Post-Doc in Dr Richard McCulloch's lab.

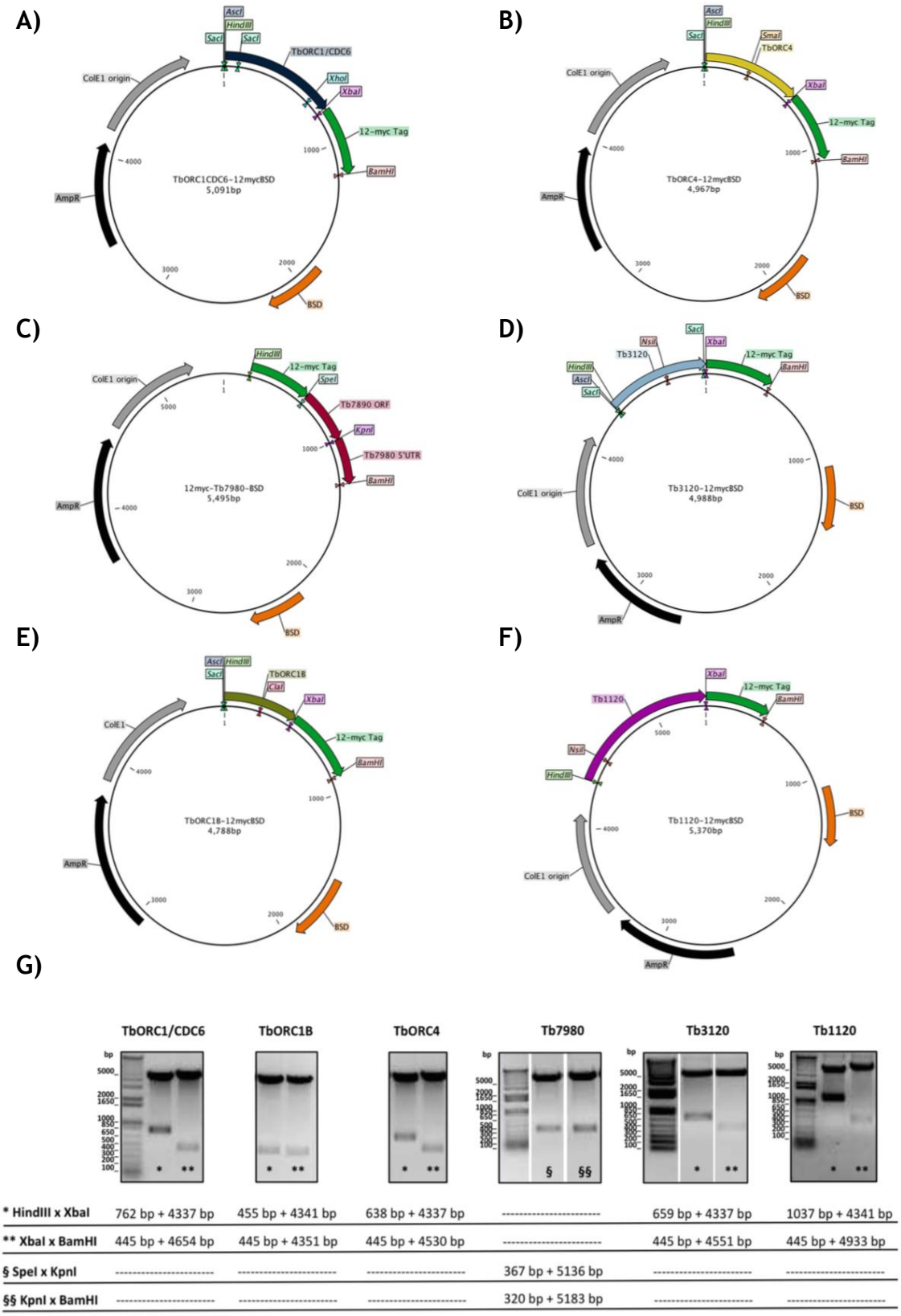


Figure 3.13. Plasmid maps and confirmation enzymatic digestions of the constructs used for endogenous tagging with N- or C-terminal 12myc.
Detailed plasmid maps of the constructs used for endogenous tagging with 12myc at the C-terminus (A, B, D, E, and F), or at the N-terminus (C). The 12myc tag is represented in green, the bacterial ampicillin resistance gene is shown as *amp*R and in black, and the bacterial origin depicted as ColE1 origin and in grey. Each gene region cloned into the construct is shown and labelled with the gene name used throughout this work, and the drug resistance gene, used for parasite clone selection, blasticidin, is represented as *BSD* and in orange. Relevant restriction sites are also shown. All constructs were confirmed through restriction enzyme digestions, shown in G, with more detailed description in the main text. In each case, the 1 Kb plus DNA ladder

(Invitrogen™) is shown as a size reference and the expected sizes, resulting from the different enzymatic digestions, are shown below each gel image: for C-terminal 12myc tag plasmids, (*) digestion with HindIII and XbaI results in the excision of the gene fragment cloned into the plasmid; (**) digestion with XbaI and BamHI results in the excision of the 12myc tag from the plasmid; for the N-terminal 12myc tag plasmids, (§) digestion with SpeI and KpnI excises the ORF region cloned into the plasmid, while (§§) digestion with KpnI and BamHI results in the excision of the 5'UTR fragment cloned into the plasmid. Note that in some cases the same gel image has been cropped to display the results from the same plasmid side-by-side, and are separated by a white space.

Generation of a construct to be used for Tb7980 C-terminal 12myc tagging was attempted several times, but was never successfully recovered (data not shown). A different forward primer (C103), containing a SacI restriction site already present in the gene sequence, was therefore designed and used for cloning, adapting the strategy above with the SacI restriction site instead. Although it was possible to obtain the correct plasmid (confirmed both through enzymatic digestion and sequencing; data not shown), it was never possible to recover any parasites after transfection with this construct (data not shown). To circumvent these problems, it was decided to tag Tb7980 endogenously with 12myc at the N-terminus instead. For this, a modified version of the pEnT6B construct (Kelly *et al.*, 2007), generated by Dr Anna Trenaman during the course of her PhD (Trenaman, 2012), and containing the *BSD* selection marker, was used. A schematic representation of the strategy used is shown in Figure 3.12, B. In this case, to allow homologous recombination into the parasite's genome, two fragments are PCR-amplified: a region from the 5' end of the gene's ORF, beginning at the second codon (and thus excluding the start codon), and a section from upstream of the ORF start codon and in the 5' intergenic region (Kelly *et al.*, 2007). As before, all pairs of primers contain restriction enzyme sites, but the key aspect in this cloning strategy is that the unique restriction site that will be used for the linearization of the plasmid prior to transfection (KpnI in this case), is included in the reverse primer used for the amplification of the gene's ORF fragment, as well as in the forward primer used for the amplification of the 5' intergenic region (Figure 3.12) (Kelly *et al.*, 2007; Trenaman, 2012). The two PCR fragments, as well as the plasmid, are then digested with the appropriate restriction enzymes, and simultaneously cloned in a "three-piece ligation" reaction. In this case, a 367 bp 5' region of the *Tb7980* ORF was PCR-amplified using the primer pair C186 (containing the SpeI restriction site) and C187 (KpnI site), while the 5' intergenic region upstream of the *Tb7980* ORF was amplified using the C185 (containing the KpnI site) and C184

(with the BamHI site) primer pair (sequences shown in the materials and methods, Chapter 2, Table 2-1). After cloning and purification from bacteria, the resulting plasmid, 12myc-Tb7980-BSD (map depicted in Figure 3.13, C), was confirmed by restriction enzyme digestion using the SpeI and KpnI combination, to extract the ORF fragment, as well as KpnI and BamHI, to extract the 5' intergenic region (shown in Figure 3.13, G). In addition, the cloned regions were further confirmed through sequencing using the AT120[i] and AT121 primers (shown in materials and methods, Chapter 2, Table 2-2).

3.4.2 Cloning constructs for endogenous tagging with HA

To analyse protein interactions, *T. brucei* cell lines expressing two proteins endogenously tagged with different tags have been generated in the past (Dang and Li, 2011; Tiengwe *et al.*, 2012b). One successful combination encompasses the use of the c-myc and HA tags, and it has been used to confirm the interaction of TbORC1/CDC6 (tagged with 12myc) with TbORC4, Tb7980 and Tb3120 (each tagged individually with HA) (Tiengwe *et al.*, 2012b). Although it was identified has a strong hit following the mass-spectrometry analysis of TbORC1/CDC6 immunoprecipitation (IP), the predicted interaction between Tb1120 and TbORC1/CDC6 was never confirmed (Tiengwe, 2010). Therefore, it was decided to endogenously tag Tb1120 at the C-terminus with 6 copies of HA (6HA), using a variant of pNAT^{x12M} where the 12myc tag and the blasticidin resistance gene have been replaced, respectively, with a 6HA tag and the bleomycin resistance gene (*BLE*) (Tiengwe *et al.*, 2012b). TbORC1B was previously shown to interact with TbORC1/CDC6 using a different combination of tags (Dang and Li, 2011): 3HA and PTP. To provide comparable data with all other putative Orc-like factors, it was also decided to tag TbORC1B with 6HA. The cloning strategy used was identical to the one described for C-terminal tagging of proteins with 12myc in the previous section (shown in Figure 3.12). Because the 6HA plasmid is a derivation of pNAT^{x12M}, the restriction sites used for cloning remained the same and, therefore, the cloning process was the same as described in section 3.4.1. Plasmid maps of the TbORC1B-6HABLE and Tb1120-6HABLE constructs are shown in Figure 3.14, together with the confirmation

[i] Primers labelled as AT were designed by Dr Anna Trenaman during the course of her PhD, and can be found in her thesis, Trenaman, 2012.

digestions: HindIII and XbaI to extract the gene insert, and XbaI in combination with BamHI to isolate the 6HA tag, of 211 bp in size.

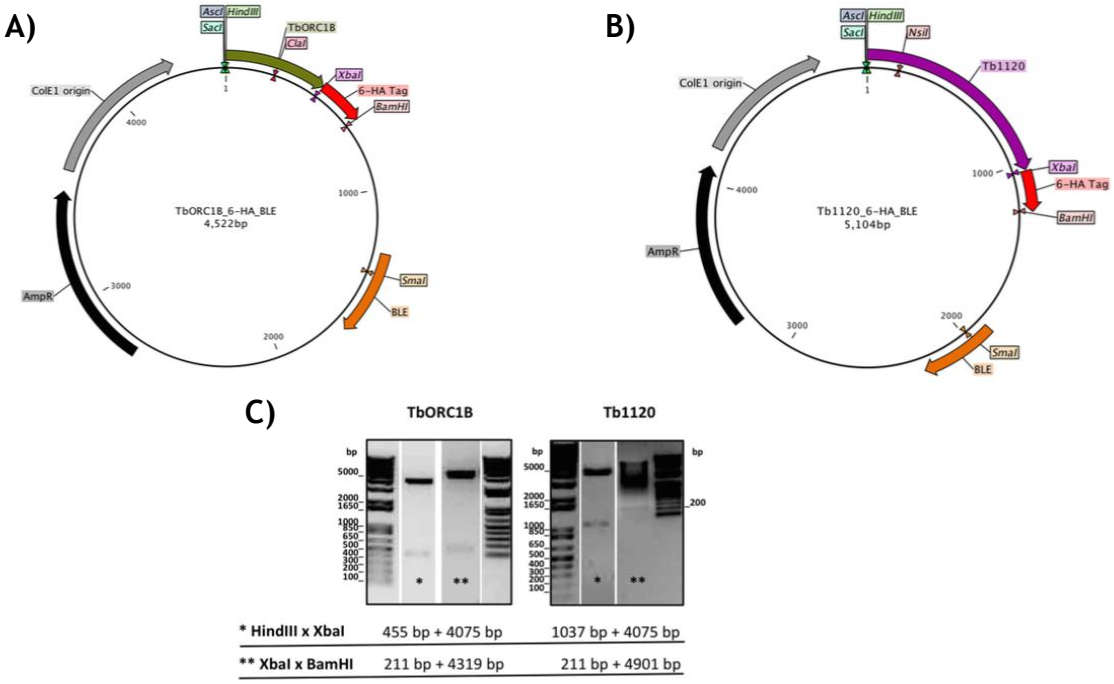


Figure 3.14. Plasmid maps and confirmation enzymatic digestion of the constructs used for endogenous tagging of proteins with C-terminal 6HA tag.
Detailed plasmid maps of the constructs used for endogenous tag TbORC1B (A) and Tb1120 (B) with 6HA at the C-terminus. The 6HA tag is represented in red, the bacterial drug resistance gene, against ampicillin, is shown as *amp*R and in black, and the bacterial origin in depicted as ColE1 origin and in grey. Each gene region cloned into the construct is shown and labelled with the gene name used throughout this work, and the drug resistance gene, used for parasite clone selection, bleomycin, is represented as *BLE* and in orange. Relevant restriction sites are also shown. The constructs were confirmed through restriction enzyme digestions, shown in C, with more detailed description in the main text. In each case, the 1 Kb plus DNA ladder (Invitrogen™) is shown as a size reference, and the expected sizes, resultant from the different enzymatic digestions, are shown below each gel image: (*) digestion with HindIII and XbaI results in the excision of the gene fragment cloned into the plasmid; (**) digestion with XbaI and BamHI results in the excision of the 6HA tag from the plasmid. Note that in the final images shown, the same gel image has been cropped so to display the results from the same plasmid side-by-side, and are separated by a white space.

3.4.3 Cloning constructs for *in situ* gene disruption

As briefly mentioned in section 3.4, it is important to confirm whether the protein fused with the tag is functional. The most commonly used strategy is to delete one allele and then tag the remaining allele, which will result in the only protein expressed in the cell being the tagged one (Kelly *et al.*, 2007). This has been successfully achieved for TbORC1/CDC6 (Tiengwe *et al.*, 2012b), and it relies on the transfection of a construct that, by homologous recombination with the gene's 5' and 3' intergenic flanking regions, will lead to deletion of the

endogenous gene by replacing it with a selective drug marker gene. In this study, a plasmid modified by Dr Marko Prorocic from the pmtl23 plasmid (gift from Professor Marshall Stark, University of Glasgow), so to harbour a drug resistance gene between two multiple cloning sites, was used (a detailed plasmid map is shown in the appendices, Figure 7.34). Several variations of pmtl23 have been generated, so to have puromycin (*PURO*), blasticidin (*BSD*), bleomycin (*BLE*), or the neomycin (*NEO*) resistance genes, and thus allow a wide choice of selective drugs that can be used. Like other constructs used for gene deletion (e.g. the modified pBluescript KS plasmid used for deletion of one of TbORC1/CDC6 alleles (Tiengwe *et al.*, 2012b), which is shown in Figure 3.16, A), a region of approximately 450-600 bp is PCR-amplified from the gene ORF's immediate 5' and 3' intergenic regions (here referred to as untranslated regions, UTRs), with primers containing specific restriction enzymes sites (Figure 7.34), to allow cloning into the plasmid flanking the drug resistance gene (Figure 3.15). Due to the presence of multiple cloning sites on either end of the selective drug marker, forward and reverse primers can be designed such that four different enzymes are used for the digestion of the 5' and 3' UTR fragments prior to ligation into the plasmid backbone (two combinations, two enzymes each PCR fragment). The parent plasmid is digested, separately, twice: once with the enzyme present in the forward primer of the 5'UTR fragment and with the enzyme in the reverse primer of the 3'UTR PCR product, generating the plasmid backbone; and once with the enzyme present in the reverse primer of the 5'UTR fragment, together with the enzyme in the forward primer used to amplify the 3'UTR region, which will result in excision of a fragment containing the drug selective marker gene. The two isolated plasmid fragments are therefore incompatible, as they were excised using different enzyme combinations and are not able to re-ligate with one another. The 5' and 3' UTR PCR products, purified after enzymatic digestion, are then added to the two plasmid fragments, and simultaneously cloned in a "four-piece ligation" reaction, which will only result in full length plasmids (and therefore bacterial colonies) if the four fragments are ligated to each other in the correct order and orientation. This strategy allows cloning of two PCR products into the plasmid in one cloning step, rather than consecutive ligation reactions. Another advantage of using this construct lies in the fact that both the forward primer used for the 5' UTR PCR amplification and the reverse primer used for the 3' UTR PCR contain, in the

cases here, a NotI restriction site, allowing excision of a fragment encompassing the two UTRs and the drug resistance cassette (to be used for transfection into the parasite cells) using a single restriction enzyme. Because this site is included in the primers, other enzyme sites can be used, giving this construct a great flexibility regarding the choice of restriction sites that can be used prior to transfection into the parasite cells.

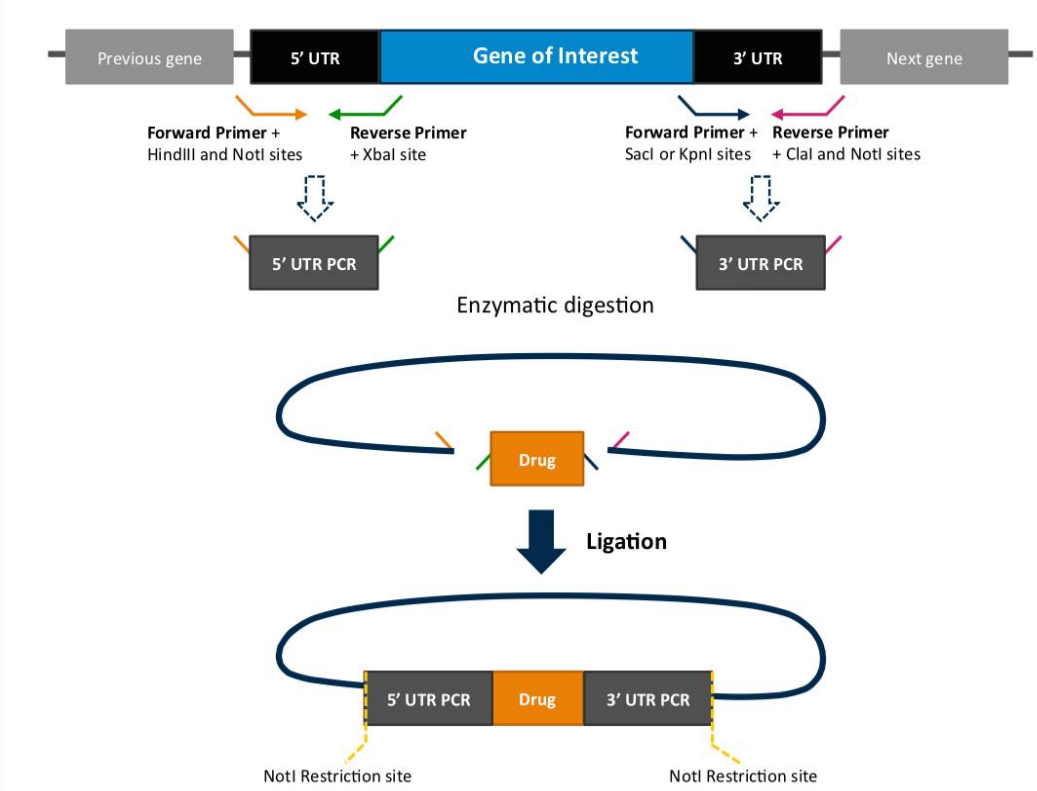


Figure 3.15. Cloning strategy to generate constructs for gene deletion.
Schematic representation of the cloning strategy used to generate the gene deletion (KO) constructs from the modified pmtl23 plasmid (a detailed map of the modified pmtl23 plasmid is shown in the appendices, Figure 7.34). The 5' UTR (untranslated region) refers to the 5' intergenic region of the gene of interest, while the 3' UTR refers to the 3' intergenic region of the gene, as both the 5' and 3' UTR regions of the genes had not been annotated in TriTrypDB at the time these constructs were designed and generated. The selective drug can be blasticidin, puromycin, bleomycin or neomycin, depending on the variation of the plasmid used. In the presented example, two combinations of restriction enzymes that can be used are depicted (HindIII and XbaI; SacI/KpnI and ClaI sites). The NotI restriction sites are inserted with the primers to allow excision of the fragment containing the drug selective marker flanked by the 5' and 3' UTR regions to be used for parasite cells transfection. Representations herein shown are not to scale.

In the present work, only the variation of the pmtl23 plasmid containing the neomycin resistance gene (pmtl23-NEO) was used. For cloning of the constructs targeting *TbORC1B*, *TbORC4*, *Tb7980* and *Tb3120* (Figure 3.16, B-E), the 5' UTR region was amplified using primer pairs containing the HindIII and NotI restriction sites in the forward primer and the XbaI site in the reverse one (*TbORC1B* primers C201 and C202; *TbORC4* primers C189 and C190; *Tb7980*

primers C193 and C194; and *Tb3120* primers C197 and C198), while the 3' UTR fragment was amplified using a forward primer with the SacI site in conjugation with a reverse primer containing both the NotI and ClaI restriction sites (*TbORC1B* primers C203 and C204; *TbORC4* primers C191 and C192; *Tb7980* primers C195 and C196; and *Tb3120* primers C199 and C200). All primer sequences are shown in the material and methods Table 2-1. For the construct targeting *Tb1120* (Figure 3.16, F), the 5' UTR region was amplified as described above (primers number C205 and C206), while the 3'UTR fragment, which has a SacI site, was amplified using a forward primer containing a KpnI restriction site instead (primer C207), and a reverse primer containing both the NotI and ClaI restriction sites (primer C208). All plasmids were confirmed by enzymatic digestion using the NotI restriction enzyme (Figure 3.16, G), as well as sequencing, using the standard vector primers M13 uni (-43) and M13 rev (-49) (sequences shown in material and methods, Chapter 2, Table 2-2). For simplicity, the resulting plasmids are here referred to as TbORC1B-KO-NEO, TbORC4-KO-NEO, Tb7980-KO-NEO, Tb3120-KO-NEO, and Tb1120-KO-NEO (*NEO*, neomycin resistance gene). The construct generated by Dr Calvin Tiengwe targeting TbORC1/CDC6 chromosome locus (Tiengwe *et al.*, 2012b) is here referred to as TbORC1/CDC6-KO-PURO (*PURO*, puromycin resistance gene).

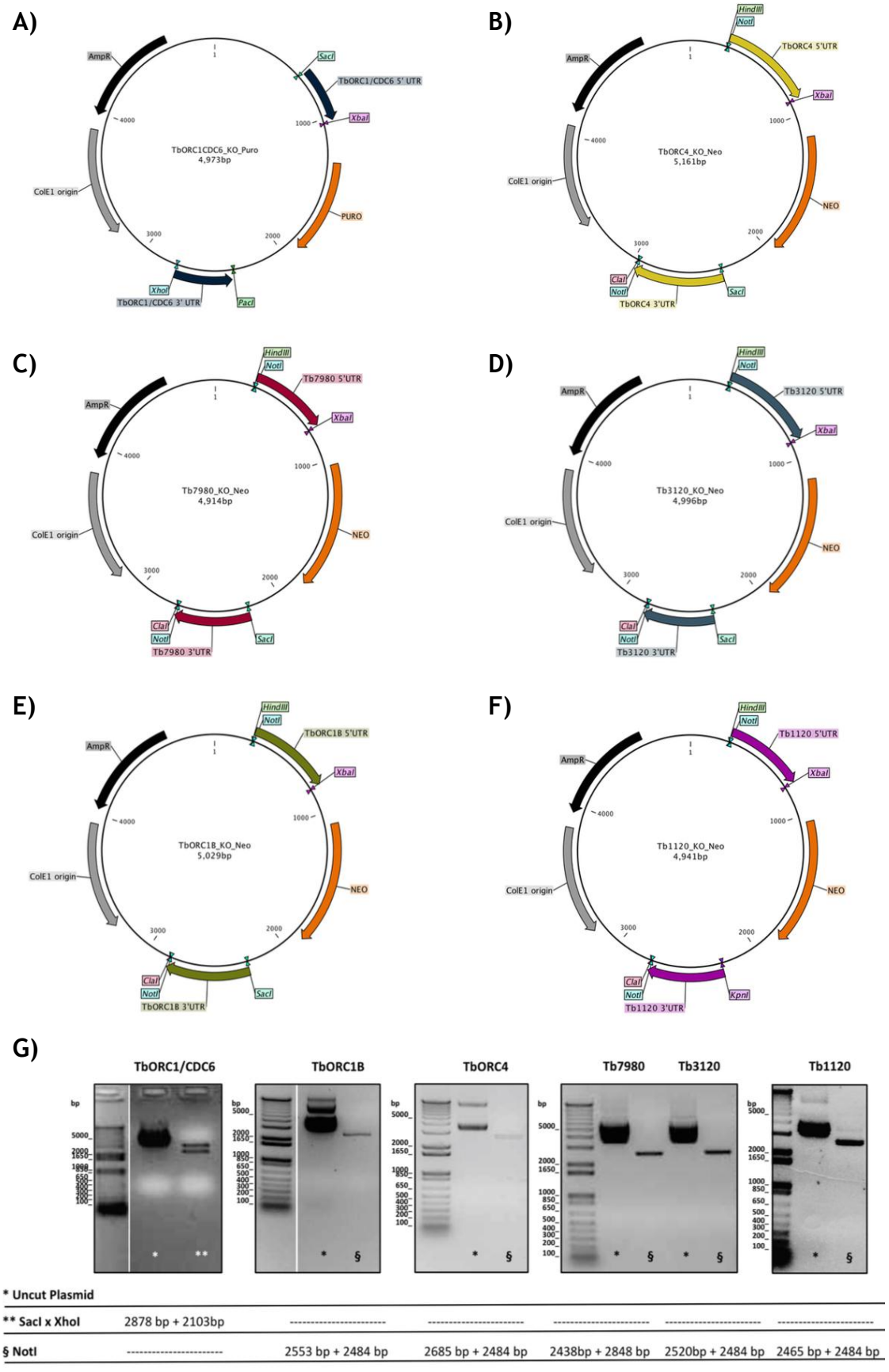


Figure 3.16. Plasmid maps of the constructs used for TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120 gene deletion.

Detailed plasmid maps of the constructs used for gene deletion of one allele of *TbORC1/CDC6* (A), *TbORC4* (B), *Tb7980* (C), *Tb3120* (D), *TbORC1B* (E) and *Tb1120* (F). Each gene's 5' and 3' UTR (intergenic) regions cloned into the construct are shown and labelled with the gene name used throughout this work, and the drug resistance genes, used for parasite clone selection, puromycin (*PURO*) in A, and neomycin (*NEO*) in B-F, are represented as in orange. The bacterial drug

resistance gene, against ampicillin, is shown as *amp*R and in black, and the bacterial origin in depicted as ColE1 origin and in grey. Relevant restriction sites are also shown. All constructs were confirmed through restriction enzyme digestions and are shown in G, with more detailed description in the main text. In each case, the 1 Kb plus DNA ladder (Invitrogen™) is shown as a size reference, and the expected sizes, resultant from the different enzymatic digestions, are shown below each gel image. Note that in some cases the same gel image has been cropped so to display the results from the same plasmid side-by-side, and are separated by a white space.

3.4.4 Generation of TbORC1B 12myc, TbORC4 12myc, Tb3120 12myc and Tb1120 12myc PCF cell lines

To allow tagging, the constructs are linearised and transfected into *T. brucei* cells. Once transfected, the constructs can integrate into the target gene through homologous recombination, guided by the 3' region of the ORF in the construct (Figure 3.17). As a result, the 12myc tag is incorporated in-frame into the 3' end of the gene's ORF, together with the remainder of the construct. This means that while the 5' UTR remains unchanged, the 3' UTR of the gene is no longer the native one (Figure 3.17, iv), having been replaced by the *T. brucei* ßα tubulin mRNA processing signal (5'), the drug resistance marker gene, and the *T. brucei* actin mRNA processing signal (3'), allowing the correct splicing of the drug resistance gene (Oberholzer *et al.*, 2006; Kelly *et al.*, 2007). In addition, it also results in the duplication of the gene fragment used in the construct downstream of the tagged locus (Kelly *et al.*, 2007). However, because this does not comprise the whole gene, and is lacking the start codon, it is presumed that this is inactive (represented in Figure 3.17, iv), as "Incomplete ORF"). In this tagging strategy, expression of the targeted gene can be altered if the control for expression is dependent on the endogenous 3' UTR, for instance through mRNA stability changes during the cell or life cycle. Estimates of mRNA abundance during the cell cycle have been made genome-wide following elutriation-based cell cycle synchronisation (Archer *et al.*, 2011), and these data are available on TriTrypDB. For each of the Orc-like genes, these data were recovered (shown in the appendices, Figure 7.35), and only TbORC1B mRNA levels displayed any evidence for mRNA variation, with ~2-fold increased abundance in 'late G1' (3 hours after synchronisation). However, whether this effect is dictated by the gene's 3'UTR or is related to replication is unknown.

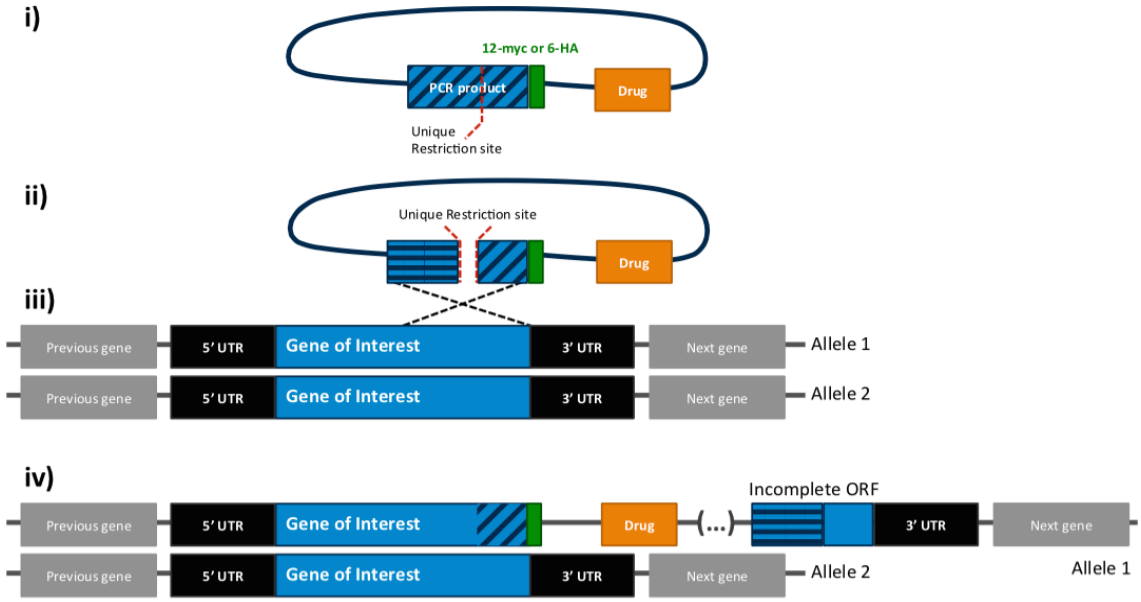


Figure 3.17. Integration of the C-terminal 12myc tag constructs into the genome by homologous recombination.
Simplified schematic representation of the process followed to tag a gene endogenously with 12myc tag. i) General representation of the C-terminal tagging plasmid, with the unique restriction site highlighted in red, approximately in the middle of the gene fragment cloned into the plasmid ("PCR product", and shown in blue). In this illustration, the 12myc tag is represented in green, and the selective drug resistance gene is shown in orange, while the other features of the plasmid are not shown. ii) After digestion of the plasmid with the unique restriction site enzyme, the linearised plasmid is transfected into the parasite cells. iii) The linearised plasmid then recombines with the 3' region of the endogenous gene by homologous recombination, resulting in the endogenously tagged gene locus, as depicted in iv). Diagrams are not to scale. The same strategy applies to the integration of the constructs designed for C-terminal endogenous tagging with 6HA.

For linearization, TbORC1B-12mycBSD was digested with ClaI, TbORC4-12mycBSD with SmaI, and both Tb3120-12mycBSD and Tb1120-12mycBSD were individually digested with NsiI. Each linearised plasmid was individually transfected into PCF cells, strain TREU 927 (see materials and methods, Chapter 2, section 2.3.4) and transformants were selected with 10 μg.ml^{-1} of blasticidin. The resulting cell lines are referred to throughout this work as TbORC1B 12myc, TbORC4 12myc, Tb3120 12myc, and Tb1120 12myc. Confirmation of correct integration of the constructs by PCR and of the expression of the 12myc fused proteins by western blot is shown, respectively, in sections 3.4.8 and 3.4.9.

3.4.5 Generation of cell lines co-expressing both TbORC1/CDC6^{12myc} and TbORC1B^{6HA}, or Tb1120^{6HA}

For co-localisation and co-immunoprecipitation studies, the 6HA C-terminal tagging constructs for TbORC1B and Tb1120 were transfected into PCF TREU 927 cell lines that already had one allele of TbORC1/CDC6 tagged with 12myc and

the other replaced with a puromycin resistance gene cassette (Tiengwe *et al.*, 2012b). This was performed as described for the C-terminal 12myc tag constructs and transformants were confirmed by PCR and western blot as described in sections 3.4.8 and 3.4.9. These lines are referred to henceforth as TbORC1/CDC6 -/12myc TbORC1B 6HA and TbORC1/CDC6 -/12myc Tb1120 6HA cell lines. The TbORC1/CDC6 -/12myc Tb1120 6HA cell line was generated with the help of Emily Ross, during her undergraduate project.

3.4.6 Generation of 12myc Tb7980 cell line

For N-terminal 12myc tagging of Tb7980, the 12mycTb7980-BSD construct was linearised with KpnI (Figure 3.13), and transfected into PCF TREU 927 cells. Transformants were selected with 10 μg.ml^{-1} of blasticidin. In this strategy, the 5' UTR and the 12myc fused ORF regions in the construct allow its integration into the gene allele (Figure 3.18). As a result, the 12myc tag is incorporated in the 5' of the ORF, meaning that the gene's 3' UTR is not modified. The gene's 5' UTR is also still present, but is separated from the gene's ORF by the drug resistance gene (itself flanked at the 5' and 3' by the *T. brucei* ßα tubulin and actin mRNA processing signals), as well as the rest of the construct. Transformants recovered were confirmed by PCR and western blotting, as discussed in sections 3.4.8 and 3.4.9.

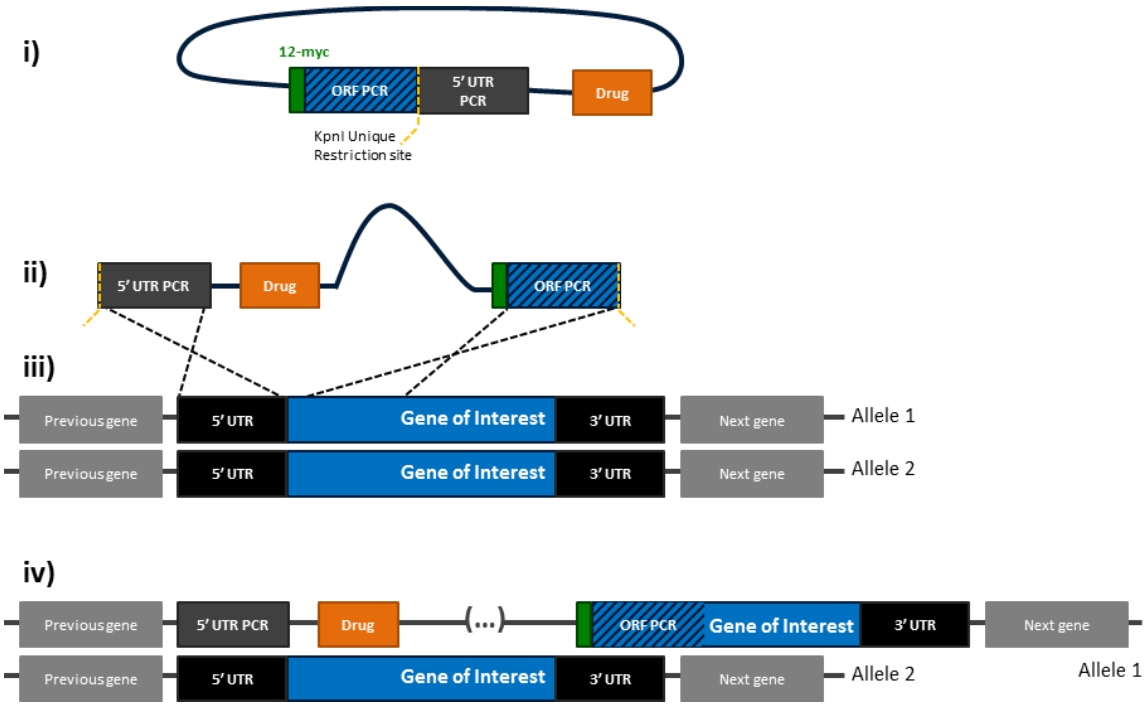


Figure 3.18. Integration of N-terminal 12myc tag constructs into the genome by homologous recombination.

Simplified representation of the process followed to tag a gene endogenously with a N-terminal 12myc tag. i) General representation of the N-terminal tagging plasmid, with the 12myc tag (green) fused with the cloned region of the gene ORF (shown in blue and as "ORF PCR"), which is separated from the 5'UTR (in grey, "5'UTR PCR"), by a KpnI restriction site (highlighted in yellow). In addition, the selective drug resistance gene is shown in orange, while the other features of the plasmid are not shown. ii) After digestion with KpnI, the plasmid is linearised and transfected into the parasite cells. iii) The linearised plasmid recombines with the endogenous gene locus, both at its 5'UTR and the 5' end of the ORF, through homologous recombination, resulting in the endogenously tagged gene locus, as depicted in iv). Representations are not to scale.

3.4.7 Generation of cell lines with one allele tagged and the other allele deleted

To attempt to confirm functionality of the tagged proteins, the gene deletion constructs (section 3.4.3, Figure 3.15 and Figure 3.16) were transformed into cognate cell lines in which the gene had previously been tagged in one allele. In this strategy, the gene's 5' and 3' UTRs were cloned flanking the drug resistance gene (already flanked by the *T. brucei* ßα tubulin and actin mRNA processing signals, which allows the trans-splicing and polyadenylation of the drug resistance gene once integrated into the parasite's genome), allowing the integration of the construct, by homologous recombination with the native allele's 5' and 3' UTRs, into the parasite's genome, leading to the complete replacement of the endogenous gene ORF with the drug resistance marker.

For transfection, TbORC1B-KO-NEO, TbORC4-KO-NEO, Tb7980-KO-NEO, Tb3120-KO-NEO, and Tb1120-KO-NEO were digested with NotI (shown in Figure 3.16), and transfected into PCF strain TREU 927 cells confirmed to have one allele endogenously tagged with 12myc (see sections 3.4.8 and 3.4.9); transformants were selected with 10 $\mu g.ml^{-1}$ of geneticin (G418®). However, despite transfecting all constructs, correct integrants were only recovered with Tb7980-KO-NEO, generating the cell line referred to here as Tb7980 -/12myc (confirmation is shown in sections 3.4.8 and 3.4.9). In case this reflected impaired function of the tagged proteins, all constructs were also transfected into non-tagged TREU 927 cells, but this was again unsuccessful. The reasons for this remain unclear, but incorrect constructs for Tb3120-KO-NEO and Tb1120-KO-NEO, and lethal effects of tagging these proteins, can be ruled out, since integrants were successfully obtained in bloodstream form *T. brucei* cells (discussed in Chapter 4). Nonetheless, attempts to mutate TbORC4 and TbORC1B, both in PCF and BSF cells, failed despite various transformation

attempts. This might suggest that both copies of these genes are needed, a hypothesis that is potentially supported by the results obtained through RNAi, as reduction in less than 50% of the mRNA levels of either TbORC4 or TbORC1B led to a severe growth phenotype (sections 3.3.2.3 and 3.3.2.5, respectively).

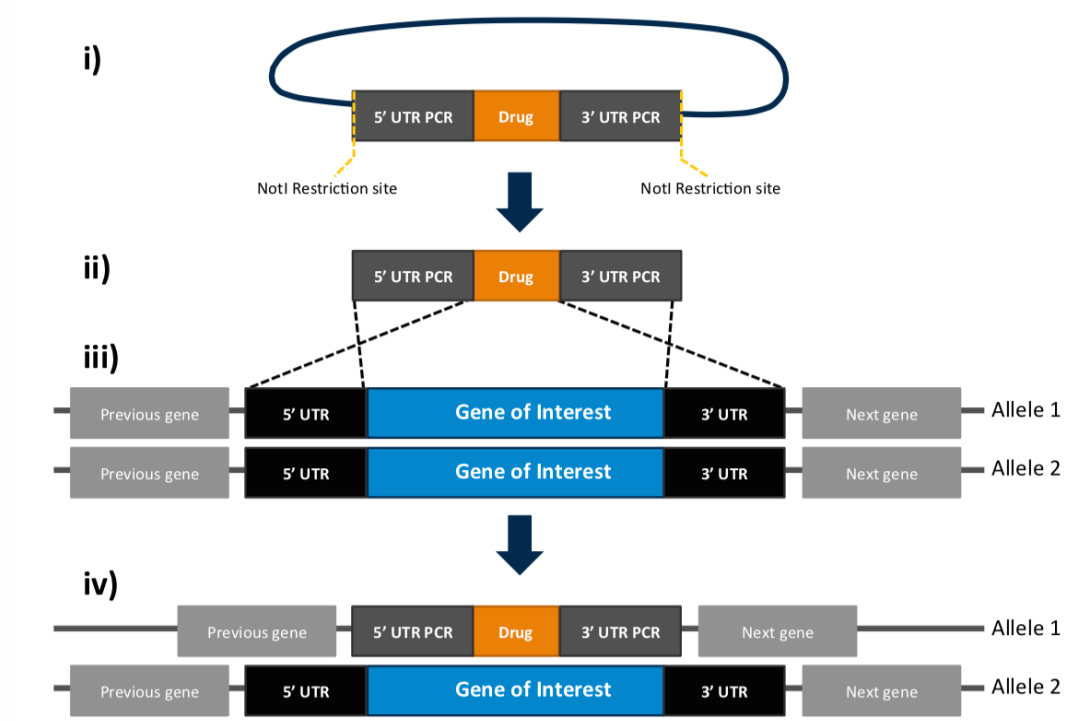


Figure 3.19. Schematic representation of the replacement of a gene copy by a drug resistance gene through the integration of the KO plasmids by homologous recombination. Schematic representation of the process followed to delete a copy of the gene, by replacing the endogenous gene ORF with a drug resistance gene. i) General representation of the gene deletion (KO) plasmid, with the NotI restriction sites (yellow), flanking the gene's 5' and 3' UTRs (in grey), enclosing the selective drug resistance gene (orange). For simplicity, other features of the plasmid are not depicted. ii) After digestion of the plasmid with the NotI, the plasmid fragment containing the UTR regions and the drug resistance gene is excised and transfected by electroporation into the parasite cells. iii) This fragment then recombines with both the 5' and 3' UTR regions of the endogenous locus, replacing the gene's ORF with the selective drug resistance gene, as represented in iv). Representations are not to scale.

3.4.8 Confirmation of the endogenously tagged cell lines by PCR

Various clones were obtained for each tagged gene. However, the results presented here refer only to the cell lines used in all experiments discussed later on in this chapter. All cell lines were routinely confirmed by both PCR and western blot (discussed in the next section) prior to performing any other experiments. These include the cell line with a TbORC1/CDC6 allele endogenously tagged with C-terminal 12myc and the other allele replaced with the puromycin resistance gene (generated by Dr Calvin Tiengwe and referred to

as TbORC1/CDC6 -/12myc) (Tiengwe *et al.*, 2012b). Therefore, primers used for its confirmation are also described.

A schematic representation of the PCR confirmation strategy is shown in Figure 3.20, A-C). Integration of the C-terminal tagging constructs, either 12myc or 6HA, was confirmed using a forward primer recognising an area upstream of the region of the gene that had been cloned into the plasmid, and a reverse primer recognising a region of the drug resistance marker gene sequence (Figure 3.20). Each forward primer is specific to each targeted gene and recognises a region absent from the construct, meaning that they can only generate a PCR product with the resistance gene primer when the construct has been integrated correctly, and will not amplify the non-tagged locus (wild type – Wt – allele) or transformed, but not integrated, plasmid. The reverse primer C40 (recognising a region of the *BSD* gene) was used in conjugation with the forward primers C102 (*TbORC1/CDC6*), C136 (*TbORC1B*), C132 (*TbORC4*), C135 (*Tb3120*), or C137 (*Tb1120*). PCR products of the expected size were generated in each case, as shown in Figure 3.20, D, panel 3 (myc PCR). For the confirmation of the C-terminal 6HA endogenous tag, in the cell lines described in section 3.4.5, the reverse primer C41 (recognising a region of the *BLE* gene) was used instead, in combination with C136 (*TbORC1B*) or C137 (*Tb1120*), and results are shown in Figure 3.20, E. For confirmation of integration of N-terminal 12myc endogenous tagging, primers were designed such that the forward primer recognises a region of the plasmid immediately upstream of the 12myc tag, and the reverse primer recognises a sequence of the tagged gene downstream of the section present in the construct (Figure 3.20, B). Again, this allows amplification only from the correctly endogenously tagged gene locus, with no amplification from the non-tagged allele or the non-integrated plasmid. For confirmation of the 12myc N-terminal tag Tb7980 cell line (named Tb7980 -/12myc), primers numbers C218 and C219 were used, and the correct size PCR product was generated, as shown in Figure 3.20, D, panel 3 (myc PCR). The successful deletion of the non-tagged allele, which was only possible for TbORC1/CDC6 and Tb7980, was confirmed using primers designed to recognise sequences upstream of the 5'UTR region cloned into the plasmid (forward primer) and within the drug resistance gene (reverse primer). For confirmation of the deletion of Tb7980, forward primer C209 was used with reverse primer C215, which recognises a sequence within the

neomycin resistance gene. The correct size PCR product generated is shown in Figure 3.20, D, panel 4 (KO PCR). Deletion of the TbORC1/CDC6 allele was confirmed with primers C213 and C214, the latter recognising a sequence within the puromycin resistance gene (Figure 3.20, D, panel 4 (KO PCR). Importantly, all cell lines were also tested for the presence of the wild type endogenous allele (Wt allele), which has not been either tagged or deleted. For this, primer pairs were designed so that only the non-modified allele (Wt allele) can be PCR-amplified. Although in all cases these primers recognise sequences also present in the endogenously tagged allele, the distance between them is always longer than 4.5 Kbp (minimum size of the constructs used for transfection). By using a PCR programme suitable only for the amplification of PCR products no longer than 2.5 Kbp, amplification of the >4.5 Kbp PCR product from the endogenously tagged allele is minimised, and only the product corresponding to the Wt allele is successfully amplified. In a cell line where one allele has been tagged and the other has been deleted, no amplification is observed using these sets of primers. For amplification of the Wt alleles the following pairs of primers were used: C101 and CTOL_12 (*TbORC1/CDC6*); C136 and C131 (*TbORC1B*); C132 and C134 (*TbORC4*); C220 and C219 (*Tb7980*); C135 and C130 (*Tb3120*); and C137 and C138 (*Tb1120*). Correct size PCR products in the different cell lines are shown in Figure 3.20, D, panel 2 (Wt allele PCR). Note that the forward primer used in the amplification of the Wt allele of *TbORC1/CDC6*, *TbORC1B*, *TbORC4*, *Tb3120* and *Tb1120* is the same one used for the confirmation of the C-terminal endogenously tagged allele, while for *Tb7980* the reverse primer is the same as the one used for the confirmation of the N-terminal endogenously tagged allele. For all PCRs, the wild type, non-tagged cell line (represented as Wt in Figure 3.20, D and E), was used both as a negative (tag integration and allele deletion PCRs) or positive control (Wt allele PCRs). Furthermore, in all cases, a PCR reaction using the primer pair C1 and C2, targeting the *TbMCM10* gene, was used as a control of the PCR reaction itself, and the same PCR conditions (PCR master mix and PCR program) were applied (Figure 3.20, D and E, panel 1 – Control PCR). All primer sequences and PCR protocols are described in materials and methods Chapter 2, section 2.2.2, Table 2-1 and Table 2-5.

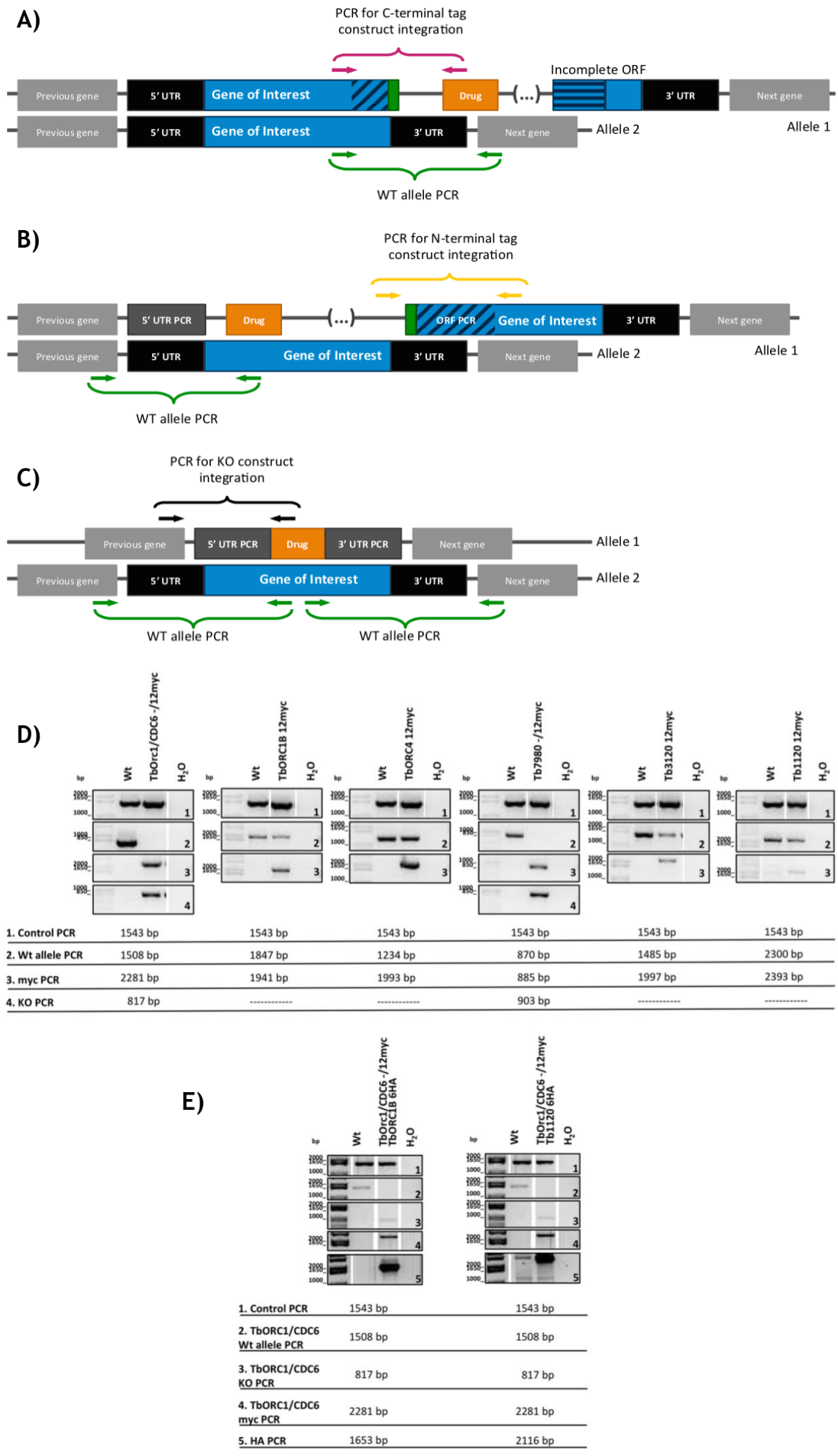


Figure 3.20. Confirmation of integration of the constructs by PCR.

Schematic representation of the resultant loci when endogenously tagged with C-terminal 12myc (or 6HA) (A), tagged with N-terminal 12myc (B), or replaced with a gene deletion construct (C). Different arrows illustrate the forward and reverse primers designed to confirm the integration of the different constructs at the target locus. Primer binding sites are also depicted for confirmation of the presence of the Wt allele (A-C). The 12myc-tagged cell lines were confirmed by PCR, which is shown in D. Panels 2-4 show the resulting products from the PCR reactions using different pairs of primers designed specifically to assess the presence of the Wt allele (panel 2), confirm integration of the C- and N-terminal endogenous tagging constructs (panel 3), and successful allele deletion (panel 4). Panel 1 shows an amplification product from *TbMCM10*, un-related to any of the modified gene loci, as a control of the PCR reaction itself. In all PCR reactions the parental, wild type, cell line was used both as a negative (myc PCR and KO PCR) or positive control (Wt allele PCR), and is here represented as Wt (far left lane). E) Confirmation by PCR of the generated 6HA-tagged TbORC1B and Tb1120 cell lines, also expressing TbORC1/CDC6 tagged with 12myc and with one *TbORC1/CDC6* allele deleted. Like in D, panel 1 shows the control PCR (primers targeting *TbMCM10*); panel 2 shows amplification of the *TbORC1/CDC6* wild type allele; panel 3 shows the PCR confirming the deletion of one *TbORC1/CDC6* allele; panel 4 shows the PCR confirming the integration of the 12myc-tag plasmid into the *TbORC1/CDC6* allele; and panel 5 shows the integration of the 6HA-tag construct into one of the alleles of *TbORC1B* (left) or *Tb1120* (right). In each case, the 1 Kb plus DNA ladder (Invitrogen™) is shown as a size reference, and the expected PCR product sizes, resultant from the different PCR reactions, are shown below each gel image. Note that in some cases the same gel image has been cropped so to display the results from the same PCR reaction, and are separated by a white space. Note that all PCR reactions, although using different primer pairs, were performed simultaneously using the same PCR program and machine.

3.4.9 Confirmation of the expression of 12myc or 6HA tagged proteins by Western Blot

In order to confirm that the tagged proteins were being expressed, around 2.5 x 10^6 cells were collected from cell cultures at approximately 1.0 x10^7 cells.ml^{-1}, lysed in protein loading buffer, and the total protein extract analysed by western blot. As detailed in materials and methods, Chapter 2, section 2.2.5.3, for detection of proteins fused with 12myc, the protein extract was separated by SDS PAGE on a 10% Bis-Tris gel, transferred onto a nitrocellulose membrane, and probed with mouse α-myc antiserum (diluted 1:7000) and detected with goat α-mouse horse radish peroxidase (HRP) conjugated antiserum (diluted 1:5000). In the case of double-tagged cell lines expressing TbORC1/CDC6^{12myc}, and TbORC1B or Tb1120 C-terminally tagged with 6HA, the membrane was stripped after myc detection and further incubated with mouse α-HA antiserum (diluted 1:10000), followed by incubation with the goat α-mouse horse radish peroxidase (HRP) conjugated antiserum (diluted 1:5000). Several clones were tested for each cell line, although only one clone of each was used for further analysis, and these are the ones shown in Figure 3.21. In all cell lines, signal reacting with the appropriate antiserum of the expected sizes for the tagged proteins was seen: TbOrc1/CDC6^{12myc}, 66.34 kDa; TbORC1B^{12myc}, 83.19 kDa; TbORC4^{12myc}, 94.99 kDa;

^{12myc}Tb7980, 66.50 kDa; Tb3120^{12myc}, 130.64 kDa; Tb1120^{12myc}, 98.4 kDa; TbORC1B^{6HA}, 72.09 kDa; and Tb1120^{6HA}, 87.30 kDa.

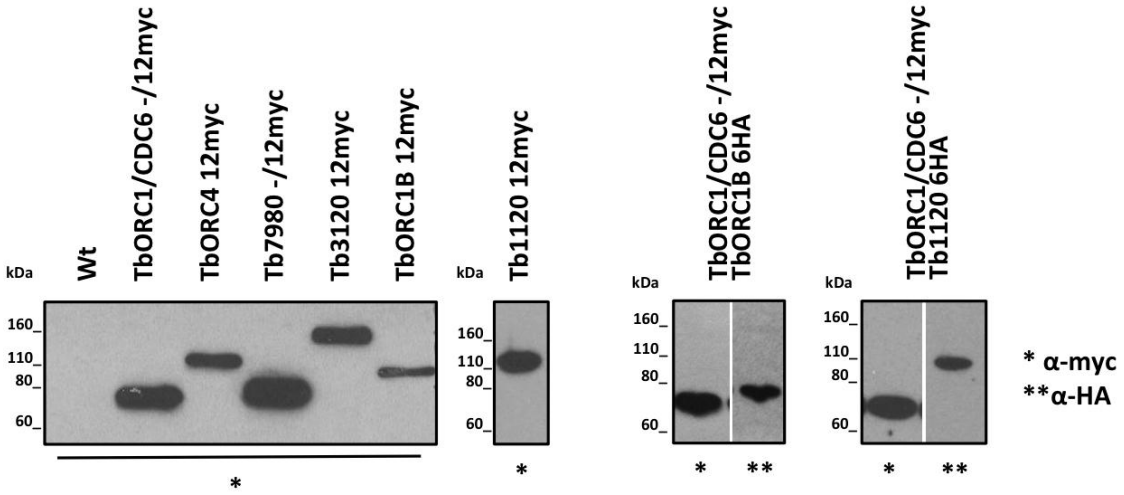


Figure 3.21. Confirmation of the expression of epitope tagged proteins by western blot. Total protein extracts from single or double-tagged cell lines were separated by SDS PAGE and analysed by western blot. The two panels on the left depict cell lines with TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 or Tb1120 tagged with 12myc, while the remaining panels show the cell lines expressing TbORC1/CDC6 tagged with 12myc and either TbORC1B or Tb1120 tagged with 6HA. (*) represents the signal obtained with the incubation of the membranes with the α-myc antiserum (α-myc); (**) represents the signal obtained with α-HA antiserum (α-HA). Note that for representation, the two right hand side panels represent two different images, one from α-myc detection and the other for α-HA exposure; these are separated by a white space. In each case, the Novex® Sharp Protein Standard was used as protein size marker.

3.5 Cellular localisation and cell cycle dynamics of TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc}, Tb1120^{12myc} and TbORC1B^{12myc} in procyclic form cells

Analysis of the subcellular localisation and dynamics throughout the cell cycle of *T. brucei* proteins by comparing them with their model eukaryotes counterparts may shed some light on their putative role in the parasite's cell cycle, as it has been shown recently for the identification of the kinetochore proteins (Akiyoshi and Gull, 2014). Following the same rationale, we aimed to investigate the subcellular localisation of TbORC1/CDC6 and interacting factors in order to better understand their biological role, infer whether they might be acting in a complex, and potentially correlate their localisation and dynamics throughout the cell cycle with already identified and described proteins from model eukaryotes.

3.5.1 Cell lines

Although TbORC1B, TbORC4, Tb7980, and Tb3120 have been shown to interact with TbORC1/CDC6 (Dang and Li, 2011; Tiengwe *et al.*, 2012b), their subcellular localisation has not been investigated. To examine this, we sought to detect these proteins using the cell lines described above, in which each protein was endogenously tagged with 12myc (section 3.4). Because our aim was to analyse factors that are hypothesised to be involved in the initiation of DNA replication, it was important to ensure that the endogenously tagged cell lines showed normal cell growth, as well as similar DNA replication levels and cell cycle profile, when compared to the unmodified, parental cell line (927 wt). As depicted in Figure 3.22, A, all cell lines showed a similar growth rate, when inoculated at a starting concentration of 1 x 10^6 cells.ml^{-1}. For the TbORC1/CDC6 -/12myc and Tb7980 -/12myc cell lines, where only the 12myc-tagged version of the protein (TbORC1/CDC6^{12myc} and ^{12myc}Tb7980, respectively) is present in the cell, unchanged growth rate (Figure 3.22, A) suggests that the tagged versions of these proteins are functional. To ask if DNA replication is altered, EdU incorporation was assessed for all cell lines, as described in detail in section 3.3.2. In all cases, approximately 20% of EdU positive cells (Figure 3.22, B) were seen after incubation for 3 hours in EdU-supplemented media at a

cell density of 1 x 10^7 cells.ml^{-1}. The same samples were also counterstained with DAPI, and cell cycle stages were assessed as described in section 3.3.2. Again, and reflecting the results obtained from both the growth curves and EdU incorporation, no cell cycle differences were evident between the 927 wt cells and any of the tagged cell lines, including TbORC1/CDC6 -/12myc and Tb7980 -/12myc (Figure 3.22, C). Because the cell lines expressing 12myc-tagged TbORC1B, TbORC4, Tb3120 or Tb1120 still have one wild type allele (as discussed in section 3.4.7), and therefore express an unmodified version of the protein, it is not possible to determine if the tagged versions of these proteins are functional. Nevertheless, if the tagged proteins impede the action of the unmodified protein or impair other proteins (for instance if they cannot interact with them), this is not reflected in any detectable changes in growth, DNA replication or cell cycle progression.

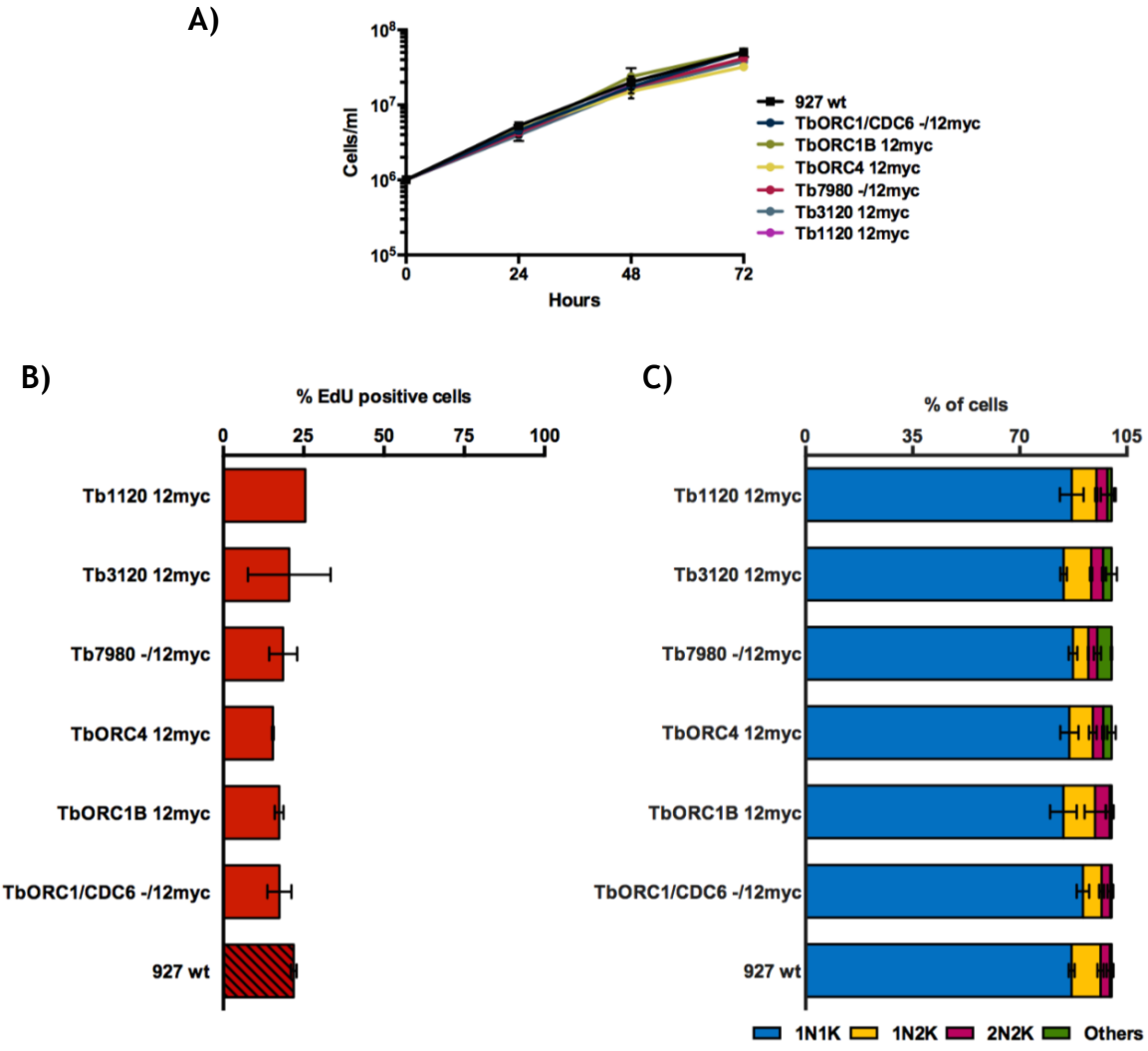


Figure 3.22. Analysis of the endogenously 12myc tagged cell lines regarding growth in cell culture, DNA replication efficiency and cell cycle progression.
Growth Curves: the different cell lines were transferred to a new culture flask with fresh SDM-79 culture media without drugs, starting at a concentration of 1 x 10^6 cells.ml^{-1}. Cell density was

assessed every 24 h, and plotted on a Log_{10} Y-axis graph, as shown in A). The individual dots on the graph represent the mean of two independent experiments (n = 2), and the error bars depict the standard error of the mean (SEM). DNA Replication: cells at a density of 1 x 10^7 cells.ml^{-1} were collected and incubated with EdU for 3 hours, and then stained for EdU detection and quantification by microscopy. A minimum of 125 cells was counted per experiment, and the percentage of EdU positive cells was calculated to from the number of cells counted per cell line. B) represents the mean % of EdU positive cells resultant from two independent experiments (n = 2), and the error bars the SEM. Cell Cycle: the cells used for EdU detection were also counterstained with DAPI, and the different cell types quantified. A minimum of 125 cells was analysed per experiment, and the percentage of cells of each cell type was calculated from the total number of cells counted per cell line. C) represents the mean of two independent experiments (n = 2) in the case of Tb7980 -/12myc, Tb3120 12myc and Tb1120 12myc, while for 927 wt, TbORC1/CDC6 -/12myc, TbORC1B 12myc and TbORC4 12myc, the mean of three independent experiments (n = 3) is shown. The error bars represent the SEM.

3.5.2 Subcellular localisation of TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc}, and Tb1120^{12myc}

The different proteins fused with 12myc were next detected by direct immunofluorescence using a commercially available anti-myc monoclonal antibody conjugated with AlexaFluor® 488 (Millipore, and used as detailed in the materials and methods, Chapter 2, section 2.5.1). This antiserum was generated from the same clone (4A6) and manufacturer (Millipore) as that used for detecting expression of the tagged proteins by western blot in section 3.4.9. As shown in Figure 3.21 (927 wt lane), this antiserum recognised specifically the myc-tagged proteins, and no cross-reactivity with any native parasite proteins was observed. Nevertheless, for each immunofluorescence assay (IFA) experiment, the parental, untagged cell line, 927 wt, was used as a negative control, also allowing the assessment of the background level of auto fluorescence or non-specific signal detected by the appropriate filter set (Alexa filter on the DeltaVision imaging system, and FITC on the Axioskop microscope). As a positive control, the cell line TbORC1/CDC6 -/12myc (Tiengwe *et al.*, 2012b) was used, as TbORC1/CDC6 has been shown to localise to the nucleus throughout the cell cycle of PCF cells by indirect immunofluorescence using a polyclonal anti-*T. cruzi* TcORC1/CDC6 antiserum (Godoy *et al.*, 2009) and it is, therefore, expected that the 12myc-tagged version of the protein (TbORC1/CDC6^{12myc}) will behave identically.

As expected, no myc signal was detected in 927 wt cells (Figure 3.23, A), and TbORC1/CDC6^{12myc} localised to the nucleus of the cell throughout the cell cycle (Figure 3.23, B), as described previously (Godoy *et al.*, 2009). These images,

acquired using a DeltaVision imaging system (and deconvolved using the ratio conservative method on the SoftWoRx software), revealed that TbORC1/CDC6^{12myc} localised to the nucleus in a punctate pattern (Figure 3.23, B), which is consistent throughout the cell cycle (i.e. indistinguishable in 1N1K, 1N2K and 2N2K cells). In addition, no TbORC1/CDC6^{12myc} signal appeared to localise to the nucleolus (Figure 3.23, B). In contrast to the observations made for TcORC1/CDC6 in *T. cruzi* promastigotes (Calderano *et al.*, 2011b), TbORC1/CDC6^{12myc} did not appear to re-localise to the nuclear periphery during the G1 to S phase transition or to remain there during S phase (Figure 3.23, B). Co-imaging of TbORC1/CDC6^{12myc} (with anti-myc antiserum) and incorporated EdU (detected with an AlexaFluor® 594-conjugate azide) (Figure 3.24) suggested that TbORC1/CDC6^{12myc} co-localised, to some extent, with newly replicated regions of the nucleus (yellow regions, myc + EdU panel). Imaging of TbORC1/CDC6^{12myc} by super resolution structure illumination microscopy (SR-SIM), using a Zeiss Elyra microscope system, allowed a better analysis of TbORC1/CDC6 localisation (Figure 3.25). These images confirmed that TbORC1/CDC6^{12myc} did not localise homogenously within the nucleus, and many discrete puncta were visible. Moreover, the number of such puncta appeared to be more abundant in S phase cells (1N1eK, and 1N2K – late S or G2 phase), although they did not appear to localise to any specific region within the nucleus (Figure 3.25). The images support exclusion of TbORC1/CDC6^{12myc} from the nucleolus, although confirmation will only be possible by counterstaining that structure with, for instance, an antibody targeting a nucleolar protein, such as TbNOG1 (Park *et al.*, 2001), or the L1C6 monoclonal antibody, which recognises an unknown antigen and has been widely used in *T. brucei* as a nucleolar maker (Durand-Dubief and Bastin, 2003; Názer and Sánchez, 2011). Discrete localisation of TbORC1/CDC6 may correlate with the genomic sites mapped previously by chromatin immunoprecipitation (ChIP) (Tiengwe *et al.*, 2012a). Co-localisation of TbORC1/CDC6^{12myc} with replicated DNA by visualising incorporated EdU by SR-SIM was attempted, but the weak signal obtained for EdU using the AlexaFluor® 594-conjugate azide was not sufficient to perform any analysis (data not shown).

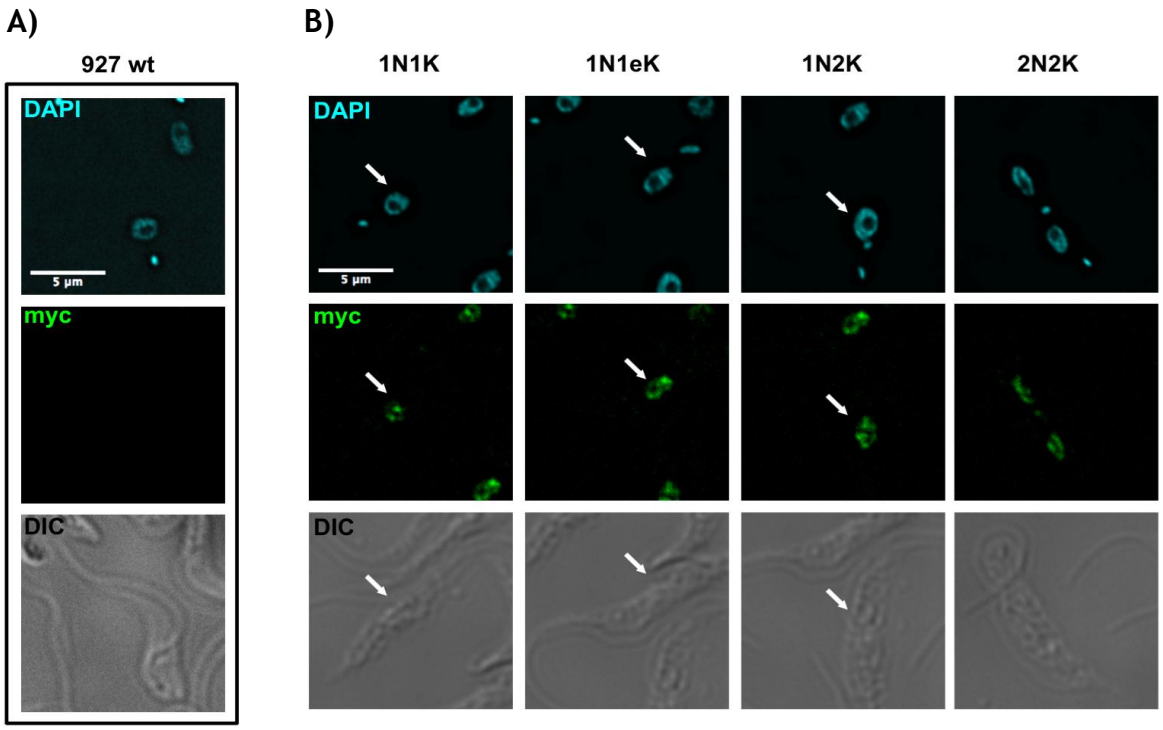


Figure 3.23. Immunofluorescence of TbORC1/CDC6^{12myc}.
A) panels inside the box show the staining of 927 wt cells with DAPI (top panel), AlexaFluor® 488-conjugated anti-myc antibody (middle panel), and the cell outline by DIC. B) panels show the staining of TbORC1/CDC6 -/12myc cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K cells (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show TbORC1/CDC6^{12myc}, recognised by the anti-myc antiserum, localising to the nucleus of the cells throughout the cell cycle, as punctate. Lower panel row shows the cells outline by DIC. Arrows direct the reader's attention to the cell in question for each cell cycle stage. Images in both A) and B) were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

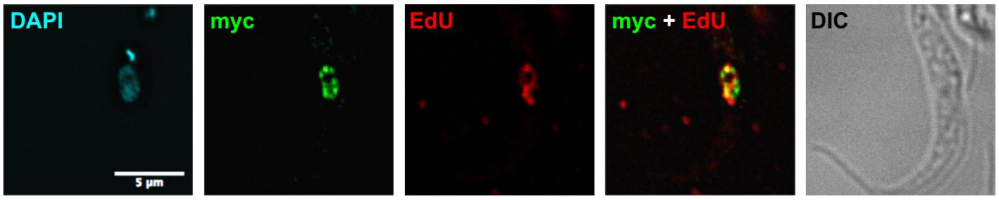


Figure 3.24. Immunofluorescence of TbORC1/CDC6^{12myc} and newly synthesised DNA.
A 1N1eK cell is shown stained with DAPI (far left panel), anti-myc antiserum (middle left panel), and AlexaFluor® 594 azide, detecting EdU (middle panel). The middle right panel depicts a merge of the myc and EdU staining, where yellow regions suggest co-localisation between TbORC1/CDC6^{12myc} and newly synthesised DNA. Cell outline is shown in the DIC image, in the far right panel. Scale bar represents 5 µm.

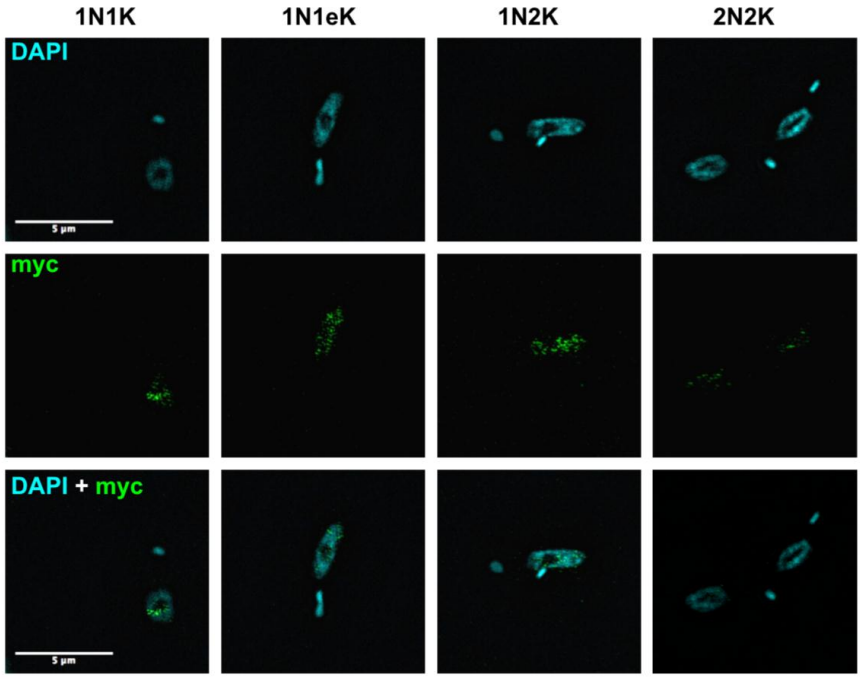


Figure 3.25. Super resolution imaging of TbORC1/CDC6^{12myc}.
Images acquired with a Zeiss Elyra super resolution microscope system. Detailed localisation of TbORC1/CDC6 within the nucleus of the cell in the different cell cycle stages. Top panel, nucleus and kinetoplast stained with DAPI; middle panel, TbORC1/CDC6^{12myc} stained with the anti-myc antiserum; bottom panel, merge of the DAPI and myc signals. Scale bar represents 5 µm.

Immunolocalisation of TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc}, and Tb1120^{12myc}, showed that these behave very similarly to TbORC1/CDC6^{12myc}: they all localised, as puncta, to the nucleus of the cell throughout the cell cycle (Figure 3.26, Figure 3.27, Figure 3.28, and Figure 3.29, respectively). The various subunits of ORC in model eukaryotes have been shown to localise to the nucleus of the cell throughout the cell cycle (Tatsumi *et al.*, 2000; Lygerou and Nurse, 1999; Pak *et al.*, 1997), which may support the idea that TbORC1/CDC6, TbORC4, Tb7980, Tb3120, and Tb1120 act together in a highly divergent ORC complex. It is perhaps worth noting that the signal for ^{12myc}Tb7980 appeared to be stronger than all the other factors (Figure 3.27; images were acquired using the same exposure time and further processed the same way using Fiji, as detailed in the materials and methods Chapter 2, section 2.5.7). This may reflect the apparently greater protein levels seen in western blot analysis (Figure 3.21), but what, if any, biological significance this might have is unclear, in particular because the role of Tb7980 in DNA replication has not been firmly validated to date. Indeed, RNAi of Tb1120 has also not been conducted, so a role in replication or even cell cycle control is speculation at this stage.

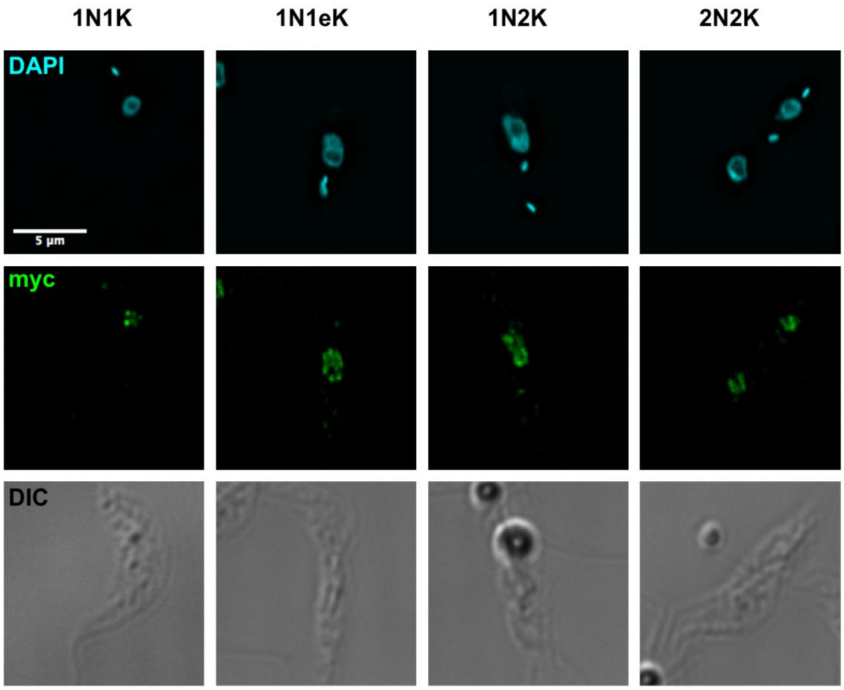


Figure 3.26. Immunofluorescence of TbORC4^{12myc}.
Panels show the staining of TbORC4^{12myc} cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K cells (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show TbORC4^{12myc}, recognised by the anti-myc antiserum, localising to the nucleus of the cells throughout the cell cycle, as punctate. Lower panel row shows the cells outline by DIC. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

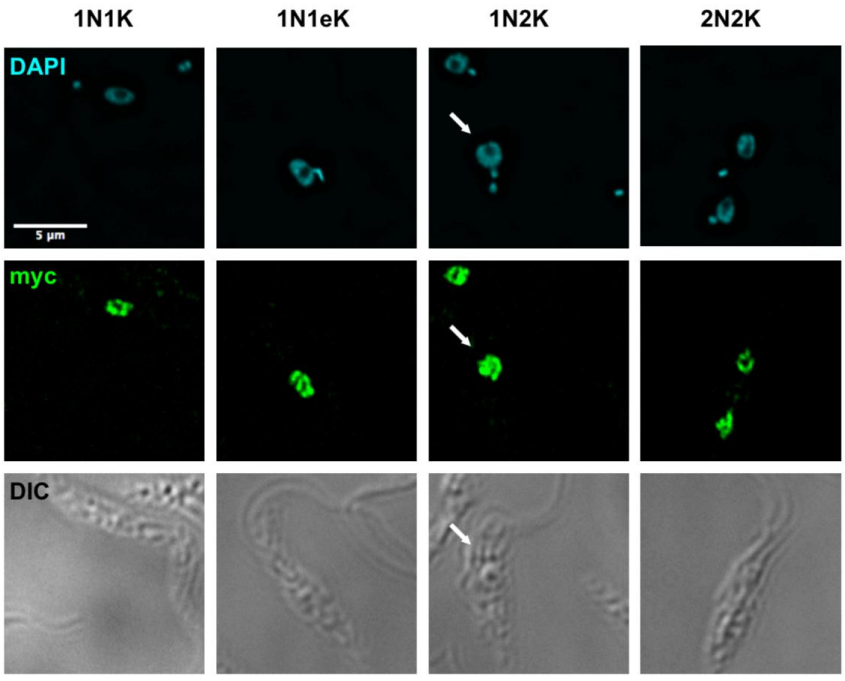


Figure 3.27. Immunofluorescence of ^{12myc}Tb7980.
Panels show the staining of ^{12myc}Tb7980 cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K cells (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show ^{12myc}Tb7980, recognised by the anti-myc antiserum, localising to the nucleus of the cells throughout the cell cycle, as punctate. Lower panel row shows the cells outline by DIC. Arrows direct the reader's attention to the cell in question for each cell cycle stage. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

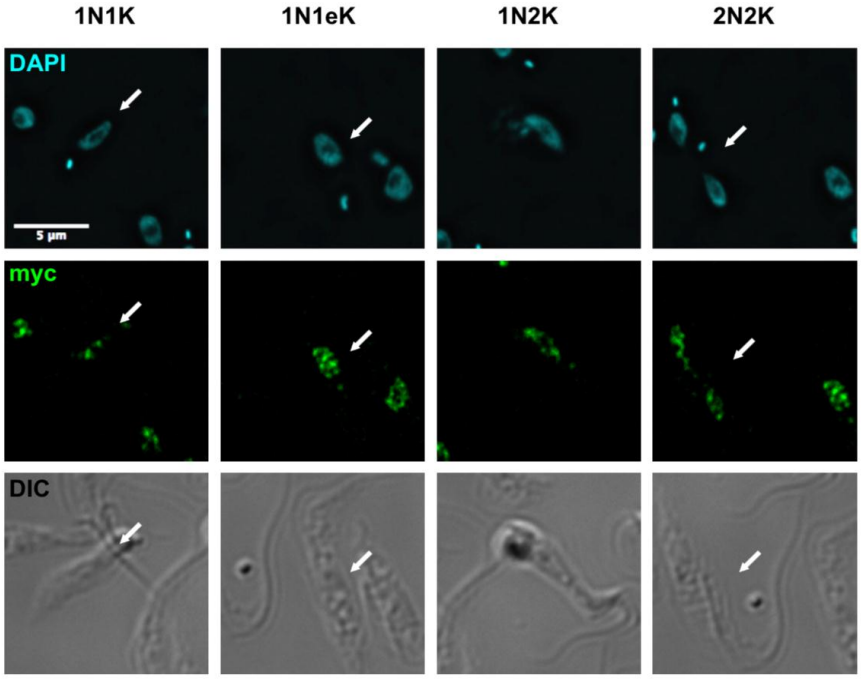


Figure 3.28. Immunofluorescence of Tb3120^{12myc}.
Panels show the staining of Tb3120^{12myc} cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K cells (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show Tb3120^{12myc}, recognised by the anti-myc antiserum, localising to the nucleus of the cells throughout the cell cycle, as punctate. Lower panel row shows the cells outline by DIC. Arrows direct the reader's attention to the cell in question for each cell cycle stage. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

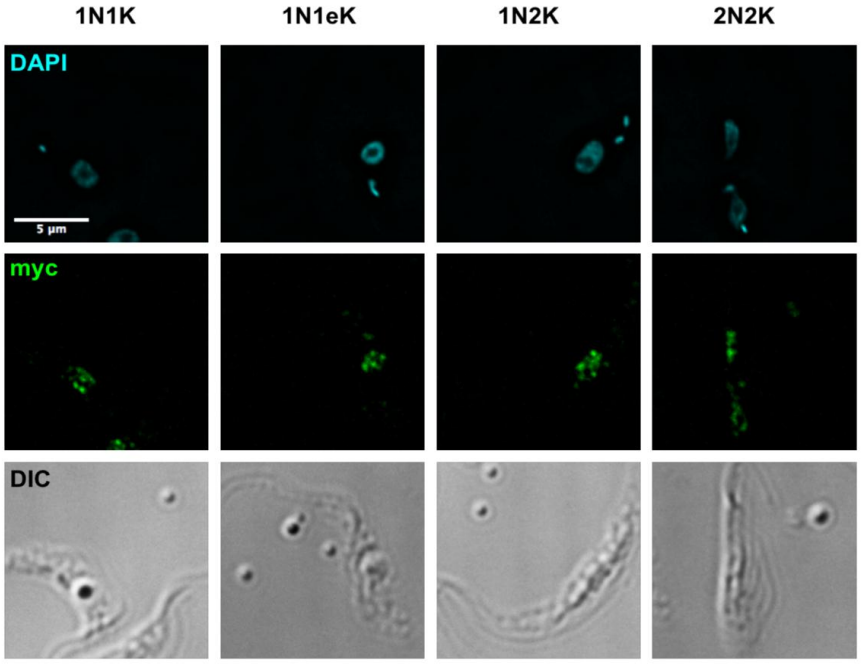


Figure 3.29. Immunofluorescence of Tb1120^{12myc}.
Panels show the staining of Tb1120^{12myc} cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K cells (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show Tb1120^{12myc}, recognised by the anti-myc antibody, localising to the nucleus of the cells throughout the cell cycle, as punctate. Lower panel row shows the cells outline by DIC. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

Quantification of the number of cells containing the individual proteins' signal in the nucleus, clearly showed that TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc}, and Tb1120^{12myc} constitutively localise to the nucleus, as in all cases ~100% of the cells showed a signal (Figure 3.30). In striking contrast, TbORC1B^{12myc} signal was detected in the nucleus of only ~33% of cells in the population (Figure 3.30). A detailed analysis of TbORC1B^{12myc} signal revealed that this factor localised to the nucleus of specific cell cycle stages, and is therefore discussed in a dedicated section (3.5.6).

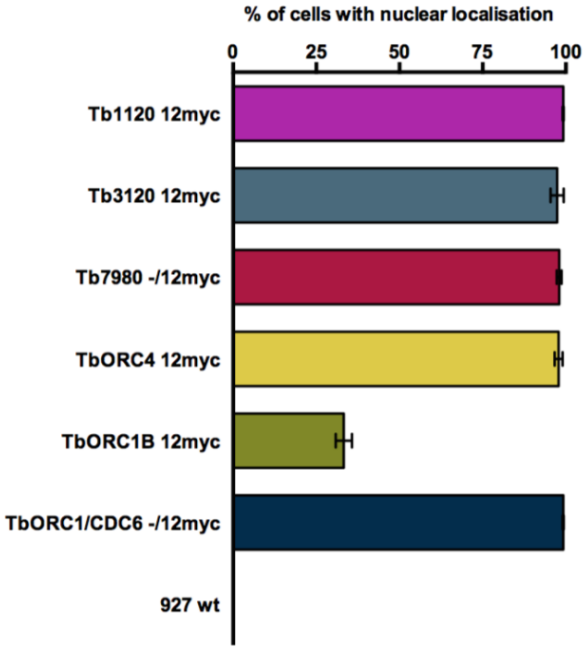


Figure 3.30. Percentage of cells containing nuclear myc signal.
Representation of the percentage of cells, per cell line, containing the respective tagged protein in the nucleus. The mean of three independent experiments (n = 3) is shown, with the error bars representing the standard error of the mean. In each experiment, at least 125 cells were counted per cell line. Percentage is calculated to the total number of cells counted per cell line in the three experiments.

3.5.3 Subcellular co-localisation of TbORC1/CDC6^{12myc} with TbORC4^{6HA}, Tb7980^{6HA}, Tb3120^{6HA}, and Tb1120^{6HA}

Although TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc}, and Tb1120^{12myc} were shown to localise to the nucleus throughout the cell cycle, in very similar punctate patterns (previous section), attempts were made to analyse co-localisation of these factors in order to test the hypothesis that they act together in a highly divergent ORC-like complex. Using the cell lines expressing both TbORC1/CDC6^{12myc} and either TbORC4, Tb7980, or Tb3120 endogenously tagged with 6HA (generated and described in Tiengwe *et al.*, 2012b), and the cell line expressing both TbORC1/CDC6^{12myc} and Tb1120 tagged with 6HA (described in section 3.4.5), co-localisation between the myc and HA signals was tested. In contrast with the 12myc tagged cells lines (as shown above), which were analysed for growth, replication and cell cycle defects, these detailed analyses were not performed for these double-tagged cell lines. Nevertheless, routine maintenance did not suggest any marked cell growth anomalies. In order to conduct co-immunofluorescence detection of both 12myc and 6HA signals, TbORC1/CDC6^{12myc} was detected as described previously with the AlexaFluor® 488 conjugated anti-myc antiserum, while the HA tagged proteins were detected by indirect immunofluorescence using a primary monoclonal anti-HA antiserum (the same used for western blot detection shown in Figure 3.21) and a secondary polyclonal anti-mouse antiserum conjugated with AlexaFluor® 594 (all detailed in the materials and methods Chapter 2, section 2.5.1). Detection of 6HA in the cells was not straightforward and failed with several batches of anti-HA antiserum, including when conjugated to the FITC fluorophore, an antibody previously described for HA detection in *T. brucei* PCF cells (Dang and Li, 2011). The level of 6HA signal that was eventually detected was lower than the one for 12myc and it was not possible to acquire high-resolution images using the DeltaVision imaging system, nor super resolution data with the Zeiss Elyra system. Nevertheless, results from wide field fluorescence imaging using the Zeiss Axioskop 2 light microscope system, allowed us to make preliminary inferences regarding co-localisation of each of the different factors relative to TbORC1/CDC6^{12myc} (Figure 3.31).

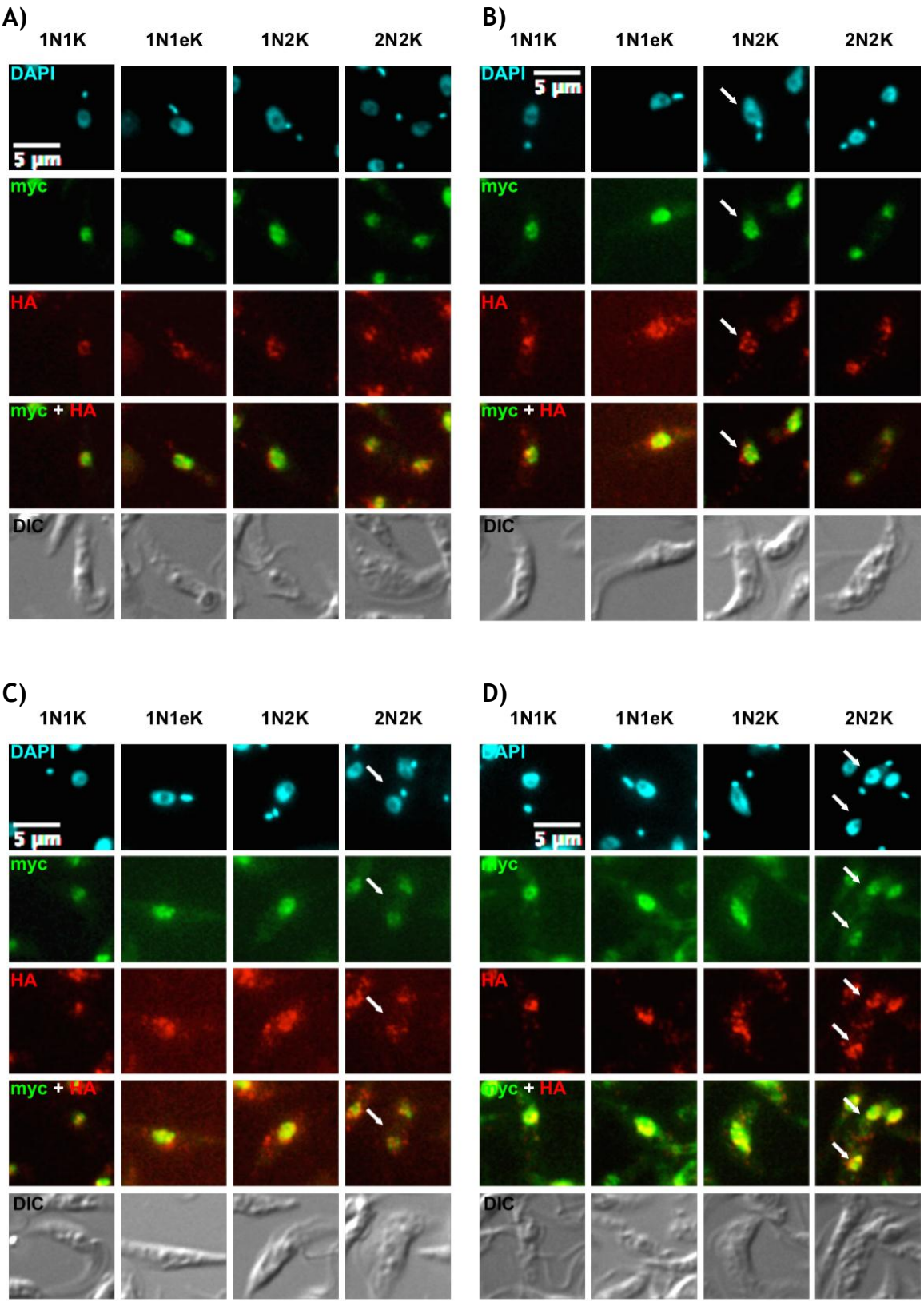


Figure 3.31. Immunofluorescence of TbORC1/CDC6^{12myc} and TbORC4^{6HA}, Tb7980^{6HA}, Tb3120^{6HA}, and Tb1120^{6HA}.

A-D) top panel shows the staining of the cells with DAPI. The second panel shows TbORC1/CDC6^{12myc}, recognised by the anti-myc antiserum. The third panel shows the signal correspondent to the proteins tagged with 6HA. The fourth panel shows a merge of the signals detected with both the anti-myc and the anti-HA antibodies. The bottom panel row shows the cells outline by DIC. A) cells expressing TbORC4^{6HA}; B) cells expressing Tb7980^{6HA}; C) cells expressing Tb3120^{6HA}; and D) cells expressing Tb1120^{6HA}. Arrows direct the reader's attention to the cell in question for each cell cycle stage. All images were acquired using an Axioskop 2 imaging system. Scale bar represents 5 µm.

In order to confirm specificity of the immunofluorescence signals for each tag, and to exclude cross-reaction between the mouse anti-myc antiserum and the mouse anti-HA primary and secondary antisera (anti-mouse), various combinations of controls were used. Because the anti-HA and anti-myc AlexaFluor® 488-conjugate antisera were both raised in mouse, the cells were first incubated with mouse anti-HA antiserum, followed by the anti-mouse antiserum conjugated with the AlexaFluor® 594 fluorophore. Next, and only after thorough washing with blocking solution, were the cells incubated with the AlexaFluor® 488 conjugated mouse anti-myc antiserum. The 927 wt cell line was used as negative control for both tags, while the TbORC1/CDC6 -/12myc cell line was used as a control for interference of the signal emitted by the anti-myc antiserum and the one emitted by the anti-HA/anti-mouse secondary antisera combination, as well as for the potential cross-reaction between the various antibodies. In addition, and to further test whether the use of this combination of antibodies would not lead to the labelling of the anti-myc antiserum by the anti-mouse AlexaFluor® 594-conjugate, or if the secondary antiserum would recognise the HA tag directly, the TbORC1/CDC6 -/12myc TbORC1B 6HA cell line was incubated with anti-myc and the anti-mouse secondary antisera alone, as an extra control. As shown in the appendices Figure 7.36, no myc or HA signal was detected in the 927 wt cell line, while only the myc signal was detected in both TbORC1/CDC6 -/12myc and TbORC1/CDC6 -/12myc TbORC1B 6HA cell lines, showing that following the designed protocol, using this combination of antibodies, lead to no cross-reactions and interference of the two signals. Nonetheless, all controls were run in parallel with the IFA of each of the double-tagged cell lines.

Similar to the observations made for their respective 12myc-tagged versions, TbORC4^{6HA}, Tb7980^{6HA}, Tb3120^{6HA}, and Tb1120^{6HA} appeared to localise to the nucleus throughout the cell cycle as puncta (Figure 3.31), although the signal was weaker than the one from the myc-tagged proteins. Nevertheless, reproduction of the localisation patterns obtained for the 12myc-tagged versions of the proteins may suggest that neither tag interferes with the proteins' functions, although it would be necessary to delete the remaining wild type allele to ultimately confirm this. The overlap of the signal obtained for TbORC1/CDC6^{12myc} and TbORC4^{6HA}, Tb7980^{6HA}, Tb3120^{6HA}, or Tb1120^{6HA} appeared

to suggest co-localisation between the proteins, although this would be more evident in images acquired with better resolution. Unfortunately, the microscope used has the inherent problem of shifting slightly when the different excitation/emission filters are changed, resulting in the images acquired with the different filters not overlapping completely (quite evident in the third panel of Figure 3.31, A). Nevertheless, these preliminary observations appear consistent with the observed interaction of TbORC1/CDC6^{12myc} with TbORC4^{6HA}, Tb7980^{6HA}, Tb3120^{6HA} (Tiengwe *et al.*, 2012b) and Tb1120^{6HA}, as shown in section 3.6.1. Again, TbORC1B^{6HA} detection and localisation in relation to TbORC1/CDC6^{12myc} will be discussed in a separate section (3.5.6.5).

It will be of value, in the future, to generate cell lines that have different combinations of these factors tagged with 12myc and 6HA, and examine whether all the factors co-localise with each other or if different interacting patterns are observed.

3.5.4 Subcellular distribution of TbORC1/CDC6^{12myc}, TbORC1B^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc}, and Tb1120^{12myc} by cell fractionation

In order to complement the information obtained by microscopy-related techniques, many studies have used cell fractionation methods to understand a protein's localisation at the subcellular level. For instance, studies from various model organisms have revealed that while the ORC subunits clearly localise to the nucleus and are mainly detected in the nuclear fraction of the cells, they can also be detected in the cytoplasmic fraction, even if such localisation is not so clearly detected by fluorescence microscopy (Tatsumi *et al.*, 2000; Lygerou and Nurse, 1999; Semple *et al.*, 2006).

In *T. brucei* PCF cells, isolation of nuclear and cytoplasmic fractions of the cell has been successfully achieved by aqueous fractionation (Zeiner *et al.*, 2003), a protocol that coupled with western blot has been used in the field for the analysis of proteins involved in various cellular processes (Trenaman *et al.*, 2013). Various characterised nuclear and cytoplasmic proteins have been used as markers for the efficiency of fractionation, including the nucleolar protein TbNOG1 (Park *et al.*, 2001) as the nuclear marker (Trenaman *et al.*, 2013; Jones

et al., 2014), and oligopeptidase B (TbOPB) (Morty *et al.*, 1999) as a cytoplasmic marker. Therefore, to assess the presence of TbORC1/CDC6^{12myc} and interacting factors in the cytoplasmic and nuclear fractions of PCF cells, sheep serum raised against *L. major* LmOPB (Munday *et al.*, 2011) (gift, Jeremy C. Mottram's Laboratory), which has been shown to recognise TbOPB (Trenaman *et al.*, 2013; Jones *et al.*, 2014), was used as the cytoplasmic marker, while the anti-TbNOG1 antibody, raised in rabbit (Park *et al.*, 2001) (gift, Marilyn Parsons' Laboratory), was used as the nuclear marker. Unfortunately, the amount of anti-TbNOG1 antibody was very limited, and only the results for Tb1120^{12myc} are shown using this marker (Figure 3.32). As an alternative, recently obtained rabbit antibody raised against *T. brucei* histone H2A (Dr Tiago D. Serafim, unpublished[j]; gift) was used (Figure 3.32). For simplicity, these antibodies are henceforth referred to as anti-OPB, anti-NOG1, and anti-H2A.

As for the microscopy assays, all the 12myc-endogenously tagged cell lines (section 3.4.9), were grown in culture at a starting concentration of 1 x 10^6 cells.ml^{-1}, and used for aqueous fractionation at a concentration of ~1 x 10^7 cells.ml^{-1}. Around 5 x 10^8 cells were collected, lysed, and fractionated as described by (Zeiner *et al.*, 2003). Immediately after extraction, both the nuclear and cytoplasmic fractions were separated by SDS-PAGE electrophoresis, alongside with a sample of the lysate used for the fractionation, as a control of the presence of the tagged protein, and were further analysed by western blot, as described in the materials and methods Chapter 2, section 2.6.3. Both the anti-OPB anti-NOG1 antisera gave very strong signals. Consequently, stripping of the nitrocellulose membrane after immunodetection was never successful, and subsequent detection of another protein was impossible (not shown). Therefore, for the aqueous fractionation experiments in Figure 3.32, each sample was equally loaded onto different SDS-PAGE gels and each probed with different antiserum (anti-myc, diluted 1:7000; anti-OPB, diluted 1:1000; anti-NOG1, diluted 1:5000; and anti-H2A, diluted 1:1000). These different antibodies were detected using goat anti-mouse (for anti-myc), donkey anti-sheep (for anti-OPB), or goat anti-rabbit (for both anti-NOG1 and anti-H2A), all IgG horseradish (HRP)-conjugated antisera, and diluted 1:5000. As shown in Figure 3.32, TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc} and Tb1120^{12myc} were

[j] The anti-H2A antibody was generated against the peptide described in Glover et al., 2012.

each detected both in the cytoplasmic (C) and nuclear (N) fractions of the cells, which, according to the signals detected with the anti-OPB and anti-H2A/NOG1 antisera, were successfully separated (i.e. no OPB signal in the nuclear fraction, and absence of NOG1 or H2A signal in the cytoplasmic fraction). These results somewhat coincide with observations made for other eukaryotes' ORC subunits, which have been detected in both nuclear and cytoplasmic fractions of the cells under study (Tatsumi *et al.*, 2000; Lygerou and Nurse, 1999; Semple *et al.*, 2006). In those studies, however, the ORC subunits were mainly detected in the nuclear fraction, with only a small proportion of the proteins being observed in the cytoplasmic fraction. This is similar to what was observed for Tb1120^{12myc} (Figure 3.32), while TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980 and Tb3120^{12myc} appeared to be equally distributed between the fractions. Why Tb1120 may differ, and whether it forms part of a putative ORC-like complex, is unclear, though it is worth noting that this analysis was only performed once with the Tb1120 12myc cell line, while the assay was repeated three times (same results, not shown) for all the others. It is nonetheless possible the less intense cytoplasmic signal merely reflects poor western blot detection.

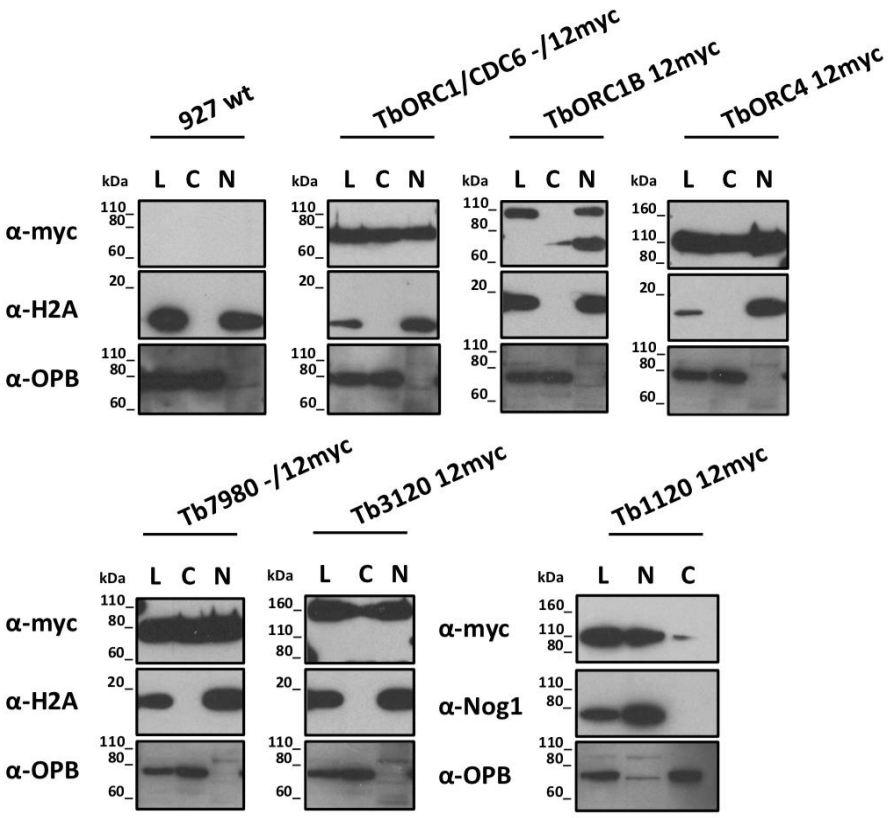


Figure 3.32. Detection of TbORC1/CDC6^{12myc}, TbORC1B^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc} and Tb1120^{12myc} in the nuclear and cytoplasmic fractions of the cell.
Untagged *T. brucei* PCF cells (927wt) or PCF cells expressing 12myc tagged variants of TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 or Tb1120 were lysed (L), and further

fractioned into cytoplasmic (C) and nuclear (N) fractions (Zeiner *et al.*, 2003). The lysate and the two fractions were then separated by SDS-PAGE and analysed by western blot. The 12myc tagged proteins were detected with the α-myc antiserum (top panel). Histone H2A (~14.2 kDa, detected with α-H2A antiserum, gift from Dr Tiago D. Serafim) or NOG1 (~74.75 kDa, detected with α-Nog1 antiserum, (Park *et al.*, 2001) were used as nuclear markers (middle panel), while oligopeptidase B (~80.5 kDa, detected with α-OPB serum, (Munday *et al.*, 2011), was used as the cytoplasmic marker (bottom panel).

Interestingly, TbORC1B^{12myc} showed a different pattern to all the other factors: full length TbORC1B^{12myc} (predicted size 83.19 kDa, but ran just below 110 kDa) was exclusively detected in the nuclear fraction (Figure 3.32, top row, third group of panels). A smaller band, of around 65 - 70 kDa was consistently detected in this analysis, and a small amount of this species was also seen in the cytoplasmic fraction, but not the lysate. The same band has also been observed in immunoprecipitation assays (shown in section 3.6.2), but it is not clear what this band might be. Since it is not seen in the other cell lines, it is most likely derived from full length TbORC1B^{12myc} and must then comprise the C-terminal end of the protein, where the 12myc tag is fused. It may simply be that TbORC1B is prone to proteolysis, which is limited in the lysate due to the fact that this sample was frozen at – 20°C immediately after lysis. Nevertheless, the different behaviour of TbORC1B, as discussed below (section 3.5.6), and perhaps reflected in the RNAi data, may mean that these observations suggest some unknown aspect of protein expression or function, though further analysis would be needed to examine this.

3.5.5 Analysis of the cell cycle dynamics of TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc}, and Tb1120^{12myc}

In previous sections it was shown that TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc} and Tb1120^{12myc} appear to localise to the nucleus of the cell throughout the cell cycle. However, it remains possible that analysis of merely counting the proportion of cells that show 12myc nuclear signal might miss dynamic expression changes of these proteins as the cell cycle progresses. In an attempt to address this, the fluorescent intensity of the nuclear 12myc signal was measured and analysed per cell cycle stage. The 12myc-tagged cell lines were used again as described in section 3.5.2, with images acquired using the Axioskop imaging system. Although the image quality and resolution is poorer when compared to the images captured using the DeltaVision microscope system, it allows the capture of images containing a larger number of cells, thus

allowing the quantification of different parameters from a large number of cells. For this analysis, cells were incubated with EdU for 3 h, and prepared for DAPI staining and anti-myc immunofluorescence, as detailed in section 3.5.2 and in the materials and methods Chapter 2, section 2.5.3. The presence of EdU, allied to N-K configuration allowed the identification of cells that were undergoing replication. Intensity of the DAPI and myc fluorescence signals was measured per individual cell by drawing a circular 21 x 21 pixel region of interest (ROI) around the cell's nucleus in Fiji (Schindelin *et al.*, 2012) (described in Chapter 2, section 2.5.7). Because light dispersion throughout an image is not always homogeneous when using the Axioskop microscope system, every analysed image was treated using the rolling ball background subtraction plugin[k], set up with a radius of 50 pixels. After background was removed, the mean pixel intensity of the DAPI and myc individual signals inside each circular ROI was measured and plotted on a graph against the corresponding cell type (1N1K, 1N1eK, 1N2K or 2N2K). A minimum of 150 cells was analysed per cell line and, in the event of a single image having more than 150 cells, a minimum of two images was still examined per cell line. Data from one experiment is shown below but, because the maximum and minimum intensities of the DAPI, myc and EdU signals can differ between experiments, an experimental repeat is shown in the appendices, section 7.6, which revealed the same dynamics profile for each of the factors.

DAPI and myc signals were first measured in the parental cell line, 927 wt (Figure 3.33), as a control. This allowed both the assessment of a normal DAPI intensity progression throughout the cell cycle and the basal level of auto fluorescence of the cell that is detected by the FITC filter set used to detect the myc signal when using the AlexaFluor® 488-conjugate antibody. Figure 3.33, A, shows representative images of normal cell types during the cell cycle, as identified by N-K ratio detected by DAPI (1N1K, 1N1eK, 1N2K and 2N2K). No detectable signal was seen in the myc images (second panel row), and EdU was detected in cells either in mid to late S phase (1N1K, 1N1eK) as well as in G2 (1N2K) phase (third panel row). DAPI fluorescence intensity (Figure 3.33, B) increased as cells progressed from G1 phase cells (1N1K; ~6 units) through S phase (1N1eK), where cells showed a wider variety of intensity values representing cells in the different stages of S phase – from early S to late S phase

[k] Available at http://fiji.sc/Rolling_Ball_Background_Subtraction

stages. The DAPI signal peaked in G2 cells (1N2K), which showed approximately double (~12 units) the signal of G1 phase cells, consistent with the DNA being completely replicated (Figure 3.33, B). After mitosis (2N2K cells), the signal of each individual nucleus of the cell decreased back to the intensity values observed for G1 cells (Figure 3.33, B), as expected upon completing nuclear division. The signal resulting from auto fluorescence of the cells in the FITC filter set showed no change during the cell cycle (Figure 3.33, C), suggesting that there is no influence of the DAPI signal (e.g. bleed through) on the signal detected with this filter setting, nor unspecific signal from incubation with the anti-myc antiserum. The average background anti-myc signal of the whole cell population, thus derived from these measurements, is represented as a dashed red line in the myc intensity plots (C) in Figure 3.34, Figure 3.35, Figure 3.36, Figure 3.37, and Figure 3.38.

In all analysed cell lines, TbORC1/CDC6 -/12myc, TbORC4 12myc, Tb7980 -/12myc, Tb3120 12myc and Tb1120 12myc, the intensity of the DAPI signal followed the dynamics described for the 927 wt parental cell line (in B of Figure 3.34, Figure 3.35, Figure 3.36, Figure 3.37, and Figure 3.38), supporting the suggestion that all these 12myc-tagged cell lines have no clear cell cycle defects (section 3.5.1). Furthermore, each of the tagged proteins was localised to the nucleus of the cells throughout the cell cycle (in A of Figure 3.34, Figure 3.35, Figure 3.36, Figure 3.37, and Figure 3.38), as expected from earlier analysis (section 3.5.1). Measurement of the intensity of the anti-myc signal revealed that virtually all cells, from each cell line, presented a higher than the measured background signal (in C of Figure 3.34, Figure 3.35, Figure 3.36, Figure 3.37, and Figure 3.38). Moreover, all the cell lines showed a similar intensity profile, which followed the dynamics of the DAPI signal: the signal increased gradually from G1 phase cells (1N1K) through S phase cells (1N1eK), peaking in G2 cells (1N2K), and reducing back to the intensity levels found in G1 phase cells in post-mitotic (2N2K) cells (in C of Figure 3.34, Figure 3.35, Figure 3.36, Figure 3.37, and Figure 3.38). Indeed, in all cases, the average signal level detected in the G2 phase cells was approximately double that seen in G1 phase and post-mitotic cells, a significant difference (p-value <0.05) also seen in the DAPI signal intensity profiles (sections B and C of Figure 3.34, Figure 3.35, Figure 3.36, Figure 3.37, and Figure 3.38). The similarity in dynamics of the DAPI and myc

intensity profiles suggests that the amount of these factors in the nucleus increased as the amount of DNA in the nucleus of the cell increased, peaking in 1N2K cells, when the genomic DNA is almost or completely replicated, rather than remaining constant throughout the cell cycle. As TbORC1/CDC6 has been shown to bind to chromatin throughout the cell cycle (Godoy *et al.*, 2009), it is possible that these data indicate that all these putative Orc-like factors might bind to newly synthesised origins of replication as S phase progresses, as has been observed in other eukaryotes (Lygerou and Nurse, 1999).

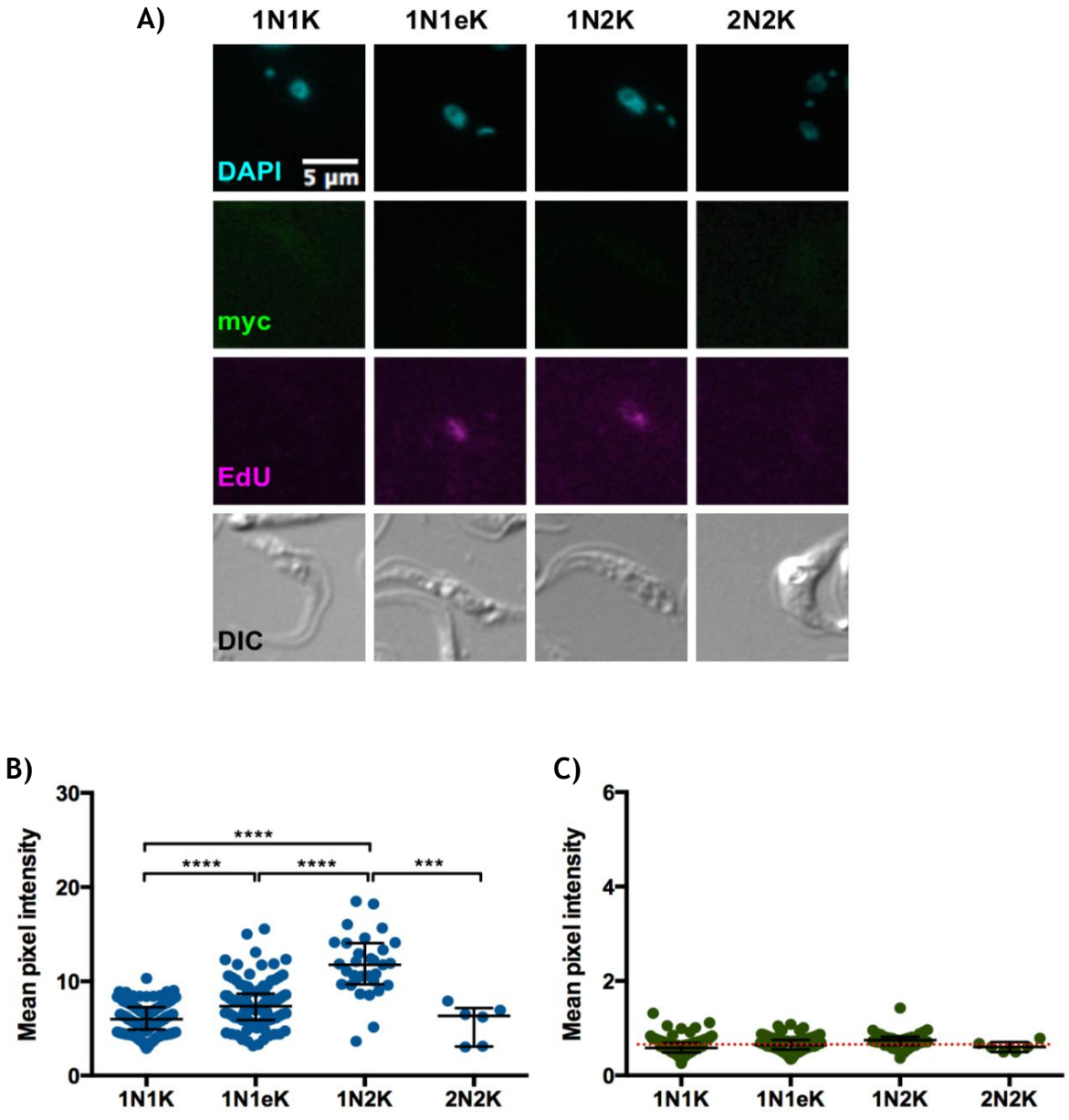


Figure 3.33. Analysis of the DAPI signal and background signal of myc detection in the parental cell line, 927 wt.
A) staining of 927 wt cells with DAPI (top panel), AlexaFluor® 488-conjugated anti-myc antibody (second panel), AlexaFluor® 555-conjugated azide (for EdU detection; third panel), and the cell outline by DIC (bottom panel). All images were acquired using an Axioskop 2 imaging system. Scale bar represents 5 µm. Intensity of the DAPI (B) and myc (C) signals is represented (dots) as the mean of pixel intensity within the circular region of interest (ROI, of 21 x 21 pixels), drawn

around each individual cell nucleus. In C), the red dotted line represents the average background signal measured in the 927 wt cell line. A total of 275 cells were analysed (n = 275). In both B) and C) the median of the values is represented, with the error bars depicting the interquartile range. Statistical significance between the different cell cycle stages was assessed through analysis using the Kruskal-Wallis non-parametric test. (***) p-value < 0.001; (****) p-value <0.0001. Intensity of the EdU signal is shown in the appendices, Figure 7.40.

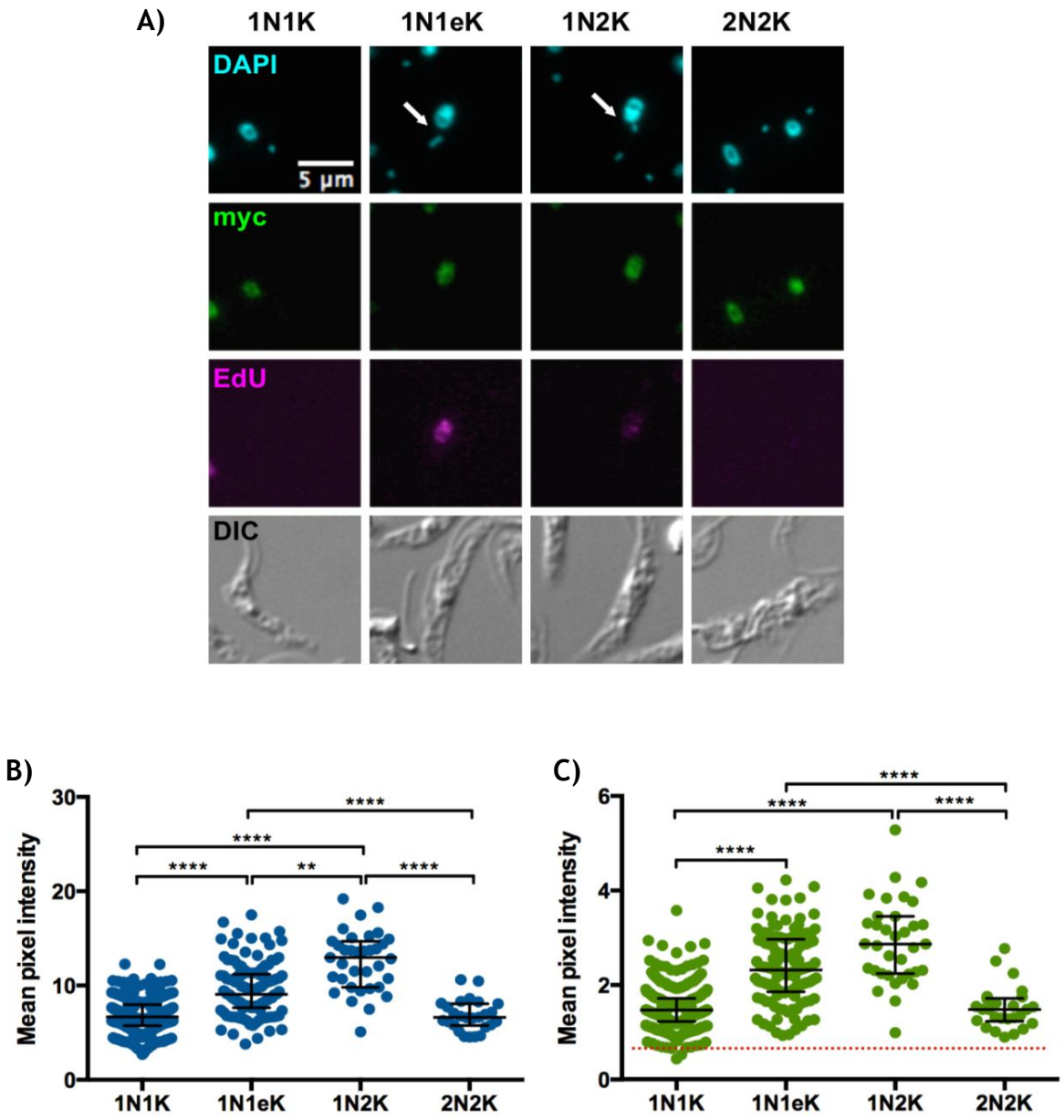


Figure 3.34. TbORC1/CDC6^{12myc} subcellular localisation and dynamics throughout the cell cycle.
A) staining of TbORC1/CDC6 -/12myc cells with DAPI (top panel), AlexaFluor® 488-conjugated anti-myc antibody (second panel), AlexaFluor® 555-conjugated azide (for EdU detection; third panel), and the cell outline by DIC (bottom panel). Arrows denote the cell in the specific cell cycle stage of interest, where other cells are present. All images were acquired using an Axioskop 2 imaging system. Scale bar represents 5 µm. Intensity of the DAPI (B) and myc (C) signals is represented (dots) as the mean of pixel intensity within the circular region of interest (ROI, of 21 x 21 pixels), drawn around each individual cell nucleus. In C), the red dotted line represents the average background signal measured in the 927 wt cell line. A total of 591 cells were analysed (n = 591). In both B) and C) the median of the values is represented, with the error bars depicting the interquartile range. Statistic significance between the different cell cycle stages was assessed through analysis using the Kruskal-Wallis non-parametric test. (**) p-value < 0.01; (***) p-value < 0.001; (****) p-value <0.0001. Intensity of the EdU signal is shown in the appendices, Figure 7.40.

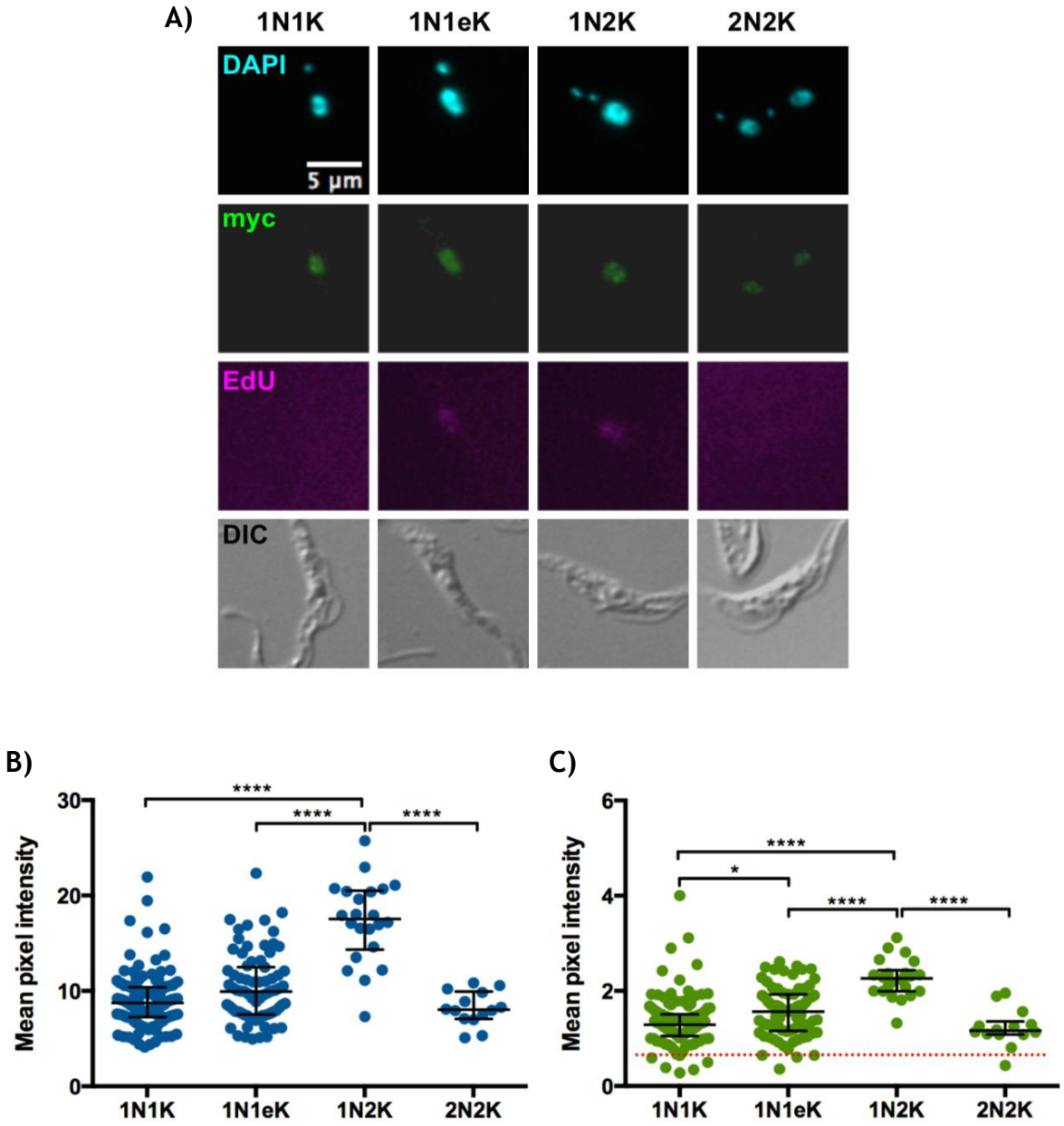


Figure 3.35. TbORC4^{12myc} subcellular localisation and dynamics throughout the cell cycle. A) staining of TbORC4 12myc cells with DAPI (top panel), AlexaFluor® 488-conjugated anti-myc antibody (second panel), AlexaFluor® 555-conjugated azide (for EdU detection; third panel), and the cell outline by DIC (bottom panel). All images were acquired using an Axioskop 2 imaging system. Scale bar represents 5 µm. Intensity of the DAPI (B) and myc (C) signals is represented (dots) as the mean of pixel intensity within the circular region of interest (ROI, of 21 x 21 pixels), drawn around each individual cell nucleus. In C), the red dotted line represents the average background signal measured in the 927 wt cell line. A total of 242 cells were analysed (n = 242). In both B) and C) the median of the values is represented, with the error bars depicting the interquartile range. Statistic significance between the different cell cycle stages was assessed through analysis using the Kruskal-Wallis non-parametric test. (*) p-value < 0.05; (****) p-value <0.0001. A), B) and C) show results obtained from the same experiment. Intensity of the EdU signal is shown in the appendices, Figure 7.40.

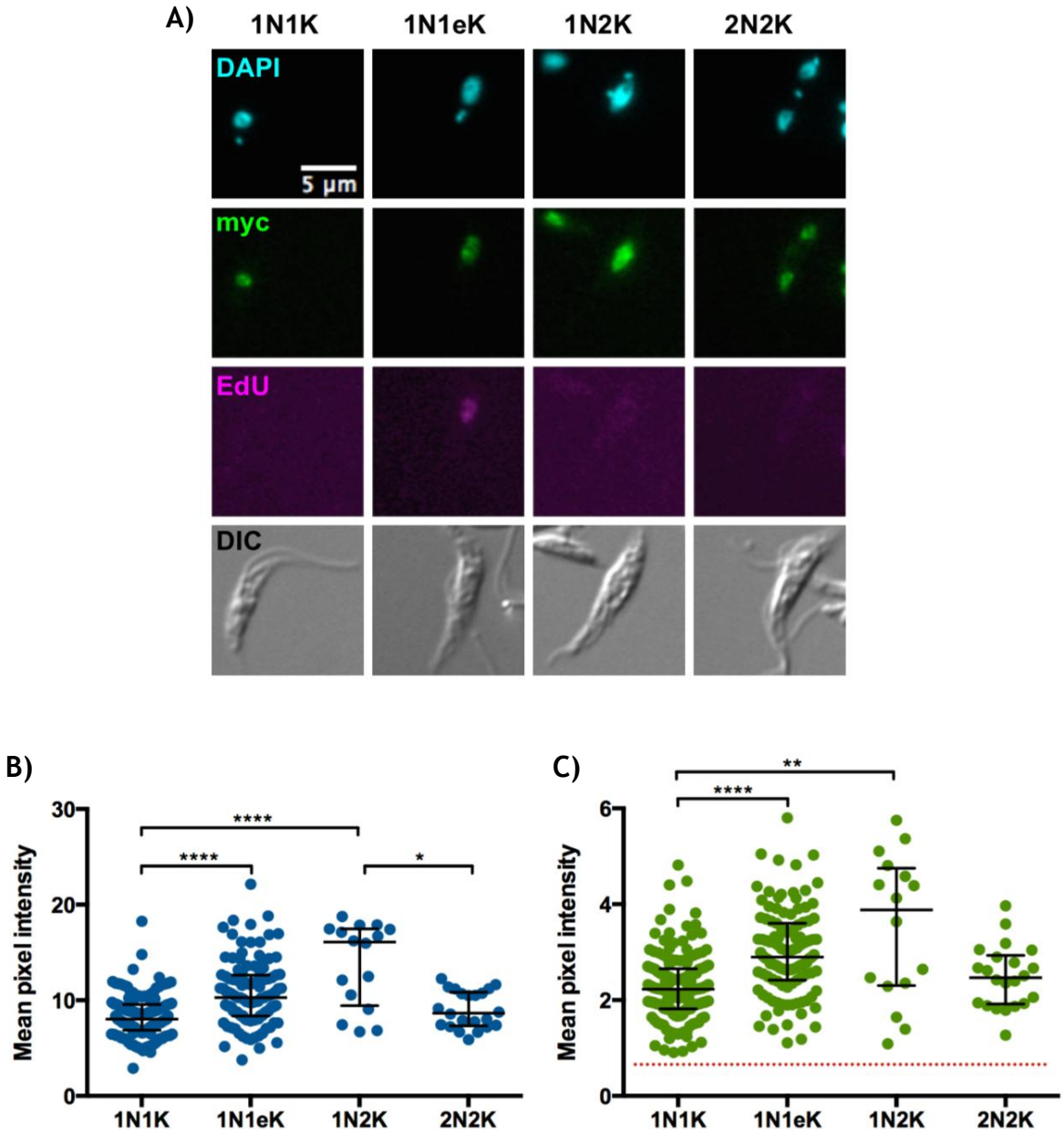


Figure 3.36. ^{12myc}Tb7980 subcellular localisation and dynamics throughout the cell cycle. A) staining of Tb7980 -/12myc cells with DAPI (top panel), AlexaFluor® 488-conjugated anti-myc antibody (second panel), AlexaFluor® 555-conjugated azide (for EdU detection; third panel), and the cell outline by DIC (bottom panel). All images were acquired using an Axioskop 2 imaging system. Scale bar represents 5 µm. Intensity of the DAPI (B) and myc (C) signals is represented (dots) as the mean of pixel intensity within the circular region of interest (ROI, of 21 x 21 pixels), drawn around each individual cell nucleus. In C), the red dotted line represents the average background signal measured in the 927 wt cell line. A total of 315 cells were analysed (n = 315). In both B) and C) the median of the values is represented, with the error bars depicting the interquartile range. Statistic significance between the different cell cycle stages was assessed through analysis using the Kruskal-Wallis non-parametric test. (*) p-value < 0.05; (**) p-value < 0.01; (****) p-value <0.0001. A), B) and C) show results obtained from the same experiment. Intensity of the EdU signal is shown in the appendices, Figure 7.40.

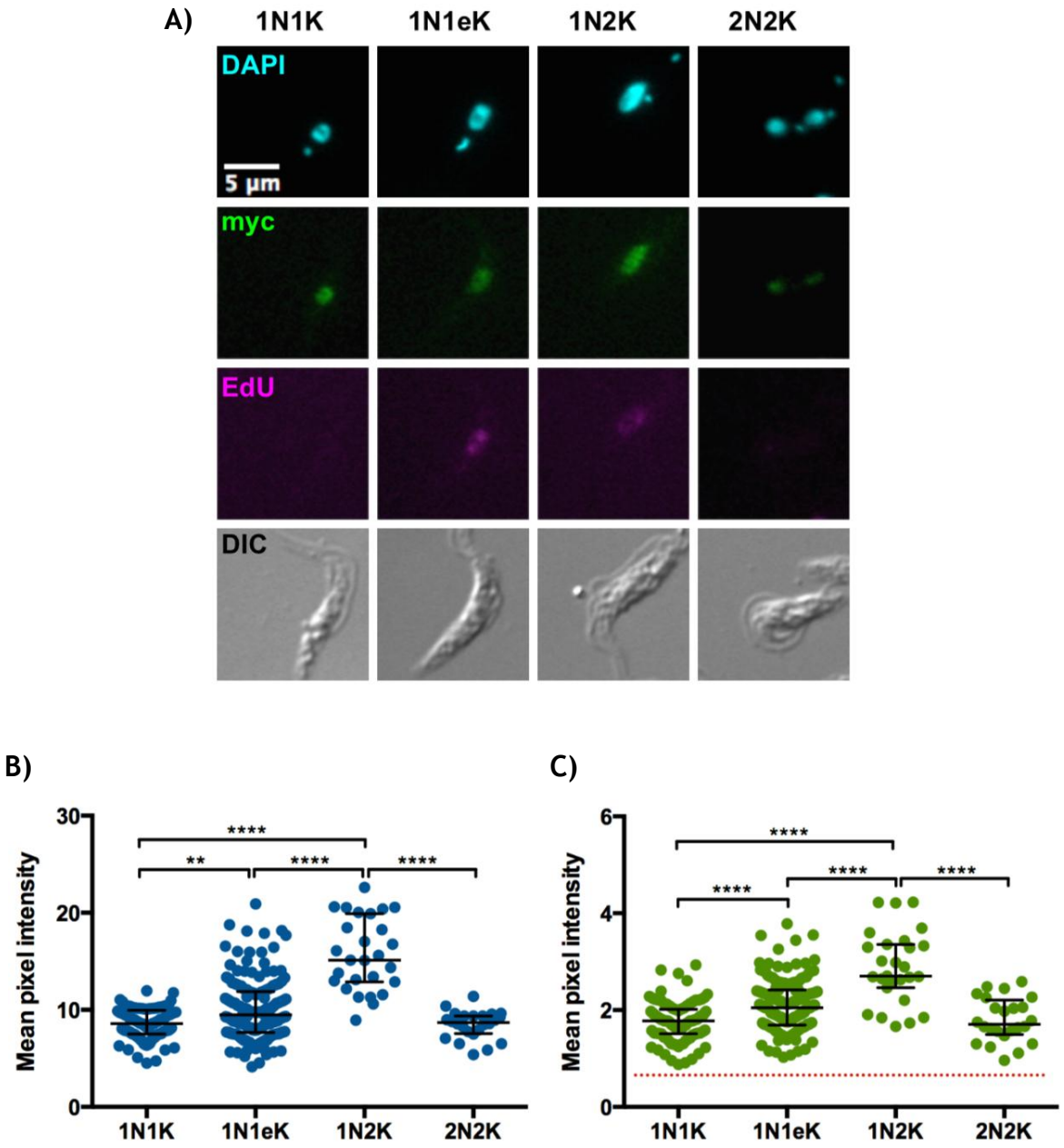


Figure 3.37. Tb3120^{12myc} subcellular localisation and dynamics throughout the cell cycle. A) staining of Tb3120 12myc cells with DAPI (top panel), AlexaFluor® 488-conjugated anti-myc antibody (second panel), AlexaFluor® 555-conjugated azide (for EdU detection; third panel), and the cell outline by DIC (bottom panel). All images were acquired using an Axioskop 2 imaging system. Scale bar represents 5 µm. Intensity of the DAPI (B) and myc (C) signals is represented (dots) as the mean of pixel intensity within the circular region of interest (ROI, of 21 x 21 pixels), drawn around each individual cell nucleus. In C), the red dotted line represents the average background signal measured in the 927 wt cell line. A total of 288 cells were analysed (n = 288). In both B) and C) the median of the values is represented, with the error bars depicting the interquartile range. Statistic significance between the different cell cycle stages was assessed through analysis using the Kruskal-Wallis non-parametric test. (**) p-value < 0.01; (****) p-value <0.0001. A), B) and C) show results obtained from the same experiment. Intensity of the EdU signal is shown in the appendices, Figure 7.40.

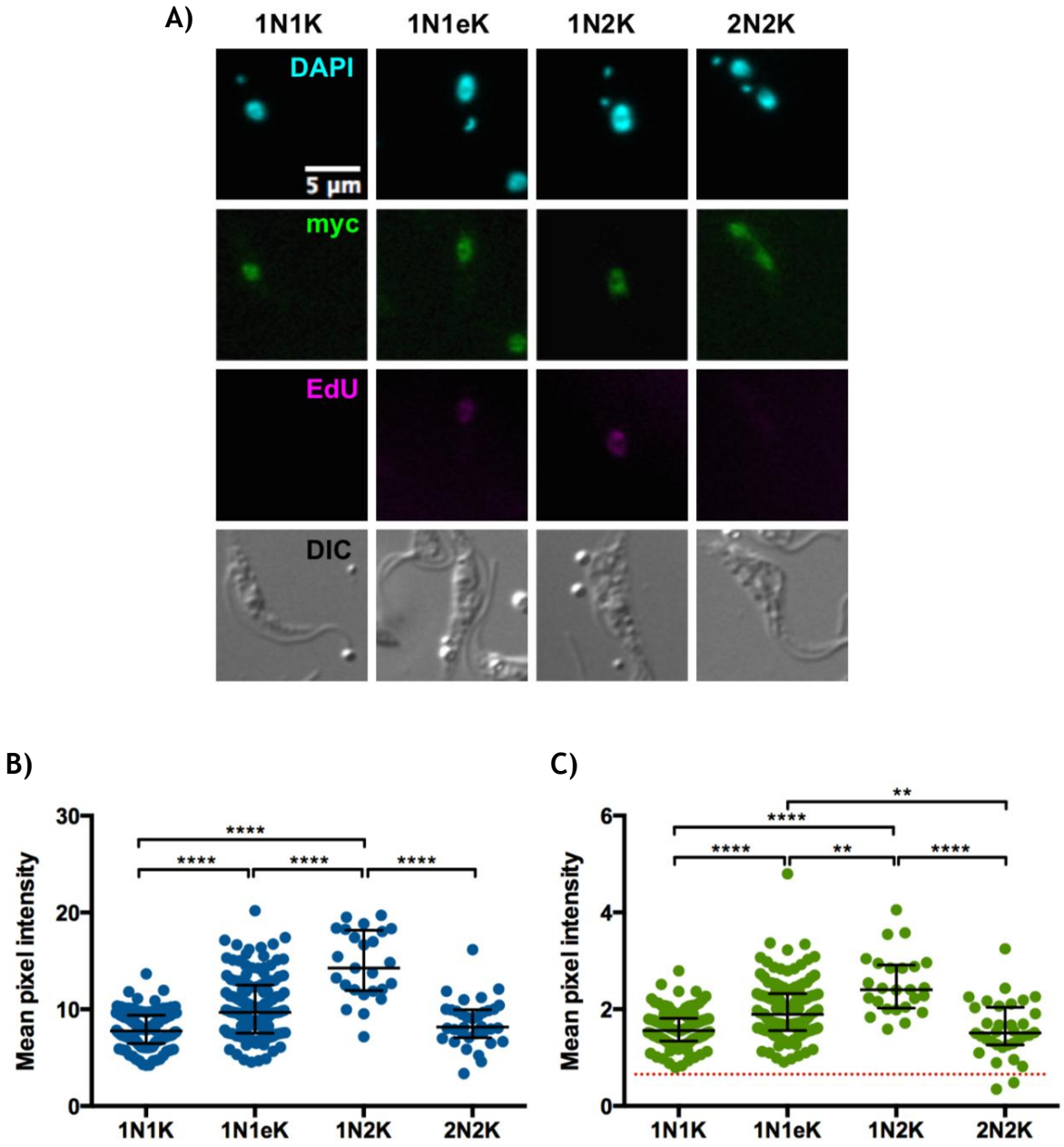


Figure 3.38. Tb1120^{12myc} subcellular localisation and dynamics throughout the cell cycle. A) staining of Tb1120 12myc cells with DAPI (top panel), AlexaFluor® 488-conjugated anti-myc antibody (second panel), AlexaFluor® 555-conjugated azide (for EdU detection; third panel), and the cell outline by DIC (bottom panel). All images were acquired using an Axioskop 2 imaging system. Scale bar represents 5 µm. Intensity of the DAPI (B) and myc (C) signals is represented (dots) as the mean of pixel intensity within the circular region of interest (ROI, of 21 x 21 pixels), drawn around each individual cell nucleus. In C), the red dotted line represents the average background signal measured in the 927 wt cell line. A total of 296 cells were analysed (n = 296). In both B) and C) the median of the values is represented, with the error bars depicting the interquartile range. Statistic significance between the different cell cycle stages was assessed through analysis using the Kruskal-Wallis non-parametric test. (**) p-value < 0.01; (****) p-value <0.0001. A), B) and C) show results obtained from the same experiment. Intensity of the EdU signal is shown in the appendices, Figure 7.40.

As in the previous sections, the results obtained for TbORC1B are shown and discussed in a dedicated section (3.5.6).

3.5.6 The exceptional case of TbORC1B: cell cycle-dependent, dynamic subcellular localisation

In contrast to TbORC1/CDC6, TbORC4, Tb7980, Tb3120 and Tb1120, which appeared to localise to the nucleus and to show a similar dynamics throughout the cell cycle, possibly suggesting that these factors may be interacting in a complex, TbORC1B seemed to behave quite differently in every performed assay. In this section, all results obtained from the assays discussed in the previous sections are shown for TbORC1B alone.

3.5.6.1 Subcellular localisation of TbORC1B^{12myc}

Subcellular localisation of TbORC1B^{12myc} was assessed using anti-myc antiserum as described in section 3.5.2. Like TbORC1/CDC6^{12myc} and all its interacting partners described above, TbORC1B^{12myc} localised exclusively to the nucleus in an apparent punctate pattern, and appeared to be absent from the nucleolus (Figure 3.39). Super resolution images (Figure 3.40) suggest that TbORC1B^{12myc} puncta resemble those observed for TbORC1/CDC6^{12myc} (Figure 3.25), possibly in approximate number and size. The striking difference between TbORC1B^{12myc} and all the other factors, however, is that it was not detected in every cell. Quantification revealed that TbORC1B^{12myc} nuclear signal was only present in approximately 33% of the cellular population (Figure 3.30, green bar; Figure 3.41, B). Categorisation of the N-K ratios of the cells showed that TbORC1B^{12myc} signal was not evenly distributed amongst the detectable cell cycle phases. In only 10% of 1N1K cells was TbORC1B^{12myc} signal detected, and no 2N2K cells displayed any signal (Figure 3.41, B). In contrast, ~70% of 1N1eK cells had detectable signal, as well as ~50% of 1N2K cells (Figure 3.41, B). Taken together, these data indicate that TbORC1B^{12myc} was only detected in the nucleus of cells that appeared to be at the onset of S phase (1N1K and 1N1eK), in S phase (1N1eK), and either in late S or G2 phases (1N2K) (Figure 3.39). It is important to note that the categorisation of 1N1eK cells is not a straightforward process: such cells are clear when the kinetoplast is evidently elongated (and perhaps approaching segregation), and it is possible that some cells categorised as 1N1K may actually be already undergoing kinetoplast replication, and thus be in late G1 or early S phases. In principle, it may have been considered that such cells would have been detected by EdU incorporation in the kDNA, but in all assays

performed, incorporation of EdU into the kDNA was never observed. In addition, nuclear EdU signal was barely detected in these cells, most likely because the amount of incorporated EdU at such early stages of nuclear DNA replication is below detection. As an alternative, staining of the newly synthesised flagellum has been shown to allow the differentiation between 1N1K cells that are in G1 phase (have one flagellum) and 1N1K cells that are in S phase (two flagella) (Godoy *et al.*, 2009), and could potentially have been of value here. As already mentioned, TbORC1B^{12myc} was detected in the nucleus of approximately half of the 1N2K cells analysed (Figure 3.41, C). Whether these cells are in late S phase, or whether TbORC1B^{12myc} is retained in the nucleus through at least part of G2 phase is difficult to evaluate. However, detailed examination suggested that TbORC1B^{12myc} signal was never detected in 1N2K cells that had an elongated nucleus and in which the two kinetoplasts were localised very distant from each other (an example is shown in Figure 3.39), suggesting that the protein becomes undetectable either late in G2 phase or when nuclear mitosis begins. It could be interesting to label the cells with a mitosis marker, such as the antibody against di-methylated histone H3 lysine 76 (anti-H3K76me2) (Janzen *et al.*, 2006) or the commercially available anti-KMX-1 (an antibody against ß-tubulin) that has been used to stain the mitotic spindle (Hu *et al.*, 2014), in order to examine these observations further.

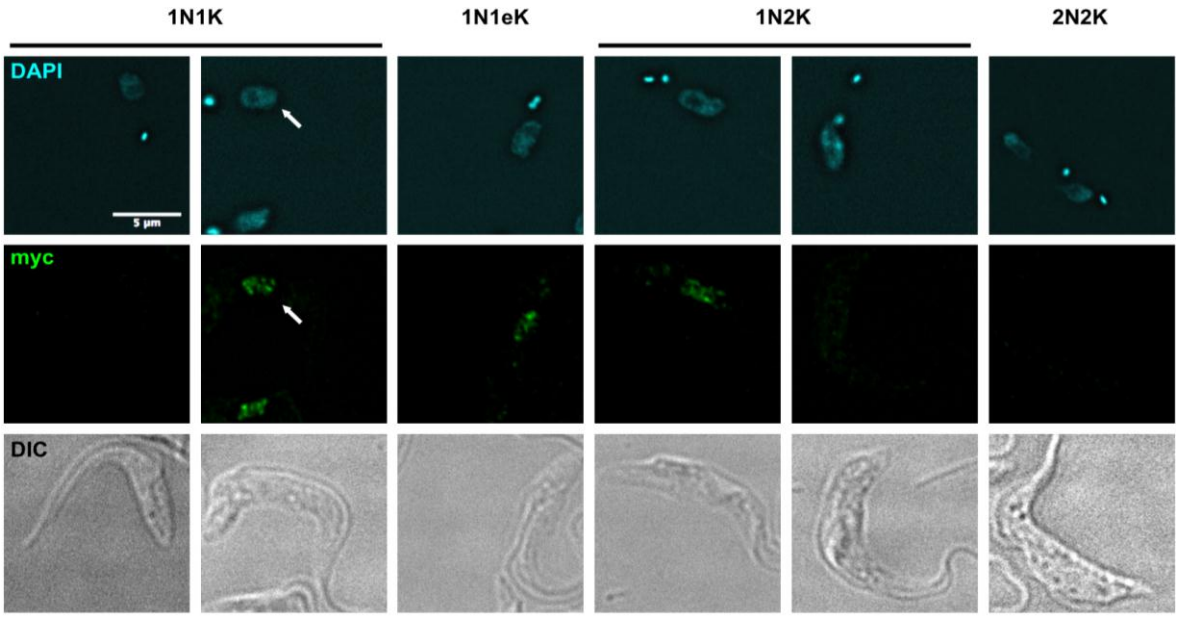


Figure 3.39. Immunofluorescence of TbORC1B^{12myc}.
Panels show the staining of TbORC1B 12myc cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K cells (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show TbORC1B^{12myc}, recognised by the AlexaFluor® 488-conjugated anti-myc antiserum. Lower panel row shows the cells outline by DIC. Arrows direct

the reader's attention to the cell of interest. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

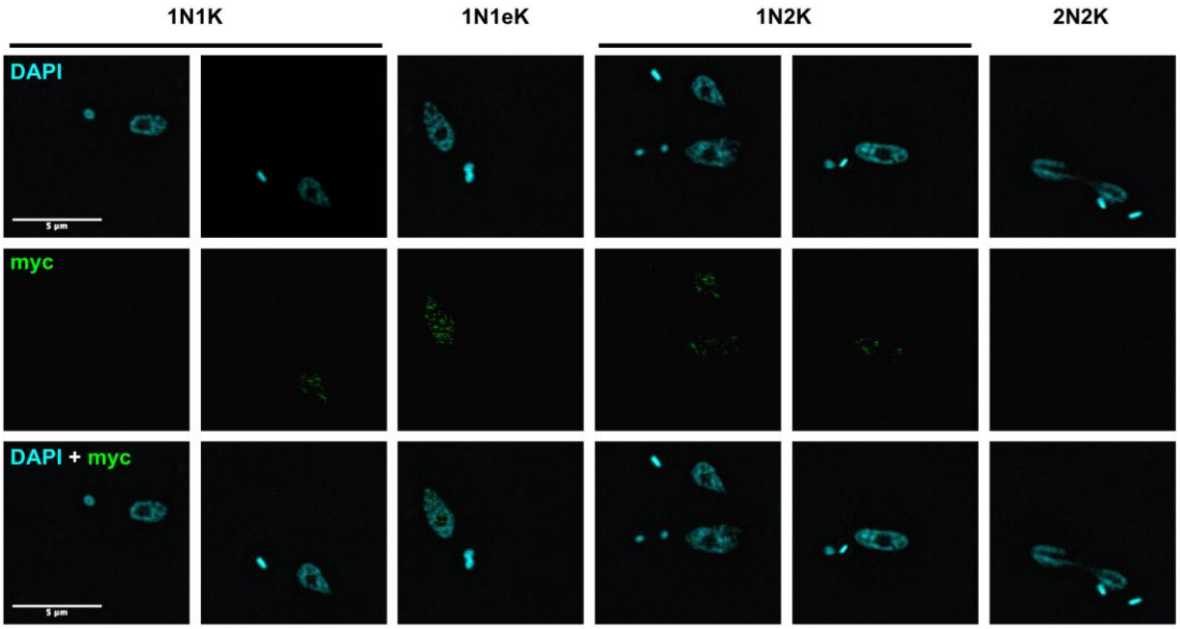


Figure 3.40. Super resolution imaging of TbORC1B^{12myc}.
Top panel, nucleus and kinetoplast stained with DAPI; middle panel, TbORC1B^{12myc} stained with the anti-myc antiserum; bottom panel, merge of the DAPI and myc signals. Scale bar represents 5 µm. Images acquired with a Zeiss Elyra super resolution microscope system.

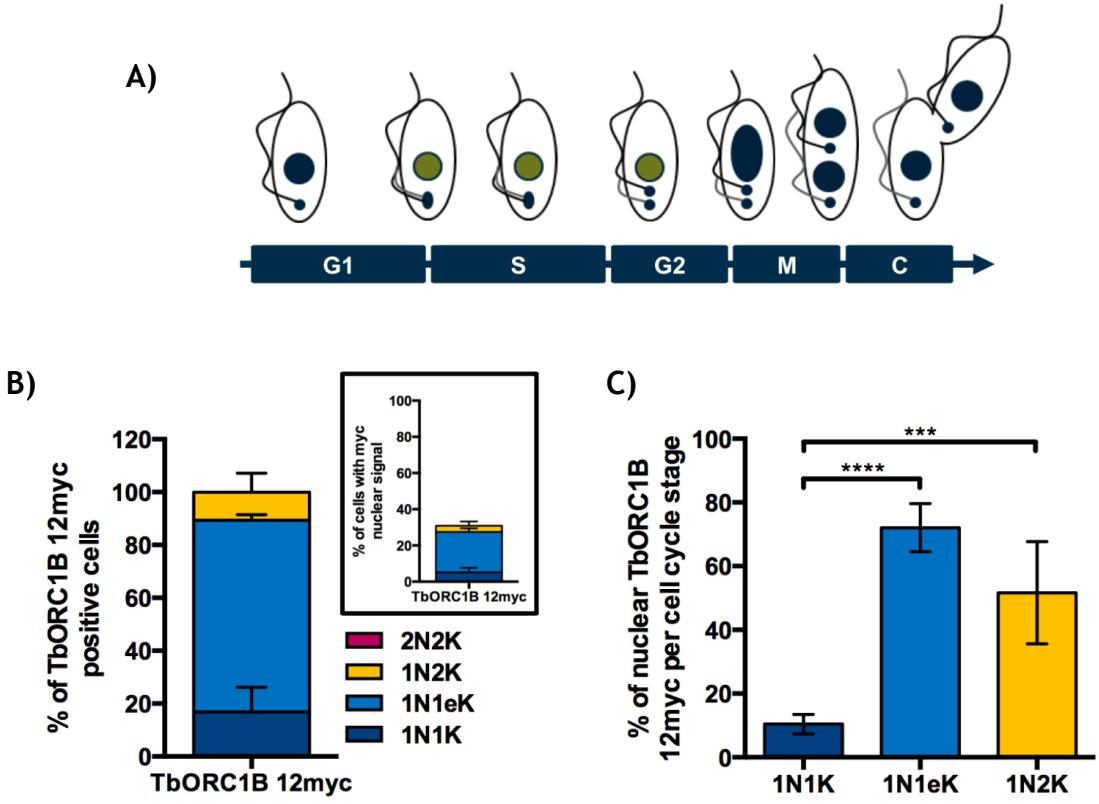


Figure 3.41. Quantification of cells containing nuclear TbORC1B^{12myc} signal.
A) schematic representation of the cells based on the ratio between nucleus (N) and kinetoplast (K) per cell: 1N1K, 1N1eK, 1N2K, and 2N2K. Bar below the cells schematics represents the different cell cycle phases to which each cell type (above) corresponds. Cells that contain TbORC1B signal in the nucleus have the nucleus coloured in green. The diagram was generated based on

information and schematics from (Field and Carrington, 2009; McKean, 2003). B) Quantification of the proportion of cells containing TbORC1B^{12myc} signal. Approximately 33% of the cells (inset box) showed TbORC1B^{12myc} signal in the nucleus, and the percentage of this total is shown by cell cycle type as determined by N-K ratio (dark blue, 1N1K cells; light blue, 1N1eK cells; yellow, 1N2K cells, red, 2N2K). Results were obtained from two independent experiments; the mean is represented (n = 2), and the error bars depict the standard deviation. C) Percentage of individual cell cycle stage cells that display TbORC1B^{12myc} signal (2N2K cells are not represented, as none showed TbORC1B^{12myc} signal in any of the experiments). The mean of four independent experiments is represented; error bars correspond to the standard deviation. Statistical significance between the different groups was assessed through analysis using the one-way ANOVA parametric test. (***) p-value < 0.001; (****) p-value <0.0001.

3.5.6.2 Cell cycle dynamics of TbORC1B^{12myc}

Overall, the above results indicate that TbORC1B localises to the nucleus of the cells at the onset and throughout S phase, but is no longer present in the nucleus of cells in late G2 or that have started mitosis (diagram in Figure 3.41, A). Results from aqueous fractionation suggested that TbORC1B^{12myc} localises exclusively to the nucleus of the cell (Figure 3.32), excluding regulation of its function by re-localisation to the cytoplasm outside S phase, as has been described for the Orc1 subunits of other organisms (Romanowski *et al.*, 1996). It is possible, however, that TbORC1B expression and availability is cell cycle regulated, as is the case for the human and *D. melanogaster* Orc1 subunits (Tatsumi *et al.*, 2000; Asano and Wharton, 1999). In order to better understand TbORC1B^{12myc} localisation dynamics, intensity of the DAPI and anti-myc signals were analysed as described in section 3.5.5. Similarly to 927 wt, and all other cell lines (Figure 3.33 to Figure 3.38), the TbORC1B 12myc cell line showed a normal DAPI intensity profile (Figure 3.42, B), supporting the idea that this cell line progresses normally through the cell cycle (Figure 3.22, C). Measurement of the intensity of anti-myc signal, however, revealed a different profile from any of the other putative Orc-like factors (Figure 3.42, C). In contrast to TbORC1/CDC6^{12myc} and the other factors, where the intensity peaked in 1N2K cells, the TbORC1B^{12myc} signal peaked in 1N1eK cells and decreased in 1N2K cells (the difference between these two cell types is not statistically significant, Figure 3.42, C). Consistent with the lack of visual detectable TbORC1B^{12myc} signal in 2N2K cells, the signal was indistinguishable from background in 2N2K cells (Figure 3.42, C). Indeed, the average signal in 1N1K cells was also at background levels, with some cells showing higher levels, consistent with TbORC1B^{12myc} being expressed in only a fraction of this cell type. The signal intensity in 1N1K and

2N2K cells (at or near background signal) was statistically different from the signal obtained in 1N1eK and 1N2K cells (which is visually detectable, Figure 3.42, C). In Figure 3.42, C, many (n >400) cells were counted and, within these, cells that were categorised visually as being TbORC1B^{12myc} positive (values used to generate Figure 3.30 and Figure 3.41, B and C) are represented as light green, whereas cells with TbORC1B^{12myc} signal below visual detection are shown as dark green, most of which had an intensity below background signal. These findings are consistent with the fact that, unlike for TbORC1/CDC6^{12myc} and its other interacting factors, TbORC1B^{12myc} signal is undetectable in many cells. Indeed, it remains possible that the 1N1eK cells that showed signal below visual detection (dark green) might represent a group of miss-categorised cells, as discussed above. These data are derived from one experiment, but a second experiment is shown in the appendices, section 7.6, for comparison.

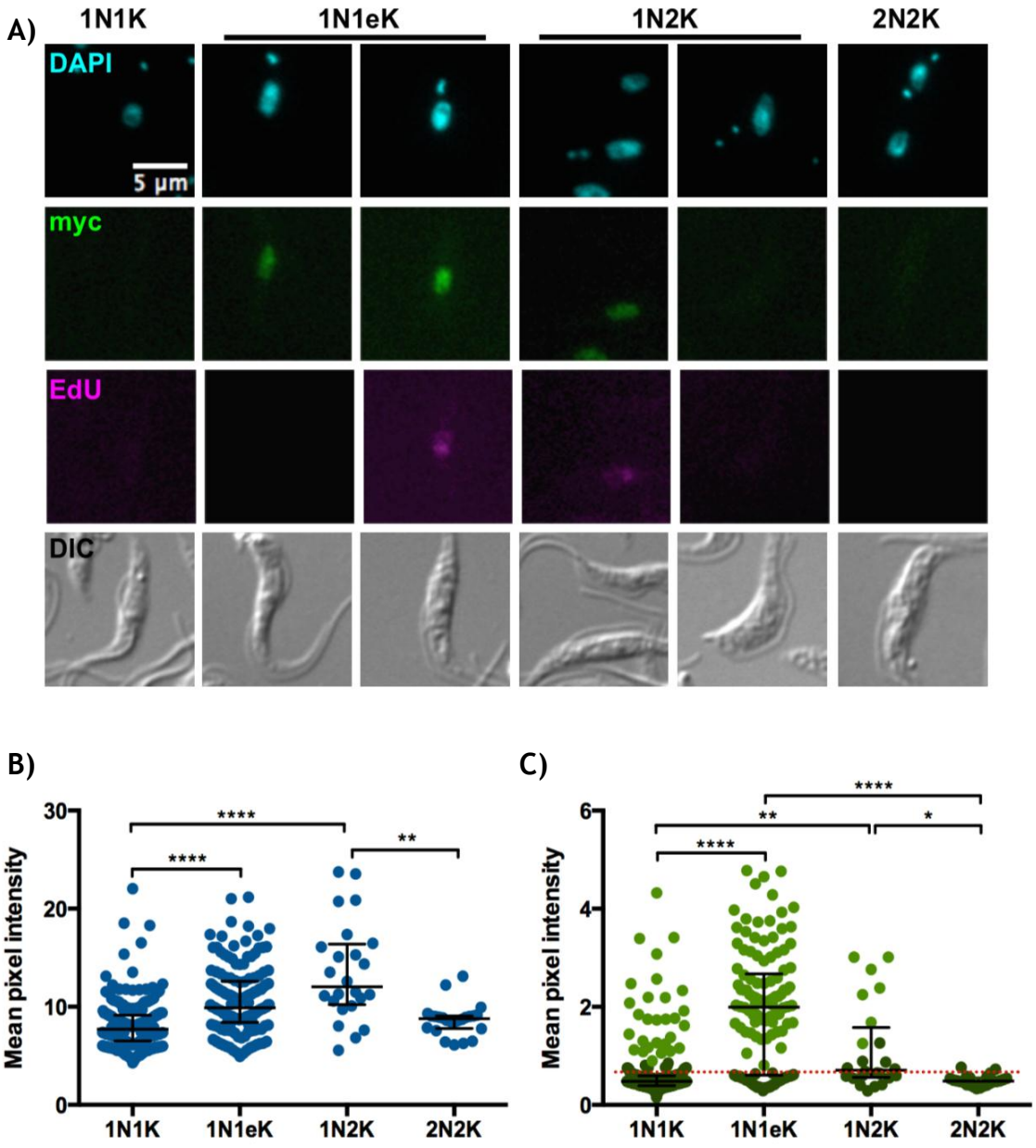


Figure 3.42. TbORC1B^{12myc} subcellular localisation and dynamics throughout the cell cycle. A) staining of TbORC1B 12myc cells with DAPI (top panel), AlexaFluor® 488-conjugated anti-myc antibody (second panel), AlexaFluor® 555-conjugated azide (for EdU detection; third panel), and the cell outline by DIC (bottom panel). All images were acquired using an Axioskop 2 imaging system. Scale bar represents 5 µm. Intensity of the DAPI (B) and myc (C) signals is represented (dots) as the mean of pixel intensity within the circular region of interest (ROI, of 21 x 21 pixels), drawn around each individual cell nucleus. In C), the red dotted line represents the average background signal measured in the 927 wt cell line, dark green dots represent the cells in which TbORC1B^{12myc} signal is not visually detected, while light green dots represent the cells with TbORC1B^{12myc} signal strong enough to be detected by eye. A total of 412 cells were analysed (n = 412). In both B) and C) the median of the values is represented, with the error bars depicting the interquartile range. Statistical significance between the different cell cycle stages was assessed through analysis using the Kruskal-Wallis non-parametric test. (*) p-value < 0.05; (**) p-value < 0.01; (***) p-value < 0.001; (****) p-value <0.0001. A), B) and C) show results obtained from the same experiment. Intensity of the EdU signal is shown in the appendices, Figure 7.40.

3.5.6.3 TbORC1B^{12myc} expression appears to be cell cycle-regulated

The data from localisation and signal intensity measurements of TbORC1B^{12myc} suggest that the expression of this factor might be cell cycle regulated, or that the protein becomes detectable only when it localises in the nucleus. In order to examine this further, different cell cycle stages were isolated and analysed by western blot for the presence of TbORC1B^{12myc}. Briefly, TbORC1B 12myc cells were set up in culture at a starting concentration of 1 x 10^6 cells.ml^{-1}, and used for the experiment when at a concentration of approximately 1 x 10^7 cells.ml^{-1}. Around 10^9 cells were collected from the asynchronous culture, fixed in 70% methanol, and stained with propidium iodide (PI), as detailed in the materials and methods Chapter 2, sections 2.7.1 and 2.7.5. The cells were then analysed and separated in G1, early S, late S and G2/M phase fractions (Figure 3.43, A), by using a BD FACS Aria I™ flow cytometer sorting system (BD Biosciences). In order to avoid contamination of the G1 phase population with cells in S phase, the G1 phase "gate" was designed to capture cells that were on the left hand side of the G1 peak of the histogram profile (Figure 3.43, A). It is important to note that the cells sorted into the G2/M phase population include both 1N2K cells that have completed DNA replication and 2N2K cells that have not completed cytokinesis, as these are detected by doubled DNA content relative to G1. The different cell cycle fractions were then analysed by western blot as described in the materials and methods, Chapter 2, section 2.2.5.3 (Figure 3.43, B). As a loading control, the translation elongation factor 1 alpha (Ef1α, gene Tb927.10.2100) was detected using a commercially available antibody. These data showed that TbORC1B^{12myc} was undetectable in the G1 (2N content) fraction, while it was strongly present in both early and late S phase fractions (Figure 3.43, B). TbORC1B^{12myc} was also detected in the G2/M (4N content) fraction (though perhaps in slightly reduced levels); since no immunofluorescence signal could be seen in 2N2K cells (previous section), these data perhaps suggest that TbORC1B^{12myc} expression continues through G2 phase (Figure 3.43, B). These data strongly support the hypothesis that TbORC1B expression might be cell cycle regulated, though at present the nature of this regulation is unknown. Transcriptome information (Archer *et al.*, 2011) suggests that *TbORC1B* mRNA levels are highest in late G1, reduce somewhat in S phase and are lowest in early G1 and G2 phases (appendices, Figure 7.35). Such fluctuations would need confirmation in the TbORC1B 12myc cell line, for

instance by RT-qPCR using primers for the 12myc tag, such as described in section 3.3.2. However, whether or not mRNA level control can explain the changing proteins levels documented here is unclear.

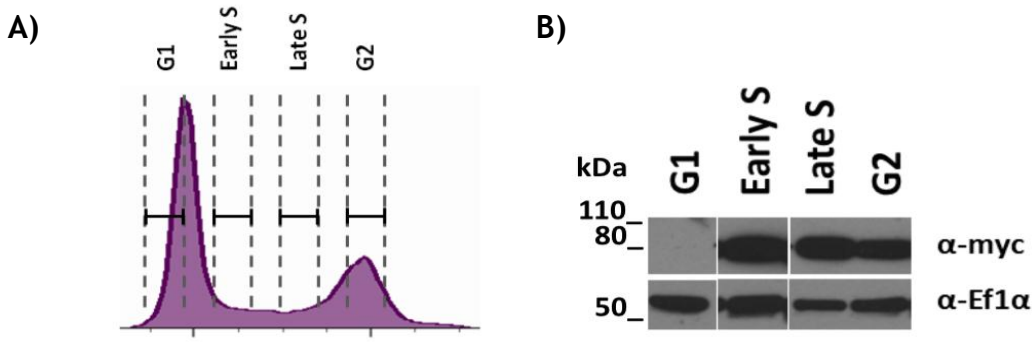


Figure 3.43. TbORC1B expression appears to be regulated throughout the cell cycle.
A) Histogram depicting the cell cycle profile of a mid-log phase PCF cell culture, reflecting the fluorescent signal emitted by the staining of DNA with propidium iodide (PI). Dashed vertical lines represent the gates defined to isolate the different cell cycle populations using a FACS Aria I sorting flow cytometer machine. The asynchronous cell population was separated into G1, early S, late S and G2/M phases. B) The sorted fractions of the cell cycle were analysed by western blot for the detection of TbORC1B^{12myc}. The transcription elongation factor Ef1α was used as a loading control for each fraction.

3.5.6.4 TbORC1B^{12myc} appears to not co-localise to newly replicated DNA

Together, the data here presented strongly suggests that TbORC1B expression and cellular localisation are cell cycle regulated, in contrast with TbORC1/CDC6 and all the other analysed factors, which localise to the nucleus throughout the cell cycle. It appears that TbORC1B expression is limited to a period of the cell cycle stretching from the beginning of (or immediately before) S phase until late S or G2. This suggests that TbORC1B might have a regulatory function rather than being a static member of a potential ORC-like complex. In section 3.5.2, it was shown that TbORC1/CDC6^{12myc} appeared to co-localise in the nucleus with at least some regions of newly replicated (EdU positive) DNA in 1N1eK cells (Figure 3.24). A similar overlay of anti-myc and newly synthesised EdU-labelled DNA in 1N1eK and 1N2K TbORC1B^{12myc} cells did not suggest the same extent of co-localisation (Figure 3.44): by eye, the extent of overlap of TbORC1B^{12myc} and EdU signals appeared less than for that seen for TbORC1/CDC6^{12myc} and EdU, although further less subjective tests would be needed to validate this.

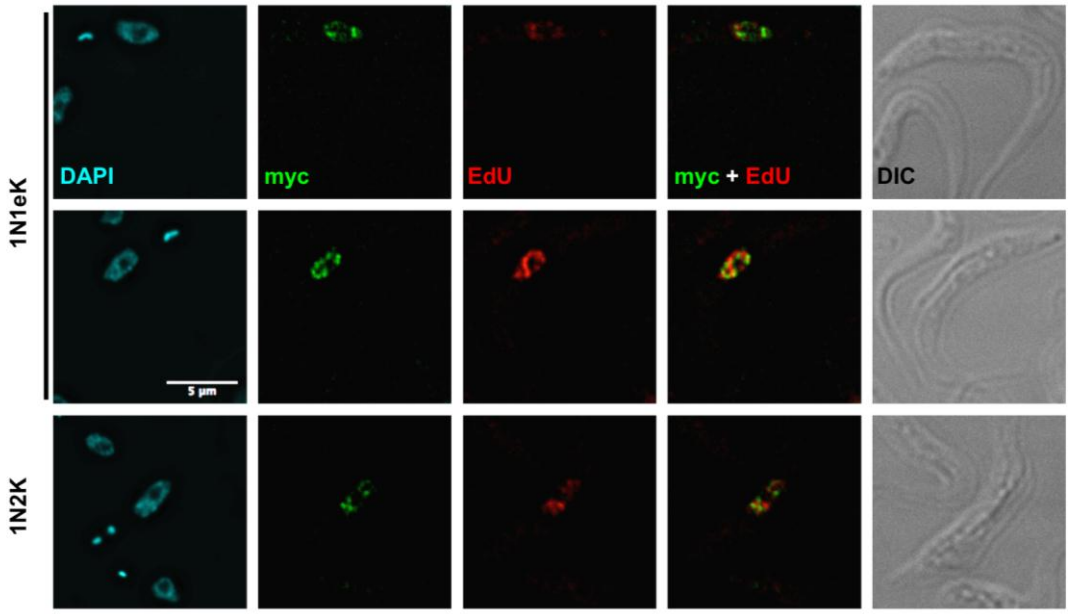


Figure 3.44. Immunofluorescence of TbORC1B^{12myc} and newly synthesised DNA.
Two representative 1N1eK cells are shown in the top panel stained with DAPI (far left panel), anti-myc antiserum (middle left panel), and AlexaFluor® 594 azide, detecting EdU (middle panel). The middle right panel depicts a merge of the myc and EdU staining, where the lack of yellow regions does not suggest co-localisation between the TbORC1B^{12myc} and newly synthesised DNA. Cell outline is shown in the DIC image, in the far right panel. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

3.5.6.5 TbORC1B^{6HA} co-localisation with TbORC1/CDC6^{12myc}

By now, it was clear that TbORC1B appears to be different from all the other factors that have also been analysed, including TbORC1/CDC6. Using a cell line with TbORC1/CDC6 endogenously tagged with 12myc and TbORC1B tagged with 6HA (described in section 3.4.5), it was asked whether these two factors co-localise, as has been described for the other factors in section 3.5.3. Again, it was not possible to acquire better resolution images with, for instance, the DeltaVision imaging system. Nevertheless, images acquired with the Axioskop system showed that TbORC1B^{6HA} presents the same subcellular pattern as the 12myc tagged version of the protein: it is detected, as puncta, in the nucleus of 1N1K, 1N1eK and 1N2K cells (Figure 3.45). Overlay of the TbORC1/CDC6^{12myc} and TbORC1B^{6HA} signals suggested that the two proteins localise to the same regions of the nucleus, but it was not clear whether the factors show as extensive overlap as TbORC1/CDC6^{12myc} displayed with the other putative Orc-like factors. Higher resolution imaging would be crucial to examine this is more detail.

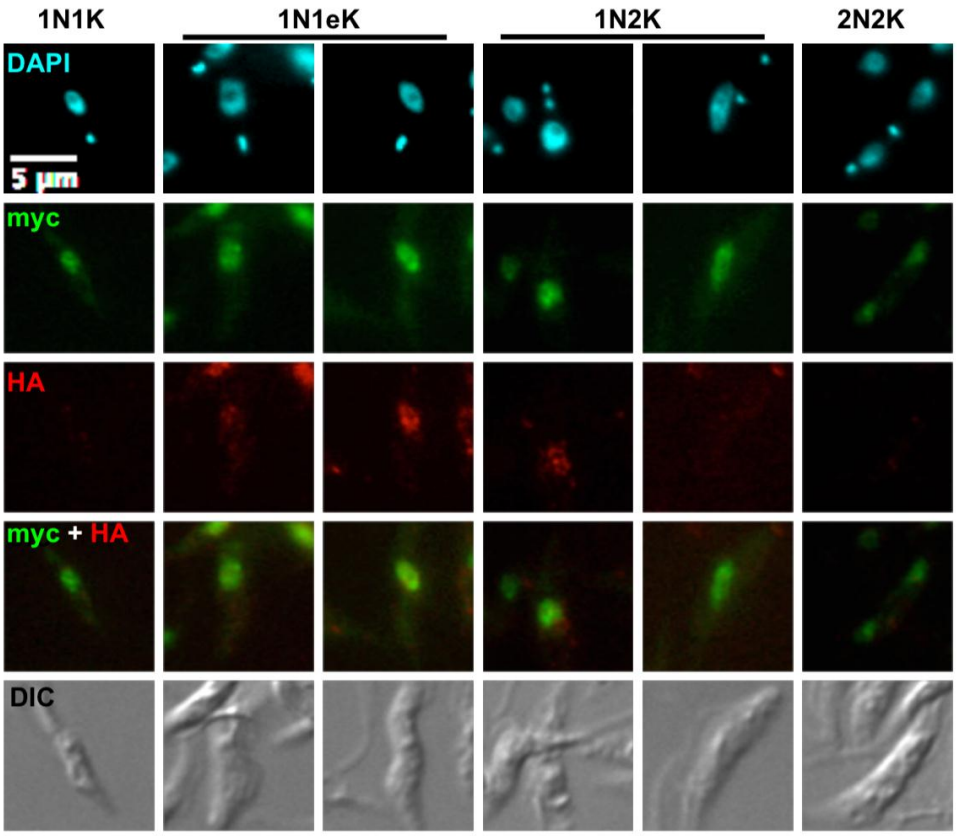


Figure 3.45. Immunofluorescence of TbORC1/CDC6^{12myc} and TbORC1B^{6HA}.
Top panel shows the staining of the cells with DAPI. The second panel shows TbORC1/CDC6^{12myc}, recognised by the anti-myc antiserum. The third panel shows the signal correspondent to the TbORC1B^{6HA}. The fourth panel shows a merge of the signals detected with both the anti-myc and the anti-HA antisera. The bottom panel row shows the cells outline by DIC. Images were acquired using an Axioskop 2 imaging system. Scale bar represents 5 µm.

3.6 Do TbORC1/CDC6, TbORC4, Tb7890, Tb3120, Tb1120, and TbORC1B interact in a complex?

To date, TbORC1/CDC6 has been shown to interact with TbORC1B, TbORC4, Tb7980 and Tb3120 (Dang and Li, 2011; Tiengwe *et al.*, 2012b). TbORC1B interaction with TbORC1/CDC6 was examined by co-immunoprecipitation (co-IP) using a cell line with TbORC1/CDC6 endogenously tagged with PTP and TbORC1B tagged with three copies of the HA tag (3HA) (Dang and Li, 2011). In contrast, TbORC4, Tb7980 and Tb3120 interaction with TbORC1/CDC6 was examined through co-IP using TbORC1/CDC6^{12myc} and each of the factors tagged with 6HA (Tiengwe *et al.*, 2012b). None of these experiments however, addressed whether all these proteins interact with each other in a complex.

From the TbORC1/CDC6^{12myc} immunoprecipitation (IP) assay (Tiengwe, 2010), more hits were retrieved than just TbORC4, Tb7980 and Tb3120. Amongst the other hits was Tb1120, which was the protein recovered that displayed the highest number of identified peptides (14 peptides) (Tiengwe, 2010). Tb1120, however, was not further investigated and its interaction with TbORC1/CDC6 was not confirmed. Interestingly, no peptides corresponding to TbORC1B were recovered from TbORC1/CDC6^{12myc} IP hits (Tiengwe, 2010).

3.6.1 Co-immunoprecipitation of TbORC1/CDC6^{12myc} with TbORC4^{6HA}, Tb7980^{6HA}, Tb3120^{6HA}, Tb1120^{6HA}, or TbORC1B^{6HA}

In order to confirm the interaction between TbORC1/CDC6 and Tb1120, co-IP using the TbORC1/CDC6 -/12myc Tb1120 6HA cell line was performed. Likewise, the interaction between TbORC1/CDC6^{12myc} and TbORC1B^{6HA} was also investigated, to check that the proteins interact with this combination of epitopes. As controls, the following cell lines were also used: the 927 wt cell line, which does not possess any tag, as a negative control for both HA and myc detection; TbORC1/CDC6 -/12myc, as a negative control for HA; Tb7980 6HA[l], as a negative control for myc; and TbORC1/CDC6 -/12myc TbORC4 6HA, TbORC1/CDC6 -/12myc Tb7980 6HA, and TbORC1/CDC6 -/12myc Tb3120 6HA, as positive controls (Tiengwe *et al.*, 2012b). The IP, using the HA-tagged proteins

[l] Cell line generated by Dr Calvin Tiengwe during the course of his PhD.

as “bait”, was performed as described in (Tiengwe *et al.*, 2012b; Tiengwe, 2010), with few modifications. Around 5 x 10^8 cells of each cell line were collected from cultures with a density of around 1 x 10^7 cells.ml^{-1}, and lysed in 1 ml whole cell extract buffer (WCE buffer), for 30 minutes at 4°C, as described in the materials and methods Chapter 2, section 2.6.2. In this case, 1 mM of DTT (as described in (Dang and Li, 2011)) and 2x protease inhibitors were added to the WCE solution described in (Tiengwe *et al.*, 2012b). The new variation of the lysis solution was tested and confirmed to not interfere with the binding of the proteins to the magnetic beads coated with anti-HA antibodies used in the assay (not shown). For each cell line, a sample was taken from the cells prior to (PL – pre-lysis) and after lysis (I – input or lysate), to be used as a positive control of the presence of the proteins of interest before and after the cells were lysed. Each lysate (input for the IP) was then incubated with magnetic Dynabeads® M-280 sheep anti-mouse IgG (which had been previously coated with mouse anti-HA antiserum) for 2 h at 4°C, as detailed in the materials and methods, Chapter 2, section 2.6.2. A sample was collected at the end of the incubation to be used as flowthrough sample (F- flowthrough), representing the protein’s fraction that had not attached to the coated beads. The beads were further washed with a washing solution identical to the WCE solution, with the exception of the amount of detergent present (0.1% Triton X-100) and the absence of DTT. As already mentioned, IP of TbORC1/CDC6^{12myc} failed to identify TbORC1B as a potential interacting partner (Tiengwe, 2010; Tiengwe *et al.*, 2012b). In the study by (Tiengwe *et al.*, 2012b), washes were performed using a high salt containing solution, with 500 mM LiCl and 0.7% sodium deoxycholate, an anionic detergent (Tiengwe *et al.*, 2012b). This contrasted with the less stringent solution used in (Dang and Li, 2011), which contained only 100 mM of NaCl and 1% NP-40. As it is possible that the high levels of salt and strong detergent in (Tiengwe *et al.*, 2012b) might have disturbed the interaction between TbORC1/CDC6^{12myc} and TbORC1B, the solution used here for washing contained only 100 mM NaCl, like the WCE solution. After washing, the proteins bound to the beads were eluted directly in protein loading buffer (E – elution sample), separated on an SDS-PAGE gel, and analysed by western blot, as detailed in the materials and methods, Chapter 2, sections 2.2.5.3. Membranes were first probed with anti-myc antiserum, and then stripped before being incubated with the anti-HA antiserum.

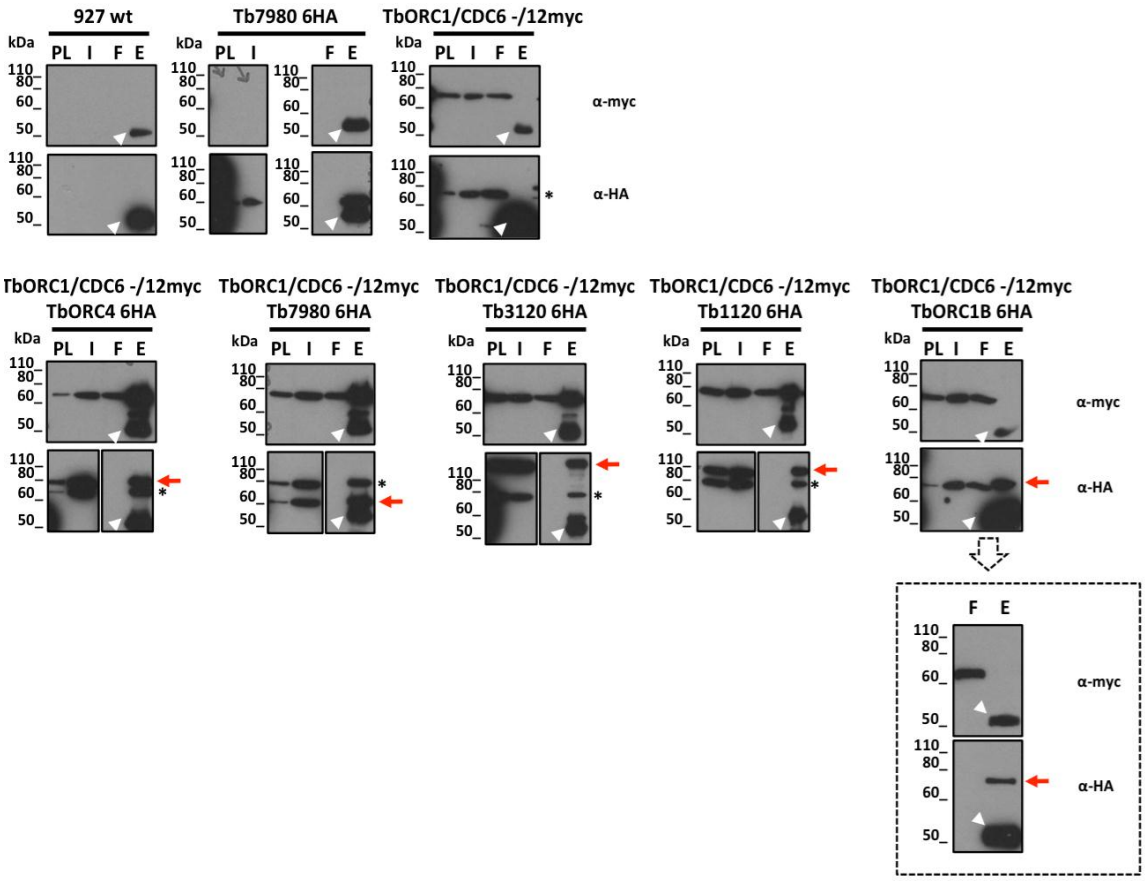


Figure 3.46. Immunoprecipitation of HA-tagged proteins, and co-detection of TbORC1/CDC6^{12myc}.
Cell lines were lysed and the HA-tagged proteins were immunoprecipitated as described in the main text and in the materials and methods chapter. PL – pre-lysis sample; I – input or lysate sample; F – flowthrough sample; E – eluate sample. Each cell line is represented separately, with the top panel depicting the membrane probed for myc detection, and the lower panel showing the membrane probed for HA detection. Anti-myc and anti-HA are represented as α-myc and α-HA, respectively. (*) highlights the detection of TbORC1/CDC6^{12myc} in the membrane probed for HA detection, still present after stripping of the membrane post-myc detection. While arrows call the attention of the reader to the anti-HA antibody heavy chain, which was eluted with the remaining proteins since the antibody was not cross-linked to the magnetic beads used for the IP. Red arrows signalise the HA-tagged proteins. A second, independently performed IP experiment is shown for the TbORC1/CDC6 -/12myc TbORC1B 6HA cell line, within the dashed box.

Figure 3.46 shows that no signal was detected for myc or HA in the lysate of the 927 wt cell line, showing again that both antibodies are specific for each of these tags. The band (~50 kDa) present in both anti-myc and anti-HA panels corresponds to the heavy chain of the anti-HA antibody used in the IP, as this was not cross-linked to the magnetic beads, and was therefore eluted together with the bound proteins (highlighted in Figure 3.46, white arrows). In the case of the cell line expressing only Tb7980^{6HA}, the protein was efficiently recognised by the anti-HA antibody bound to the magnetic beads (E fraction), and was not detected in the membrane probed with anti-myc, showing that the anti-myc antiserum did not recognise the HA tagged protein. The absence of signal in the flowthrough (F) fraction may suggest that most of the HA-tagged protein was

successfully attached to the antibody bound to the beads during the IP. On the other hand, IP using the TbORC1/CDC6 -/12myc cell line showed that TbORC1/CDC6^{12myc} was not detected in the eluate fraction in both membranes probed either with anti-myc or anti-HA, confirming that it did not bind non-specifically to the anti-HA-coated beads. Because the same nitrocellulose membrane was used for sequential myc and HA detection, it was stripped between incubation with the two antisera. Because both anti-myc and anti-HA antibodies used in this study were raised in mice, the same anti-mouse HRP-conjugated antiserum was used to detect both primary antibodies. Despite the fact that the membrane was incubated with the stripping buffer, it is clear that this was not entirely successful, and some anti-myc antibody was still attached to TbORC1/CDC6^{12myc} when the membrane was probed for the detection of the HA tag. The outcome was the detection of TbORC1/CDC6^{12myc} in the PL, L and F fractions after probing for HA. This was also observed for all other cell lines containing TbORC1/CDC6^{12myc}, and is highlighted in Figure 3.46 with an asterisk. Results from the IP of TbORC4^{6HA}, Tb7980^{6HA} and Tb3120^{6HA}, confirm that these three factors interact with TbORC1/CDC6^{12myc}, as the latter is detected in the elution (E) fractions in the anti-myc probed membrane (Figure 3.46), and each individual HA-tagged factors is detected in the E fraction of the anti-HA probed membrane (highlighted with a red arrow, Figure 3.46). Likewise, Tb1120^{6HA} was shown to interact with TbORC1/CDC6^{12myc} (Figure 3.46), validating the IP and mass spectrometry detection of this interaction (Tiengwe, 2010). Surprisingly, no TbORC1/CDC6^{12myc} could be detected in the E fraction from the IP of TbORC1B^{6HA}, although the latter was detected in this fraction, confirming that TbORC1B^{6HA} was successfully attached to the anti-HA-coated beads (Figure 3.46). Because the TbORC1B^{6HA} protein is only slightly larger than TbORC1/CDC6^{12myc}, it is not clear whether the bands seen in the PL, L and F fractions in the membrane probed with anti-HA represent detection of TbORC1B^{6HA} or remnants of TbORC1/CDC6^{12myc} still bound by anti-myc antibodies and not removed by the stripping treatment of the membrane. Nevertheless, a repeat of the same experiment failed to co-IP TbORC1B^{6HA} and TbORC1/CDC^{12myc}, as shown in Figure 3.46 (panels within the dashed box). These results support the non-detection of TbORC1B by (Tiengwe *et al.*, 2012b), and question the proposed interaction between TbORC1B and TbORC1/CDC6, which was reported using similar lysis and washing conditions (Dang and Li, 2011) as the ones used here.

3.6.2 Immunoprecipitation of TbORC1/CDC6^{12myc}, TbORC4^{12myc} and TbORC1B^{12myc}

One potential way to ask if the putative Orc-like factors interact together in a complex would be to tag them in pairs, e.g. with 12myc and 6HA, and perform co-IPs with all different combinations. However, this would require generating cell lines expressing all the possible combinations between the known factors, which would be very time consuming. As an alternative, it was decided to IP selected factors, separate the eluted samples on SDS-PAGE gels, and identify interacting proteins by mass spectrometry, as was done previously for TbORC1/CDC6^{12myc} (Tiengwe *et al.*, 2012b). It was decided to perform the IPs using the 12myc tagged cell lines because detection of myc-tagged protein has always been easier than HA-tagged ones (perhaps due to the efficiency of the antisera), and to focus on TbORC4 and TbORC1B.

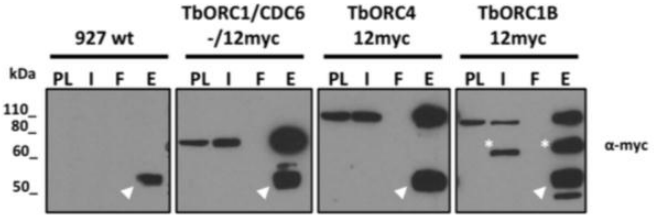
As a starting point, IP of TbORC1/CDC6^{12myc} was performed as described before (Tiengwe *et al.*, 2012b), using the TbORC1/CDC6 -/12myc cell line, in parallel with the 927 wt cell line, used as a negative control in order to identify non-specific binding of proteins to the magnetic beads coated with anti-myc antibody. Unfortunately, no differences were seen between the two IP samples when compared on coomassie-stained gels, despite the fact that the IP worked, as confirmed by western blot (data not shown). Other attempts were performed using different amounts of cells and concentrations of salt (both NaCl and LiCl), without improving specificity. In an attempt to decrease the non-specific binding of proteins to the beads, the cell lysates were incubated with magnetic beads that had not been coated with the anti-myc antiserum, prior to incubation with the anti-myc coated magnetic beads. Again, this did not allow the identification of specific bands once the eluate samples had been separated in SDS-PAGE gels and stained with coomassie blue (not shown). Finally, the IP conditions described in the previous section were used but this time the resulting SDS-PAGE gel was stained with SYPRO® Ruby, instead of coomassie blue; though there was still considerable overlap between the banding patterns from the 927 wt and myc-tagged cells, some apparently specific protein bands were seen in the latter IPs (Figure 3.47, B). The IP conditions used were those described for the co-IP experiment discussed in the previous section, though 1.5 x 10^9 cells were used and lysed in 3 ml of WCE solution (each 1 ml of lysate incubated with a set of

coated-beads), and the magnetic beads were coated with anti-myc antibody, as described in the materials and methods, Chapter 2, section 2.6.1. The success of the IP was confirmed by western blot, with proteins of the sizes expected for the myc-tagged variants recovered from the cognate cell line and absent from the 927 wt cells (Figure 3.47, A). As discussed in section 3.5.4, two bands were detected in the TbORC1B 12myc cell line IP: one most likely corresponding to the full size TbORC1B^{12myc} protein (running between 80 and 110 kDa), and a smaller band, just larger than 60 kDa (Figure 3.47, A, highlighted by a white asterisk). The smaller band was not detected in the pre-lysis sample (PL), suggesting that it may have arisen due to degradation during the experiment course. Due to the large numbers of bands in each gel, a section of each lane (Figure 3.47, B, highlighted by the parentheses), instead of individual bands, was excised and sent for nanoflow liquid chromatography-electrospray tandem mass spectrometry (nLS- ES MS/MS) analysis (Glasgow Polyomics, University of Glasgow; Performed and initial analysis by Dr Richard Burchmore, Dr Christina Naula and Dr Stefan Weidt). A preliminary analysis was performed, but a thorough analysis will be needed in the future.

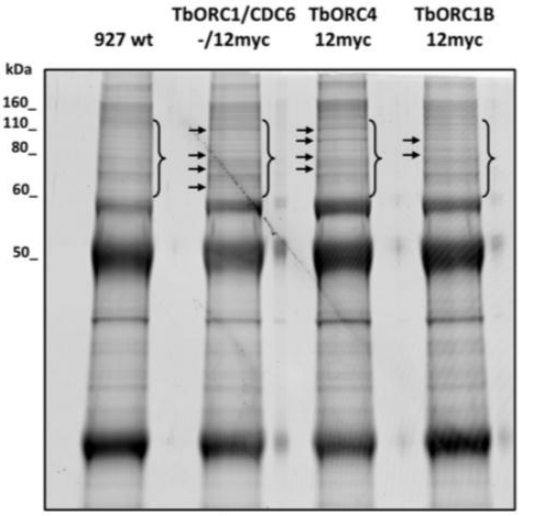
Figure 3.47, B, shows a summary of the "hits" obtained for each IP that were not present in the non-tagged cell line sample (927 wt), and thus were precipitated most likely due to unspecific binding to the coated beads, but were precipitated by each of the 12myc tagged proteins. Because the section of the gel excised and sent for analysis enclosed only proteins larger than 60-65 kDa, the number of hits resultant from the IP of TbORC1/CDC6^{12myc} here shown was smaller than in (Tiengwe, 2010). Nevertheless, it was possible to identify Tb3120, and a set of other proteins that did not appear in the previous analysis (Tiengwe, 2010) (Figure 3.47, B). It was not possible, however, to identify peptides from TbORC4 and Tb1120, though these proteins are larger than 60 kDa. Moreover, TbORC1/CDC6 was not detected in the sample, though TbORC1/CDC6^{12myc} is ~66 kDa in size. IP of TbORC4^{12myc} was able to pull down TbORC4 itself as well as Tb1120 and Tb3120, besides other hits (Figure 3.47, B). These data support the idea that TbORC4, Tb1120, Tb3120 and TbORC1/CDC6 interact in a complex. However, neither TbORC1/CDC6 nor TbORC1B were retrieved from this analysis, most likely because as TbORC1/CDC6 is ~48 kDa and TbORC1B is ~65.5 kDa, and were therefore excluded from the sampled region of the gel. TbORC1B^{12myc} IP

did not retrieve any of the Orc-like factors, perhaps indicating that it is not a part of a stable complex with these factors. Nevertheless, IP of each of the three proteins, TbORC1/CDC6^{12myc}, TbORC4^{12myc} and TbORC1B^{12myc}, retrieved a common hit, which was absent from the non-tagged control sample (927 wt): Tb927.10.2240, an hypothetical protein according to data on TriTrypDB v9.0 (as of February 2015). InterPro analysis (as described in section 3.2.3) suggests that it has a nuclear transporter factor 2 (NTF2) domain at its N-terminus, as well as a RAS GTPase-activating protein-binding protein domain, covering the protein's full length. Proteins from these groups have been involved in various roles including nuclear transport of other proteins. It will be necessary to validate the interaction of this factor with TbORC1/CDC6, TbORC4 and TbORC1B, and assess whether it may be an extra member of the putative ORC-like complex, and which role might it have in DNA replication. It is nevertheless surprising that it was not detected in the previous TbORC1/CDC6^{12myc} IP, in which the whole gel lane was analysed.

A)



B)



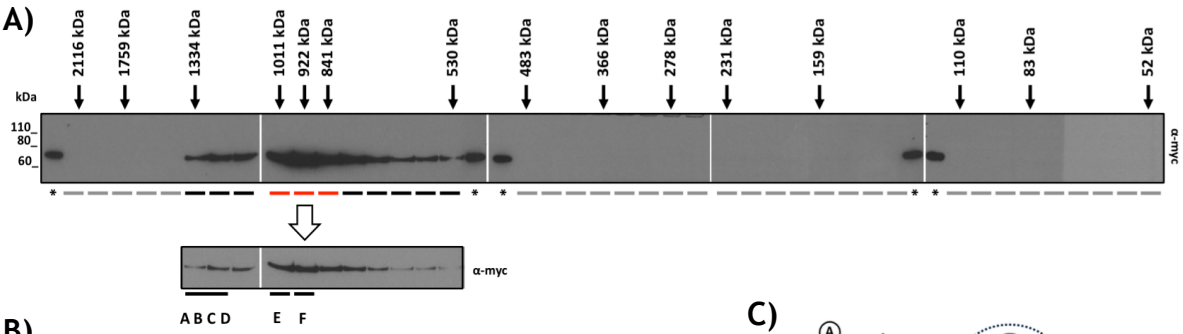
TbORC1/CDC6 12myc			TbORC4 12myc			TbORC1B 12myc		
Gene ID	Name	Mw (kDa)	Gene ID	Name	Mw (kDa)	Gene ID	Name	Mw (kDa)
Tb927.10.2240	Hypothetical protein	66.35	**Tb927.10.13380**	**TbORC4**	**77.29**	Tb927.11.6280	Pyruvate phospate dikinase (PPDK)	100.45
Tb927.8.7590	Receptor-type adenylate cyclase GRESAG 4	139.52	Tb927.10.2240	Hypothetical protein	66.35	Tb927.10.3360	Hypothetical protein	102.76
Tb927.9.4530	**Tb3120**	**112.94**	**Tb927.6.1120**	**Tb1120**	**80.70**	Tb927.10.2240	Hypothetical protein	66.35
Tb927.6.2550	RNA-binding protein (TRRM2)	79.85	**Tb927.9.4530**	**Tb3120**	**112.94**	Tb927.5.930	NADH-dependent fumarate reductase (FRDg)	123.69
Tb927.8.4320	Hypothetical protein	167.70	Tb927.1.200	Retrotransposon hot spot protein (RHS, pseudogene)	236.16	Tb927.3.2620	Hypothetical protein	216.87
Tb927.10.11780	Hypothetical protein	131.275						
Tb927.1.1710	Transmembrane protein	195.44						

Figure 3.47. Immunoprecipitation of TbORC1/CDC6^{12myc}, TbORC4^{12myc} and TbORC1B$^{12\ myc}$. A) Confirmation of the success of the IP using magnetic beads coated with anti-myc antibody by western blot. PL – pre-lysis sample; I – input or lysate sample; F – flowthrough sample; E – eluate sample. Each cell line is represented separately. White arrows call the attention of the reader to the anti-myc antibody heavy chain (band just above the 50 kDa size marker), which was eluted with the remaining proteins since the antibody was not cross-linked to the magnetic beads used for the IP. White asterisks call the attention of the reader to the second, smaller band, observed in the TbORC1B 12myc sample. B) SDS-PAGE gel stained with SYPRO® Ruby showing the eluate fractions obtained from the IP. The dark bands around 50 kDa and close to the end of the gel correspond to the antibody, which was eluted with the proteins attached to the beads. Black arrows pinpoint potential specific bands not present in the untagged cell line, 927 wt. Parentheses demark the region of each lane that was extracted for further mass spectrometry (MS) analysis. Results from MS analysis, showing hits not present in the non-tagged cell line (927 wt), but identified in each of the other proteins' IP samples. TbORC4 is highlighted in yellow, Tb3120 in blue and Tb1120 in pink. A common hit to the three cell lines IP was found, and is highlighted in green.

3.6.3 Size exclusion chromatography, or Gel Filtration

Despite the common hits from the different IPs described above, the approach used does not address if the proteins interact in a complex or individually. To address this, a gel filtration analysis was conducted. Approximately 7.5 x 10^8 cells were collected from a TbORC1/CDC6 -/12myc cell line culture at a concentration of 1 x 10^7 cells.ml^{-1} and lysed in 2 ml of WCE solution, as described in the materials and methods, Chapter 2, section 2.6.4.3. The lysate was then centrifuged and filtered, in order to remove any particulate material, and the resultant supernatant was then subjected to size exclusion chromatography through a HiLoad 16/60 Superdex 200 Prep Grade column (GE Healthcare Life Sciences) using a flow rate of 500 µl per minute, as described in materials and methods, Chapter 2, section 2.6.4.3. Fractions of 1 ml were collected and analysed for the presence of TbORC1/CDC6^{12myc} by western blot using anti-myc antiserum; as a positive control, a sample taken from the lysate was run per gel. All steps related to the gel filtration technique, including column preparation, system calibration, and system set up, were performed by Mr Alan Scott who, together with Dr Jaspreet Grewal and Dr Jeziel Damasceno, helped with the analysis of the elution profile. Based on the profile recovered (shown in the appendices, Figure 7.41), appropriate samples to be analysed by western blot were chosen. These samples corresponded to volumes from immediately before the void volume (corresponding to ~2116 kDa; peak at 43 ml), until samples corresponding to 48 kDa (84 ml peak), according to the size standards used (Figure 3.48, A). TbORC1/CDC6^{12myc} was detected in fractions corresponding to proteins ranging from ~1334 kDa to 530 kDa, having a peak of intensity in the fractions corresponding to ~1011 kDa to 841 kDa (Figure 3.48, A). Similar to what was reported for *S. pombe* SpOrc1 (Lygerou and Nurse, 1999), the monomeric form of TbORC1/CDC6^{12myc} (around 66 kDa) was not detected (Figure 3.48, A), suggesting that the protein is mainly (or exclusively) present in a complex with other factors. The broad range of TbORC1/CDC6^{12myc} signal may suggest complexes (or sub-complexes) of various sizes. However, the predominance of signal at around 900-1000 kDa suggests a predominant multiprotein species. The composition of this putative complex requires further analysis, but if TbORC1/CDC6^{12myc} formed an ORC-like complex with only TbORC4, Tb7980, Tb3120 and Tb1120, the resulting size would be only of 386

kDa, smaller than what was seen. If TbORC1B were also part of this putative complex, a species of 451 kDa would be predicted, again smaller than the ones observed (Figure 3.48, A). The individual sizes of several further replication-related factors, including the TbMCM2-7 helicase complex, TbCDC45, the TbGINS complex, and TbMCM10 are shown in Figure 3.48, B, and then combined with the putative ORC-like complex to predict the resulting molecular size of potential complexes (Figure 3.48,C). Amongst these, the putative pre-RC, composed of "ORC" and the MCM-2-7 helicase would be ~1003 kDa to 938 kDa in size, depending on whether TbORC1B is included or not. In turn, putative pre-IC complexes, containing "ORC", the TbMCM2-7 helicase, TbCDC45, TbGINS and either containing or lacking TbMCM10, would range from 1123 kDa to 1267 kDa.



B)

		Potential Complexes					
	Size (kDa)§	A	B	C	D	E	F
TbORC1/CDC6 12myc	66.34	+	+	+	+	+	+
TbORC4	77.29	+	+	+	+	+	+
Tb7980	48.8	+	+	+	+	+	+
Tb3120	112.9	+	+	+	+	+	+
Tb1120	80.7	+	+	+	+	+	+
TbORC1B	65.49	+	-	+	-	+	-
MCM 2-7 complex	522.15	+	+	+	+	+	+
TbCDC45	78.2	+	+	+	+	-	-
GINS complex	107.5	+	+	+	+	-	-
TbMCM10	77.8	+	+	-	-	-	-
Size (kDa)		1267.2	1201.7	1189.4	1123.9	1003.6	938.2

C)

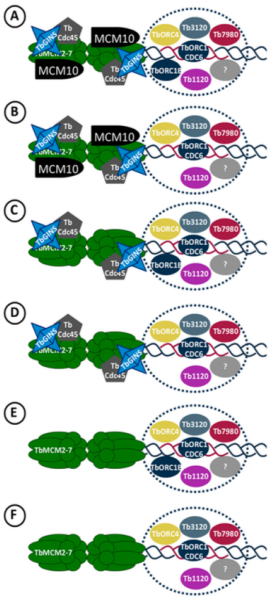


Figure 3.48. Gel filtration of TbORC1/CDC6 -/12myc.
A) Detection of TbORC1/CDC6^{12myc} with the anti-myc (α-myc) antiserum in the different fractions resulting from gel filtration of TbORC1/CDC6 -/12myc cell line lysates. The estimate molecular weight corresponding to relevant fractions is shown next to the corresponding black arrow. Dashes underneath the western blot panels depict the lanes with negative (grey), positive (black) and most intense (red) signals. (*) pinpoints the lanes loaded with the lysate sample, as a positive control in each western blot membrane. The white arrow draws attention to a panel showing the positive lanes, where the signal was acquired with a shorter exposure time. A-F highlight the bands that might correspond to the potential complexes proposed in (B). B) Table showing the molecular weight (in kDa) of the presently known interacting partners of TbORC1/CDC6, the TbMCM2-7 helicase complex, TbCDC45, the TbGINS complex, and TbMCM10, as available in TriTrypDB v9.0 as of January 2015 (§). Columns A to F suggest potential correspondence between these putative

complexes and the signals detected by western blot shown in (A). The different combinations of factors present in the potential complexes are represented as (+, presence) and (-, absence). The colour scheme corresponds to the one used in the diagrams shown in (C). C) Schematic representations of the potential complexes highlighted in (B); not to scale.

3.7 Discussion

This chapter describes experiments that address three central questions in the field of initiation of nuclear DNA replication in *Trypanosoma brucei*, each leading to an overarching question: does an origin recognition complex (ORC) exist in this parasite, or not? The first question was whether TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120, are recognisably related to the six subunits of conventional eukaryotic ORCs. All these factors have been shown to interact with TbORC1/CDC6 in PCF cells of *T. brucei*, but at the outset of this work it was unknown if they interact together in a complex involved in nuclear DNA replication. The second question was: are TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120 involved in DNA replication? Only TbORC1/CDC6 had been directly implicated in DNA replication, both through genome-wide localisation studies and examining the effects of RNAi on BrdU incorporation (Tiengwe *et al.*, 2012a; Benmerzouga *et al.*, 2013). Therefore, similar phenotypes for each of these proteins could be suggestive of the factors acting in the same biological process. The third question addressed asked about the localisation of each factor: was this consistent with a nuclear function, and did the localisation display evidence for cell cycle dependency that might be related to replication? The work described here is limited to PCF *T. brucei*, building on the available data on replication in *T. brucei* and related kinetoplastids, which has been obtained from studies investigating the insect stages of the parasites' life cycles. For most of the questions addressed, only partial answers were revealed. These questions and their answers are discussed in in the next few sections, where potential models are suggested, and new questions are raised.

3.7.1 A fresh analysis into TbORC1/CDC6 and its interacting partners: are these Orc-like proteins?

A new analysis of the protein sequences of TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, and Tb3120, has allowed refinement on the information available for these factors (Godoy *et al.*, 2009; Tiengwe *et al.*, 2012b; Dang and Li, 2011;

Tiengwe *et al.*, 2013). In addition, new analysis was performed on Tb1120, a TbORC1/CDC6 interacting partner that has never been analysed in depth (Tiengwe, 2010). BLAST analysis has confirmed that all these factors are well conserved in other kinetoplastid organisms, such as *T. cruzi* and various species of *Leishmania* (Table 3-1), and that the genes for each are syntenic between, at least, *T. brucei*, *T. cruzi*, and *Leishmania major* (appendices, Figure 7.1).

Altogether, re-analysis of TbORC1/CDC6 and TbORC4 protein sequences further support the categorisation of these factors as Orc-like proteins. TbORC1/CDC6 was recognisably related to Orc1 factors from other eukaryotes and has demonstrated ATPase activity, as well as being able to complement the lack of Cdc6 activity in yeast (Godoy *et al.*, 2009). TbORC4 remained classified as a divergent Orc4 protein, with an apparent degenerate AAA+ ATPase domain containing a putatively conserved arginine finger (Tiengwe *et al.*, 2013). It is most likely that TbORC4 does not possess ATPase activity (Tiengwe *et al.*, 2013), similar to other eukaryotic Orc4 subunits. It has been suggested that Orc4 does not bind ATP itself, most likely because it has a modified P-loop (Walker A motif), which has been shown to be required for Orc1 and Orc5 binding to ATP (Davey *et al.*, 2002b). It has, however, been hypothesised that the conserved arginine finger of Orc4 is necessary for Orc1 ATP hydrolysis, by supplying the arginine finger to Orc1's ATP-binding site (Davey *et al.*, 2002b; Duncker *et al.*, 2009), a scenario consistent with AAA+ ATPases requiring an arginine finger from a binding partner (Iyer *et al.*, 2004). This has been recently supported by the *S. cerevisiae* and *D. melanogaster* ORC structure, derived from cryo-electron microscopy (cryo-EM) and negative-stain electron microscopy, respectively (Sun *et al.*, 2012; Bleichert *et al.*, 2013). In *S. cerevisiae*, the upper lobe of ScORC is suggested to be composed of Orc1, Orc4 and Orc5, the ScORC subunits containing conserved AAA+ ATPase domains, in which the putative arginine finger of Orc4 ATPase domain is facing the putative ATP-binding site of Orc1, although at a distance too remote to allow cooperative ATP hydrolysis. It has been proposed that this is resolved upon ATP hydrolysis by Cdc6, which leads to conformational changes in the ORC complex, bringing the Orc4 arginine finger closer to Orc1's ATP-binding site (Sun *et al.*, 2012). It is possible that a similar core interaction between TbORC4 and TbORC1/CDC6 is conserved in *T. brucei*, and it may be possible to investigate this by using purified proteins. One piece of

evidence supporting a close functional relationship between TbORC1/CDC6 and TbORC4 is that individual silencing of either factor results in strikingly similar phenotypes (see below). Even if this picture is correct, it is complicated by uncertain homology in factors beyond TbORC1/CDC6 and TbORC4.

One area of uncertainty is whether *T. brucei* possesses a Cdc6 orthologue. As mentioned earlier, TbORC1/CDC6 has been shown to have ATPase activity and to complement Cdc6 activity in yeast (Godoy *et al.*, 2009). It is therefore plausible that TbORC1/CDC6 may have double functionality, performing the roles of both Orc1 and Cdc6. Why this separation of function is needed in other eukaryotes but could be absent in *T. brucei* is unclear, though archaeal organisms accommodate all ORC and Cdc6 functions in a single protein, Orc1/Cdc6 (Barry and Bell, 2006). BLAST analysis of TbORC1B mainly retrieved Cdc6 hits, which could suggest it to be the *T. brucei* Cdc6-like factor. However, the failure to detect ATPase activity (Dang and Li, 2011) in purified TbORC1B argues against this. Moreover, though in other eukaryotes Cdc6 is clearly a focus for regulation of ORC and pre-RC function, the pattern of expression of TbORC1B is highly atypical of that described for Cdc6 (see below). Nevertheless, alignment of TbORC1B with other eukaryotic Cdc6 protein sequences allowed the mapping of a putative winged-helix domain (WHD) towards its C-terminus, similar to TbORC1/CDC6. Whether one or both factors can bind to DNA through these putative motifs is unknown, though TbORC1/CDC6 has been shown to bind to the chromatin throughout the cell cycle (Godoy *et al.*, 2009). It is not known however, whether TbORC1B binds to chromatin.

An even greater area of uncertainty is how, if at all, the other putative *T. brucei* Orc-like factors might map onto the ORC architecture as proposed in *S. cerevisiae* (Sun *et al.*, 2012) and *D. melanogaster* (Bleichert *et al.*, 2013; Bleichert *et al.*, 2015). Previously, Tb7980 was suggested to have an AAA+ ATPase domain, though no characteristic signatures of the domain were identified besides a potentially conserved Walker A motif (Tiengwe *et al.*, 2012b). The new analysis suggested that this conservation may be more extensive: weak (indeed non-significant) homology with Orc5 proteins at the C-terminus of Tb7980 allowed alignments with other eukaryotic Orc5 sequences and revealed a putative Walker B motif, as well as a potential arginine finger

(albeit not within the predicted AAA+ ATPase domain). It will be valuable to test if Tb7980 possesses ATPase activity, but at this stage it remains premature to label it as an Orc5 subunit, in particular because it was not possible to investigate whether it is involved in DNA replication (see below). When first identified, the only suggestion of Tb3120 being an ORC-like protein relied on very weak evidence suggesting structural similarity with other archaeal Orc1/Cdc6 proteins. No evidence for an AAA+ ATPase domain was found (Tiengwe *et al.*, 2012b). Fresh analysis based on Pfam domain analysis revealed very weak evidence for two regions of similarity to an Orc2 domain at the C-terminus of the protein, together with a putative WHD domain. From this, alignment of Tb3120 with Orc2 subunits from different eukaryotes revealed the presence of relatively conserved Walker A and Walker B motifs within the putative Orc2 domain, though these did not follow the canonical sequences associated with these motifs, but instead the diverged motif sequences characteristic of various eukaryotes Orc2 subunits (Speck *et al.*, 2005). One aspect of the alignment of Tb3120 with other eukaryotic Orc2 subunits was notable: the protein sequences of Tb3120 and its orthologues in *T. cruzi* (Tc3120) and *L. major* (Lm3120) are considerably larger (1018 aa, 1049 aa, and 1487 aa, respectively) than all the Orc2 subunits analysed (ranging from 576-620 aa). If these kinetoplastid proteins are indeed Orc2-like factors, they appear to have undergone an expansion of their N-termini. Analysis of this region may provide a route to understand if these factors provide kinetoplastid-specific adaptation of Orc2 factors, perhaps affecting the configuration of an ORC-like complex, or influencing other aspects of DNA replication. Finally, sequence analysis of the previously uncharacterised Tb1120, which has now been shown to interact with TbORC1/CDC6, and also possibly with TbORC4 and Tb3120, was fruitless: no replication-related homology was retrieved from any of the sequence analysis performed. Tb1120 remains the most diverged putative Orc-like factor to date, with no evidence linking it to any Orc subunit. RNAi targeting Tb1120 will be important to test if it acts in DNA replication.

One final aspect of sequence divergence is that TbORC1/CDC6 (and indeed TbORC1B, if it is also an Orc1-related factor) (Dang and Li, 2011), lacks the bromo-adjacent homology domain (BAH) in its N-terminus (Godoy *et al.*, 2009), a domain that is assumed to be universally conserved amongst eukaryotic Orc1

subunits (Costa *et al.*, 2013). In the structure of ScORC, it is proposed that ScOrc1's BAH domain is displayed to the 'back' of the complex, where it may allow ScOrc1 interaction with transcription silencing factors and histones, without interfering with the ORC complex bound to the origin (Sun *et al.*, 2012). It remains unclear what the absence of a BAH domain in TbORC1/CDC6 means for *T. brucei* (and indeed kinetoplastid) biology, and for the overall structure of a potential ORC-like complex.

Together, this re-analysis of the sequences of TbORC1/CDC6 and its interacting-partners was able to confirm and enhance the available data, suggesting that TbORC1/CDC6, TbORC1B and TbORC4 are most likely Orc-like proteins. Additionally, very tentative data suggest the possibility of Tb7980 being an Orc5-like subunit, and Tb3120 an Orc2-like factor. Allied to the new RNAi and gel filtration analyses (see below), greater weight can be placed on the hypothesis that a putative ORC-like complex in *T. brucei* is likely to exist, though whether all canonical six subunits are present remains unclear. The analysis of 36 genomes from organisms across the various eukaryotic supergroups revealed that the typical six-subunit ORC complex is not found in all eukaryotic groups (Aves *et al.*, 2012). Indeed, none of the genomes from organisms of the Excavata supergroup (to which *T. brucei* belongs) appeared to possess the Orc3 and Orc6 subunits of ORC, as well as Cdc6 and Cdt1 (Aves *et al.*, 2012), perhaps explaining why various studies have failed to identify these factors over the years (El-Sayed *et al.*, 2005a; Tiengwe *et al.*, 2012b; Tiengwe *et al.*, 2013; Dang and Li, 2011). It is, nevertheless, possible that other proteins, highly diverged in terms of protein sequence, but functionally conserved, are performing the role of these missing factors. For instance, in archaea, organisms from the Methanococcales group do not possess Orc1/Cdc6 homologs, but a sequence-unrelated factor performs the same role in DNA replication in these organisms (Raymann *et al.*, 2014). Whether this is the case of Tb1120, and potentially Tb2240 (identified in section 3.6.2), as well as of other presently unidentified factors, it is not known.

3.7.2 TbORC1/CDC6, TbORC4 and Tb3120 are involved in nuclear DNA replication

Only TbORC1/CDC6 has been directly shown to be involved in DNA replication (Benmerzouga *et al.*, 2013), as downregulation of its expression by RNAi led to a decrease in the number of replicating cells. Although silencing of TbORC4, Tb7980 or Tb3120 expression resulted in similar growth and cell cycle defects, these were slow to arise and had not been confirmed to result from impairment in nuclear DNA replication (Tiengwe *et al.*, 2012b). The effects of TbORC1B or Tb1120 downregulation have never been investigated. The results presented here confirm the role of TbORC1/CDC6, TbORC4, Tb3120 and TbORC1B in *T. brucei* PCF DNA replication, while the involvement of Tb7980 and Tb1120 in this biological process remains unclear. Due to its peculiarity, TbORC1B is discussed separately in section 3.7.4.

Induction of RNAi silencing of either TbORC1/CDC6 or TbORC4 resulted in a strikingly similar phenotype: growth defects arising after ~2-3 cell doublings, accompanied by severe cell cycle anomalies comprising mainly the substantial increase in the number of enucleated cells that retain a kinetoplast (0N1K, zoids) (Robinson *et al.*, 1995). Though similar to results reported in previous studies (Tiengwe *et al.*, 2012b; Godoy *et al.*, 2009; Benmerzouga *et al.*, 2013), the time frame for the effects to become noticeable, as well as the severity of the phenotypes, were more pronounced in this study. In both TbORC1/CDC6 and TbORC4 RNAi, zoid cells accumulated to around 60-70% of the population at 48-96 h post-induction, when compared to only 20-25% of zoids observed previously at the 96 h time point, and a maximum of 30% at 144 h (Tiengwe *et al.*, 2012b; Godoy *et al.*, 2009; Benmerzouga *et al.*, 2013). The amount of 1N1K cells decreased almost proportionally with the increase in zoids, and virtually no 2N2K cells (post-mitotic and undergoing, or about to start, cytokinesis) were detected, while 1N2K cell numbers were also reduced. It is not clear why the RNAi cells used in this study responded so dramatically (Tiengwe *et al.*, 2012b; Godoy *et al.*, 2009; Benmerzouga *et al.*, 2013), especially as mRNA levels assessed by RT-qPCR suggest that knockdown efficiency was not markedly greater than those described by (Tiengwe *et al.*, 2012b) (here ~60% for TbORC1/CDC6 RNAi and only 20% in TbORC4, compared with 90% and 75%, respectively, albeit assessed ~72 h later). The appearance of zoids has been a

hallmark phenotype in studies of DNA replication and mitosis in PCF *T. brucei* cells, as in this life cycle stage of the parasite cytokinesis appears not to be dependent on completion of mitosis or DNA synthesis taking place, since the mitosis to cytokinesis checkpoint is lacking (Ploubidou *et al.*, 1999; Hammarton, 2007). For instance, treatment of PCF cells with Aphidicolin, which inhibits nuclear DNA synthesis (Ploubidou *et al.*, 1999), results in a rise in the number of zoids, which are thought to generate as 1N2K cells, that should be theoretically in G2 phase, but have not replicated their nuclear DNA, divide into a zoid and a 1N1K cell (Figure 3.49, bottom panel). Because the failure in replicating their nuclear DNA results in an inhibition of mitosis, but not cytokinesis, the 2N2K cell population disappears (Ploubidou *et al.*, 1999). The increase in the number of zoids is almost equivalent to the reduction in the numbers of 1N1K cells, as from one 1N2K cell only one 1N1K cell is returned to the cell population rather than the usual two that would result from a normal cell division (Ploubidou *et al.*, 1999). In the case of inhibiting mitosis but not DNA replication, for e.g. by incubating the cells with an anti-microtubule agent such as rhizoxin (Ploubidou *et al.*, 1999) or by targeting by RNAi TbCYC6, a mitotic cyclin (Hammarton *et al.*, 2003), G2 phase 1N2K cells do not undergo mitosis, generating a zoid and a 1N1K cells with a fully replicated nucleus, which would be detected as G2 phase cells by flow cytometry. In the present study, the cell cycle phenotypes arose following impaired nuclear DNA replication, since for both TbORC1/CDC6 and TbORC4 assessment of EdU (thymidine analogue) incorporation showed that reduced DNA replication was already detected 24 h post RNAi induction (when EdU incorporation decreased by ~50%), which preceded the emergence of the growth and cell cycle defects. It should also be noted that the very large number of zoids that emerged after RNAi of either factor argue that replication of the kDNA is largely unaffected, suggesting that the roles of these factors are strictly nuclear (consistent with their localisation) and that there may be little communication between the nuclear and kinetoplast replication processes, at least in PCF cells.

Silencing of Tb3120 resulted in similar phenotypes, though not as fast or pronounced as observed for TbORC1/CDC6 and TbORC4: a growth defect was only noticeable at day four post-induction, in parallel with an increase in the number of zoids to 30% of the cell population. Thereafter, the phenotype

became more severe than described in (Tiengwe *et al.*, 2012b), with near growth arrest by 7 days post-induction, when the number of zoids reached >40% of the cell population (compared with to 20% in (Tiengwe *et al.*, 2012b)). Like for TbORC1/CDC6 and TbORC4, replication impairment preceded the cell cycle and growth defects, with ~50% reduced EdU uptake taking place 24 h earlier. Together, these results confirm that Tb3120 is also involved in DNA replication. Why the phenotypes of this loss are delayed relative to TbORC1/CDC6 and TbORC4 is unclear, but two possible explanations might be considered. The first is that Tb3120 might not play such a central role in the process of nuclear DNA replication and, therefore, the delayed phenotype, although similar in outcome, reflects the fact that loss of Tb3120's role is not detected for several cell divisions after RNAi is firstly induced. What such a secondary role might be is unclear, however, and that would be unlikely if Tb3120 formed part of the *T. brucei* ORC-like complex. The second, perhaps more likely explanation, is that Tb3120 protein is more stable than TbORC1/CDC6 and TbORC4, and consequently, requires longer RNAi induction to lower the protein levels in the cell. This may be consistent with the fact that all three proteins are constitutively expressed (see below), and therefore coordinated turnover rates are not needed.

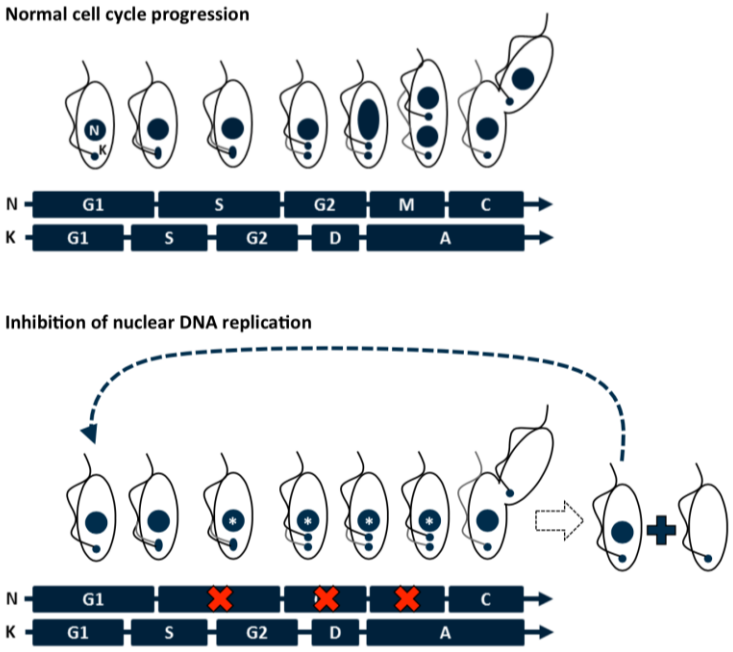


Figure 3.49. Schematic representation of cell cycle progression in PCF cells.
(Top panel) normal cell cycle progression in PCF cells. (N) Nucleus, and nuclear phases; (K) Kinetoplast, and kinetoplast phases. (G1) G1 phase; (S) S phase; (G2) G2 phase; (M) Mitosis; (C) Cytokinesis; (D) Division; (A) Anaphase. Diagram based on the descriptions and diagrams from (McKean, 2003; Field and Carrington, 2009). (Lower panel) Progression of a 1N1K cell in which nuclear DNA replication has been impaired. Red crosses mark the nuclear stages that were

blocked. White asterisks demark nucleuses that were supposed to be replicating or replicated, but are not. The result is a 1N1K cell and a 0N1K zoid. It is possible that the resultant 1N1K cell goes through the process again, as indicated by the dashed arrow. Diagram based on the description from (Ploubidou *et al.*, 1999)

3.7.3 A closer look at TbORC1/CDC6, TbORC4, Tb7980, Tb3120 and Tb1120: nuclear, no matter which stage of the cell cycle

Analysis of protein localisation at the subcellular level is critical to understanding the biological process a protein acts in, its relationship with interacting partners and to ask questions about its regulation (Brand *et al.*, 2007). Moreover, it is an important piece of evidence in support of information obtained by biochemical and molecular biology methodologies on protein-protein interactions, such as immunoprecipitations and yeast-two-hybrid assays (Brand *et al.*, 2007). For instance, recently, a group of *T. brucei* proteins have been identified as putative kinetochore factors through the comparison of their subcellular localisation dynamics during the cell cycle, despite the fact that none appear to be related to previously characterised eukaryotic kinetochore complex members (Akiyoshi and Gull, 2014). Although TbORC1B, TbORC4, Tb7980 and Tb3120 have been shown to interact with TbORC1/CDC6, their localisation in the cell has never been investigated, or indeed their potential co-localisation with TbORC1/CDC6, which has been examined previously (Godoy *et al.*, 2009). Below, the subcellular localisation of all putative *T. bruce*i Orc-like factors is described and compared with other eukaryotic proteins involved in DNA replication. Again, results obtained for TbORC1B will be discussed in a dedicated section (3.7.4).

TbORC1/CDC6 localises to the nucleus of PCF cells throughout the cell cycle, during which is remains bound to chromatin (Godoy *et al.*, 2009), similar to what has been reported for ORC in both budding (Liang and Stillman, 1997) and fission yeasts (Lygerou and Nurse, 1999), where the ORC complex remains intact and chromatin-bound at all times. This contrasts with observations in other multicellular eukaryotes, in which at least one ORC subunit, usually Orc1, has been shown to be cell cycle-regulated by various different mechanisms, like selective degradation (Li and DePamphilis, 2002) and nuclear exclusion (Romanowski *et al.*, 1996), during S phase (reviewed in DePamphilis, 2005). In *T.*

cruzi, TcORC1/CDC6 shows exactly the same cell cycle characteristics as TbORC1/CDC6 in replicating epimastigotes (insect stage of the parasite) (Godoy *et al.*, 2009), although a more recent study has suggested that TcORC1/CDC6 sub-nuclear localisation is not constant and homogeneous, but cell cycle dependent (Calderano *et al.*, 2011b). During most of the cell cycle, TcORC1/CDC6 is dispersed throughout the nucleus, while in the G1 to S phase transition it is re-located to the nuclear periphery (Calderano *et al.*, 2011b), matching previous observations suggesting that in *T. cruzi* the chromosomes become restricted to regions of the nuclear periphery during DNA replication (Elias *et al.*, 2002). Such re-localisation has not been reported in *T. brucei*, and the data here shown did not suggest such a pattern for TbORC1/CDC6^{12myc}, although a clear punctate pattern is seen in all parts of the nucleus (with an apparent exclusion of the nucleolus) throughout the cell cycle. Likewise, the ORC1/CDC6 orthologue in *Leishmania major*, LmORC1/CDC6, was shown to localise to the nucleus throughout the cell cycle, with no intra-nuclear specific re-localisation being reported (Kumar *et al.*, 2008). Nevertheless, the data in *L. major* was obtained using cells expressing LmORC1/CDC6 fused to GFP from an episomal plasmid, meaning that the number of copies of the protein in the analysed parasites might have been variable (Kumar *et al.*, 2008). Further evidence for the difference in replication between *T. brucei* and *T. cruzi* comes from the localisation of the replication-sliding clamp PCNA. Like TcORC1/CDC6, TcPCNA has been shown to localise to the nucleus throughout the cell cycle of epimastigote cells, but to also re-localise to the nuclear periphery in S phase (Calderano *et al.*, 2011b). In *T. brucei* PCF cells, TbPCNA localisation was shown to be cell cycle-dependent, being only detected in the nucleus of S phase cells, where it was seen as many puncta throughout the nucleus (Kaufmann *et al.*, 2012). In *L. donovani* promastigotes (insect stage), LdPCNA is seen in the nucleus throughout the cell cycle, with the levels of the protein peaking in G1 and S phases, although it is not clear whether the protein is re-located to any particular region of the nucleus during DNA replication (Kumar *et al.*, 2009). PCNA orthologues in all three parasites have been shown to co-localise with replicating DNA, potentially at replication 'factories', but no work has asked whether there is any overlap with the punctuate localisation described for TbORC1/CDC6. Why the localisation pattern and dynamics of these replication-

related proteins differ between the kinetoplastids is unclear, including whether this reflects differing mechanisms involved in nuclear DNA replication.

Localisation of TbORC1/CDC6^{12myc} by super resolution imaging resolved the puncta visible at lower resolution, and showed that the protein localises in a large number of discrete points, apparently throughout the nucleus and not in any specific intra-nuclear regions. The number of discrete TbORC1/CDC6^{12myc} points appeared to increase in S and G2 phase cells. What these abundant points correspond with is unknown. It is possible that they might correspond to TbORC1/CDC6 bound to the origins of replication in the genome, which have been estimated at ~42, though it is unclear what number of points were present. Alternatively, the points might correspond with the much larger number of TbORC1/CDC6 binding sites, some of which might co-localise in discrete sub-nuclear domains. Addressing these questions requires further work. For instance, it will be interesting to co-localise TbORC1/CDC6 with specific regions by fluorescence *in situ* hybridisation (FISH), using a probe recognising one of the origins, for instance. An alternative would be to co-localise TbORC1/CDC6 with histone H4 acetylated in lysine 10 (H4K10ac) (Siegel *et al.*, 2009), which is enriched at transcription start sites, and which have been shown to lie close to TbORC1/CDC6-binding sites in the genome (Tiengwe *et al.*, 2012a). As far as we know, this is the first time that an ORC protein has been analysed by super resolution imaging.

Analysis of TbORC4, Tb7980, Tb3120 and Tb1120, each individually endogenously tagged with 12myc revealed the same subcellular localisation pattern as observed for TbORC1/CDC6^{12myc}: all localise to the nucleus as discrete puncta throughout the cell cycle, without any noticeable changes in intra-nuclear localisation related to a specific cell cycle stage. Again, none of these factors seemed to localise to the nucleolus. In general, although some exceptions exist, it appears that the Orc2-6 subunits remain bound to chromatin throughout the cell cycle (DePamphilis, 2005; McNairn *et al.*, 2005), which coincides with the observations made for TbORC4, and support, to some extent, the suggestions that Tb7980, Tb3120 and Tb1120 might also be Orc-like factors. An attempt to co-localise these factors, this time endogenously tagged with 6HA, suggest co-localisation of each factor with TbORC1/CDC6^{12myc}, although better resolution

images, including super resolution, as well as optimisation of the detection of the HA signal, will be needed to confirm these observations and quantify the extent of overlap. Nevertheless, these results may suggest that these factors act in a complex, though it would be valuable to ask whether TbORC4, Tb7980, Tb3120 and Tb1120 are, like TbORC1/CDC6, bound to the chromatin throughout the cell cycle (Godoy *et al.*, 2009). Subcellular fractionation revealed that TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc} and Tb1120^{12myc} all display similar levels of protein in the nucleus and in the cytoplasm, as has been reported for various ORC subunits in other eukaryotes (Lygerou and Nurse, 1999; Tatsumi *et al.*, 2000; Semple *et al.*, 2006). Why the cytoplasmic pool for any of the *T. brucei* factors cannot be detected by microscopy is unclear. However, it has been shown in human cells that ORC assembly is a dynamic process where, for example, Orc1 and Orc6 are transported into the nucleus independently of the other ORC members, which assemble in the cytoplasm as a sub-complex comprising Orc2-5 (Ghosh *et al.*, 2011). The human Orc2-5 sub-complex is then imported to the nucleus where it accumulates, but only binds to chromatin when later associated with Orc1 (Ghosh *et al.*, 2011). One potential way to ask about the relationship between the putative *T. brucei* ORC and that of mammals would then be to ask whether TbORC1/CDC6, TbORC4, Tb7980, Tb3120 and Tb1120 interact in the cytoplasm and in the nucleus and if they form any sub-complexes. This could be done by performing immunoprecipitations using the different cellular fractions, besides potentially also reveal regulation processes involved in nuclear DNA replication in *T. brucei*.

Although TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc} and Tb1120^{12myc} localise to the nucleus throughout the cell cycle without a clear evidence of cell cycle-dependent regulation, measurement of the signal intensity of each of these factors obtained by microscopy revealed that their levels are not static. The levels of TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc} and Tb1120^{12myc}, followed the same trend: starting at a basal level in G1 phase cells, the signal increased progressively as the cells entered and progressed through S phase, finally peaking in G2 phase cells, when the DNA has been completely duplicated, and later decreasing by half, back to G1 phase levels, in each individual nucleus of post-mitotic cells. In budding and fission yeast, where ORC has been shown to bind to the chromatin throughout the cell

cycle (Liang and Stillman, 1997; Lygerou and Nurse, 1999), it has been presumed that new ORCs bind to newly synthesised origins of replication as S phase progresses (DePamphilis, 2005; Arias and Walter, 2007). Nevertheless, re-replication does not take place because regulatory mechanisms, such as phosphorylation of the Orc2 and Orc6 subunits in budding yeast (Nguyen *et al.*, 2001), ensure that the remaining members of the pre-RC complex do not bind to ORC until the late mitosis to G1 phase transition. The common pattern of TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc} and Tb1120^{12myc} levels throughout the cell cycle is further evidence that these factors might act together. Indeed, the potential co-localisation of TbORC1/CDC6^{12myc} and EdU seems to suggest that TbORC1/CDC6 associates with newly synthesised DNA, but more detailed analysis is needed, besides performing the same analysis for TbORC4, Tb7980, Tb3120 and Tb1120.

Altogether, the data here presented support the idea of TbORC1/CDC6, TbORC4, Tb7980, Tb3120 and Tb1120 might be acting in the same biological process, but it will be essential to investigate with more detail the co-localisation of these factor with each other, as well as their cell cycle dynamics, and eventually, their regulation.

3.7.4 The odd case of TbORC1B: is TbORC1B a regulatory factor?

Initially identified as a second Orc1-like protein in *T. brucei* (Dang and Li, 2011), TbORC1B has revealed itself to be the most intriguing TbORC1/CDC6-interacting factor here analysed. Besides its interaction with TbORC1/CDC6 and lack of ATPase activity, although the fact that it appears to have a quite conserved AAA+ ATPase domain (Dang and Li, 2011), nothing was known about TbORC1B prior to this study. Despite this lack of knowledge, it was suggested that TbORC1B most likely acts simply as one component of the *T. brucei* ORC-complex, along with TbORC1/CDC6 (Li, 2012).

The first piece of evidence suggesting a distinct role for TbORC1B relative to TbORC1/CDC6 came from the RNAi assay. The effects of TbORC1B expression silencing by inducible RNAi arose at least twice as quickly as those seen for

TbORC1/CDC6 (and TbORC4), and were more severe. A clear growth defect was seen at 24 h post-RNAi induction, which was accompanied by a severe cell cycle defect, with around 60% of the cell population being zoids. Despite this, the level of knockdown of *TbORC1B* mRNA was equivalent to that of *TbORC1/CDC6* in the corresponding cell line, albeit as measured by RT-qPCR. Because of the rapidity with which cell cycle defects arose, the effects of TbORC1B RNAi on replication were measured at early time points after tetracycline addition to the media, revealing that as early as 6 h post RNAi induction (estimated as half of the cell cycle in this cell line) there was already a 60% loss of EdU incorporation. Importantly, this impairment in replication again preceded the cell cycle defects, since a small increase in 1N2K and 2N2K cells was observed at 6 h, which was followed at 12 h by a clear increase in the number of 1N2K cells, a decrease in 1N1K and 2N2K cells, and the subsequent appearance of zoids. Although an increase in 1N2K cells could suggest that these cells stall in the G2 phase, flow cytometry showed that there is actually a decrease in G2 phase cells at these time points, suggesting that nuclear DNA replication is severely and rapidly affected by loss of TbORC1B, although the kinetoplast is still replicated and segregated. A hypothesis for the rapidity of the emergence of these defects is suggested below, after the results on TbORC1B localisation and dynamics are discussed.

TbORC1B subcellular localisation was analysed by endogenously tagging it with 12myc, similar to TbORC1/CDC6 and the other factors. Unlike 12myc tagged TbORC1/CDC6, TbORC4, Tb7980, Tb3120 and Tb1120, TbORC1B^{12myc} was detectable in the nucleus of only ~33% of the cells in the cell population. Cell cycle characterisation showed that TbORC1B^{12myc} was seen in the nucleus of only cells that are in late G1 to late S or early G2 phase, since no TbORC1B^{12myc} signal was detected in the majority of 1N1K cells, in any 2N2K cells, nor in 1N2K cells that had enlarged nuclei and separated kinetoplasts, suggesting they were approaching or had begun mitosis. The majority of the cells containing TbORC1B^{12myc} nuclear signal were 1N1eK cells, which are assumed to be in S phase. These predictions were supported by measurements of the intensity of the TbORC1B^{12myc} nuclear signal, which showed that TbORC1B^{12myc} did not follow the pattern of TbORC1/CDC6^{12myc} or any of the other interacting factors, since the signal peaked in 1N1eK cells, was markedly reduced in 1N2K cells and

completely absent from 2N2K cells. Finally, TbORC1B^{12myc} cells were FACS sorted into early-mid G1, early S, late S and G2 phases and the protein levels of TbORC1B^{12myc} were assess by western blot. TbORC1B^{12myc} was found in both S phase samples, as well as in G2, although at lower levels, but was not detected in G1, showing that the microscopic detection does not simply reflect cell cycle regulated transport into the nucleus, but cell cycle regulation of protein expression instead. This clearly shows that TbORC1B^{12myc} expression is cell cycle regulated, being expressed mainly during S phase cells. Evidence from a cell cycle transcriptome study (Archer *et al.*, 2011) suggests that TbORC1B mRNA levels peak in late G1, and then reduce somewhat in S phase and further still in early G1 and G2 phases[m]. In the work described here, insertion of the 12myc-tagging construct into the endogenous locus of *TbORC1B* alters the gene's 3' UTR region, potentially interfering with such controls. If so, the fact that protein levels and localisation of TbORC1B^{12myc} clearly change during the cell cycle argue for further expression controls based on the protein, though the basis for this is unknown. Taken together, all these data strongly suggest that TbORC1B expression is cell cycle regulated, implicating this factor as being central to control of *T. brucei* DNA replication.

Once detected in the nucleus, TbORC1B^{12myc} displayed the same punctate staining that was seen for TbORC1/CDC6^{12myc} and the other non-cell cycle regulated proteins examined. Indeed, super-resolution microscopy revealed the same: abundant puncta throughout the nucleus (except the nucleolus) as seen for TbORC1/CDC6^{12myc}. This may argue that TbORC1B is recruited to the putative ORC once in the nucleus, but the co-localisation studies attempted to date were somewhat unclear. Indeed the inability to detect interaction between TbORC1B and TbORC1/CDC6 by co-IP, despite this being reported previously, is perhaps surprising in this regard. However, the limited expression of TbORC1B in the cell cycle may make detection of such interaction problematic. One other problem in this study is that it has not proved possible to validate that the 12myc-tagged version of TbORC1B is essential in PCF cells. Attempts to delete the remaining untagged allele in the TbORC1B 12myc cell line were made, but unfortunately, all of these failed to recover any transformants. However, it also proved

[m] Data available on TriTrypDB, gene page for Tb927.9.2030, the gene ID for TbORC1B, and in the supplementary information of Archer *et al.,* 2011.

impossible to remove one allele of *TbORC1B* in the wild type cell line 927 wt (including in wild type BSF cells, Chapter 4), where both alleles are unmodified, perhaps suggesting that a single functional allele is insufficient for cell survival.

If TbORC1B interacts with ORC, the data presented here show that it is not a static component of a putative ORC-like complex, and instead any interaction with such a complex is most likely transitory and occurs at most through late G1 to early G2 phase of the cell cycle. This would suggest that TbORC1B acts as a regulatory factor. If so, does this explain the speed and nature of the RNAi effects? A hypothesis for the course of the events following TbORC1B RNAi is as follows. In a normal asynchronous culture, cells progress through the cell cycle normally, and TbORC1B localises to the nucleus of some 1N1K cells, most or all 1N1eK, and some 1N2K cells (Figure 3.50, A, TbORC1B presence depicted in green). At 6 h post-RNAi induction, when RT-qPCR suggests *TbORC1B* mRNA is already reduced by ~40%, no clear cell cycle differences are observed by DAPI staining, although flow cytometry suggests a decrease in the number of G2 phase cells. The ~60% reduction in EdU incorporation 6 h post RNAi induction shows that loss of TbORC1B begins to prevent replication of the DNA. Since TbORC1B is expressed only from late G1 to late S or early G2, and is potentially degraded at other times, one can hypothesise the following: if a cell had TbORC1B at the time tetracycline was added, or had finished DNA replication (Figure 3.50, B, 0 h row) but not completed cytokinesis, then RNAi induction may not have an effect, and the cells would progress into G1 (Figure 3.50, B, 6 h row). On the other hand, if TbORC1B were not present in the cell – in most 1N1K cells that are in G1 phase – RNAi induction would prevent synthesis of TbORC1B as they reach S-phase, and these cells would not be able to start DNA replication. From (Ploubidou *et al.*, 1999), it is known that in PCF cells, blocking DNA replication can inhibit mitosis, but not cytokinesis. Moreover, because the kinetoplast S phase is independent from the nuclear one (reviewed in McKean, 2003; Hammarton, 2007), the cells would still replicate and segregate their kinetoplast, generating a cell that would be morphologically 1N2K (suggestive of being in G2 phase), but with the nuclear content of a cell in G1 phase (Figure 3.50, B, 6 h row, nucleus with white asterisk). Because the majority of cells in an asynchronous population are in G1 phase, in which TbORC1B is not largely expressed, unlike TbORC1/CDC6 and TbORC4, this might explain the more rapid

onset of replication-associated defects in the TbORC1B RNAi cell line, as well as the appearance of the unusual 1N2K cells with a nucleus in G1 phase (Figure 3.50, B, 12 h row). As the time course of RNAi induction progresses, the 1N2K cells with a non-replicated nucleus would undergo cytokinesis and generate a 1N1K cell and a zoid, while the 1N1K cells would generate a 1N2K cell with a non-replicated nucleus (Figure 3.50, B, 12 h row). This scenario can explain the transient increase in the number of 1N2K cells (but lowered numbers of G2 phase cells in the flow cytometry analysis), as well as the rapid loss of 2N2K cells at early time points, and the slight increase in zoid cells at the same time (as suggested by both DAPI counting and flow cytometry analysis). Thereafter, the number of zoids would continue to increase and 1N2K cells would be lost, as eventually no cells would be able to replicate their nuclear DNA (Figure 3.50, B, 24 h row). Indeed, broadly the same effects are seen following RNAi of the other putative Orc-like factors, but take longer to emerge and are less easy to capture, as samples were not taken as frequently. It is not known how long a zoid cell survives in culture, or whether it can further divide to generate two zoids. Nevertheless, some of the cells categorised as “others” had no nucleus but two or more kinetoplasts, suggesting this may be possible. Of course, to validate these suggestions it will be essential to tag TbORC1B or raise antisera against it in order to assess protein levels at the different time points after RNAi.

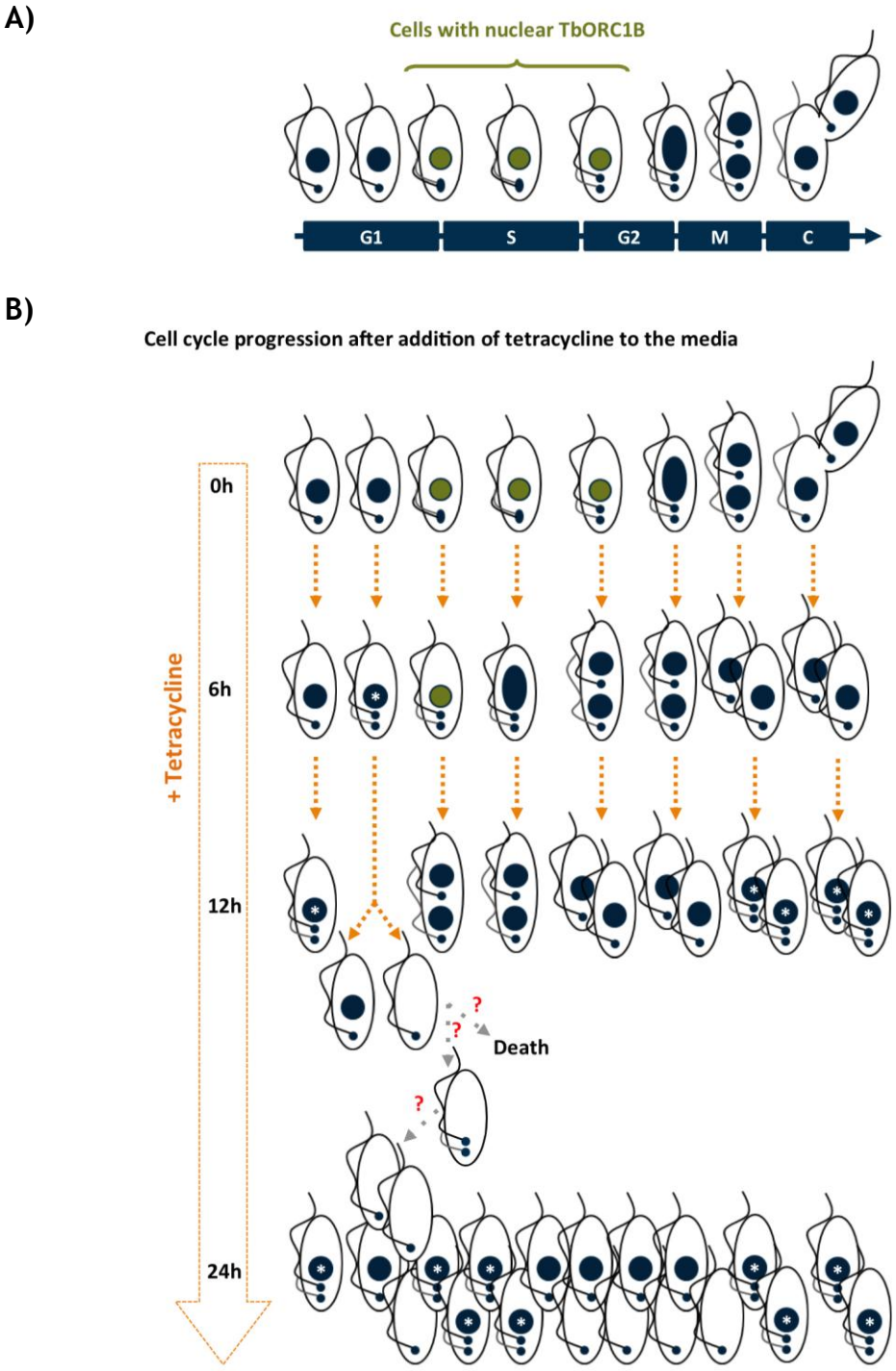


Figure 3.50. Hypothetical model of TbORC1B RNAi.
A) Schematic representation of the different cell cycle stages, according to their nucleus to kinetoplast ratio. Same diagram shown in Figure 3.41. B) Schematic representation of what might occur upon TbORC1B expression silencing by RNAi induction. Detailed description in the main text. Green nucleus – cells with nuclear TbORC1B signal; white asterisks – cells with a non-replicated nucleus.

If TbORC1B acts as a regulatory factor in nuclear DNA replication in *T. brucei*, does its expression and localisation compare with any known replication regulators in other eukaryotes? One possibility is that TbORC1B acts as Cdc6, in particular because TbORC1B BLAST hits were mainly Cdc6 proteins, but also

because Cdc6 expression and subcellular localisation is cell cycle regulated in other organisms (reviewed in Costa *et al.*, 2013; Siddiqui *et al.*, 2013). The context for this regulation is that, in characterised eukaryotes, the pre-RC complex (ORC-MCM2-7 with Cdc6 and Cdt1) is loaded onto origins from late mitosis through G1 phase. Upon entering S phase, the origins are activated, but must only fire once per cycle, and re-replication is prevented by some pre-RC disassembly. For instance, in mammalian cells, Cdc6 localises to the nucleus in G1 phase but at least some is transferred to the cytoplasm as the cell enters S phase (Delmolino *et al.*, 2001; Petersen *et al.*, 1999). Cdc6 in budding yeast and in fission yeast doesn't display re-localisation, but instability (Kearsey and Cotterill, 2003; Piatti *et al.*, 1995): transcription of both genes is limited to mitosis to G1 phase and the proteins are degraded as the cells enter S phase (Luo *et al.*, 2003). Cdc6 (named Cdc18 in *S. pombe*) removal appears to be regulated via ubiquitylation and consequent degradation by the proteasome (Arias and Walter, 2007; Borlado and Mendez, 2008; Diffley, 2010; Li and Jin, 2010). Thus, despite the potential sequence homology, TbORC1B's behaviour is quite unlike Cdc6, primarily because TbORC1B is absent in G1 phase cells and localises to the nucleus from only late G1, if not at the onset of S phase. In addition, TbORC1B may not have ATPase activity, present in Cdc6, perhaps due to the lack of the signature arginine finger (Dang and Li, 2011). Nevertheless, it is possible that TbORC1B is also regulated by ubiquitylation upon entering G2 phase; attempts were made in the present work to address this (data not shown), but were hampered by problems regarding the anti-ubiquitin antiserum used. Nonetheless, it will be of value to test this again to get insight into TbORC1B potential regulation.

Regulation of the activity of ORC subunits has also been reported in eukaryotes, and in particular the Orc1 subunit, to which TbORC1B also resembles. In general, it appears that the Orc2-6 subunits remain bound to chromatin throughout the cell cycle (DePamphilis, 2005; Bell *et al.*, 2013; McNairn *et al.*, 2005). One exception has been described in *Xenopus laevis*, where the ORC complex interaction with chromatin weakens during S phase. In mammals Orc1 function has been shown to be cell cycle-regulated by different mechanisms, including selective degradation (Li and DePamphilis, 2002) and nuclear exclusion (Romanowski *et al.*, 1996). However, these alterations, which occur during S

phase (reviewed in DePamphilis, 2005; Bell *et al.*, 2013), differ dramatically from the nuclear accumulation we see for TbORC1B at this cell cycle stage. The closest analogy seen between TbORC1B expression dynamics and ORC in any other eukaryote is with DmOrc1 of *D. melanogaster*. In the fly, DmOrc1 is detectably expressed only from late G1 phase through S until G2, and is degraded at mitosis and during G1 phase by the Anaphase Promoting Complex (APC) (Araki *et al.*, 2003) - expression patterns that are strikingly similar to TbORC1B. *D. melanogaster* DmOrc1 levels are to a large extent controlled by ubiquitin-mediated degradation through signals in the protein's N-terminus. Perhaps surprisingly, however, mutation of *D. melanogaster* DmOrc1 to prevent mitosis to G1 degradation has no effect on cell cycle progression (Park and Asano, 2011), and it has been suggested that the expression control may be needed for lineage-specific replication-mediated gene amplification (Park and Asano, 2012). It is unclear if *T. brucei* possesses any such discrete replication reactions. Furthermore, *D. melanogaster* DmOrc1 is highly conserved relative to 'model' eukaryotic Orc1 subunits, including retention of an N-terminal BAH domain, and interacts with a conventional six subunit ORC that bears substantial structural homology with that of yeast (Bleichert *et al.*, 2013; Clarey *et al.*, 2008; Bleichert *et al.*, 2015). TbORC1B is more diverged in sequence from Orc1 than TbORC1/CDC6 and lacks any equivalent N-terminal sequence to that used in *D. melanogaster* DmOrc1 degradation. Thus, the analogy with *D. melanogaster*, while intriguing, is limited, and it remains likely that TbORC1B S phase restriction is central to the control of replication in *T. brucei*.

Another candidate for functional analogy with TbORC1B is Cdt1, despite the fact that the proteins share no sequence homology. Nevertheless, this association is possible because TbORC1B has been shown to interact with the MCM2-7 helicase, at least via MCM3, which has also been reported to interact with TbORC1/CDC6 (Dang and Li, 2011). In other eukaryotes, Cdt1 interacts with MCM2-7 (reviewed in Bell and Kaguni, 2013) and mediates the interaction of the helicase with ORC, apparently through Orc6 (Takara and Bell, 2011; Bell and Kaguni, 2013). If Cdt1 is absent in *T. brucei*, then it is possible that TbORC1B acts to mediate loading of MCM2-7, possibly through MCM3, onto TbORC1/CDC6, and potentially the ORC-like complex, bound to the chromatin. Again, however, expression dynamics of TbORC1B appear incompatible with what has been described in other eukaryotes

(reviewed in Costa *et al.*, 2013; Siddiqui *et al.*, 2013), where Cdt1 acts in loading the MCM2-7 helicase to origins bound by ORC from late mitosis through G1 phase, and up to the onset of S phase. Indeed, like Cdc6, Cdt1 function is repressed by various mechanisms during S phase in many eukaryotes: e.g. by nuclear exclusion in *S. cerevisiae* (Tanaka and Diffley, 2002), by degradation in *S. pombe* (Nishitani *et al.*, 2000), and activity suppression (through binding Geminin) in mammalian cells (Wohlschlegel *et al.*, 2000; Nishitani *et al.*, 2001) (reviewed in Siddiqui *et al.*, 2013).

TbORC1B nuclear localisation dynamics are, somehow similar to one protein in *T. brucei*: TbPCNA (Kaufmann *et al.*, 2012), the sliding clamp involved in DNA polymerase association with the DNA, which is only detectable in the nucleus of PCF cells from late G1 phase to late S phase. Indeed, TbPCNA localisation in the nucleus was analysed by super-resolution microscopy and showed similar large numbers of puncta (Kaufmann *et al.*, 2012) as seen for TbORC1B. This striking overlap between expression profiles fits with the hypothesis that TbORC1B acts as a positive regulator of DNA replication, and it will be interesting to test the co-localisation between these two factors. TbPCNA localisation has also been partly overlapped with newly replicated DNA, assessed by EdU labelling (Kaufmann *et al.*, 2012). Whether the TbORC1B puncta overlap with replicated DNA was rather less clear, and requires more investigation.

It will be of interest to explore TbORC1B further, and to elucidate its role in replication in detail, including in *T. cruzi* and *L. major*, where replication organisation appears to differ in the nucleus. The complete absence of TbORC1B in G1 phase cells may suggest that it is important for the start of DNA replication and therefore, only acts at the onset of S phase. If TbORC1B acts to activate the initiation of DNA replication, it will be important to ask whether it binds to the chromatin and, if so, whether this is cell cycle dependent, and if it only binds to the origins of replication that are activated, or whether it interacts with TbORC1/CDC6 at origin-active and inactive sites in the genome (Tiengwe *et al.*, 2012b).

3.7.5 Does *T. brucei* have an ORC complex?

In the course of this study, it was confirmed that TbORC1/CDC6 interacts with TbORC4, Tb7980 and Tb3120, as well as with Tb1120, and clear functions for TbORC1B, TbORC4 and Tb3120 in replication were shown. Of course it remains possible that more interacting partners exist in the putative ORC-like complex, and one new candidate is Tb2240 (Tb927.10.2240), though no validation of this has been attempted. It remains unknown, however, if TbORC1/CDC6, TbORC4, Tb7980, Tb3120 and Tb1120 interact together in a complex or not. However, gel filtration of extracts from PCF cells expressing TbORC1/CDC6^{12myc} provides the first evidence that at least this Orc-like protein interacts relatively stably with a number of other factors, perhaps in a complex. From this experiment, which should be repeated, the size of the putative complex (most likely ~900-1100 kDa) appears to be too large to be composed of only TbORC1/CDC6 (66.34 kDa with the 12myc tag), TbORC4 (77.29 kDa), Tb7980 (48.8 kDa), Tb3120 (112.9 kDa) and Tb1120 (80.7 kDa) that, with or without TbORC1B (65.49 kDa), sums up to a complex of 386 kDa or 415 kDa, respectively, at least if each interacts in a 1:1 ratio. This might suggest that TbORC1/CDC6 interacts more stably with further, unidentified factors than those recovered above by IP or, more likely, these Orc-like factors form a highly diverged ORC that interacts with further proteins, which are still to be identified. Though highly speculative, the large size of the putative TbORC1/CDC6^{12myc}-containing complex is big enough to represent the putative ORC-like complex bound to the *T. brucei* TbMCM2-7 helicase, which would represent the pre-RC, as well as potentially the TbGINS complex, TbCDC45 and TbMCM10, and thus the pre-IC complex. Indeed, the most abundant sizes of the putative complex (~900-1100 kDa) are big enough to harbour the five Orc-like factors here described as well as the TbMCM2-7 helicase, thus the pre-RC; it makes sense that this is possibly the most abundant complex, since the cells used in this experiment were collected from an asynchronous culture, where ~80% of the cells are 1N1K, assumed to be in G1 phase, when the pre-RC is assembled in all other studied eukaryotes (reviewed in Costa *et al.*, 2013). Further experiments will be needed to test these predictions, but the ability to detect such a complex also provides an experimental route to begin dissecting the regulation of *T. brucei* DNA replication initiation.

4 Analysis of initiation of DNA replication in *T. brucei* bloodstream form cells

4.1 Introduction

Trypanosoma brucei has a complex life cycle, where the parasite has to undergo drastic developmental changes in order to infect two distinct hosts, the mammal and the tsetse fly. Two replicative life cycle stages are routinely cultured: the long slender bloodstream form (BSF), which causes pathology in the mammal, and the procyclic form (PCF) that replicates in the fly midgut. These two developmental stages were adapted to culture and have been shown to differ in many aspects of their cellular biology, such as morphology (reviewed in Field and Carrington, 2009; Fenn and Matthews, 2007; Ooi and Bastin, 2013), proteins expressed at the cell surface (reviewed in Dyer *et al.*, 2013; McCulloch *et al.*, 2014), metabolic pathways (reviewed in Creek *et al.*, 2012), cell division checkpoints (reviewed in Hammarton, 2007; McKean, 2003; Li, 2012), nuclear architecture and chromatin structure (reviewed in Daniels *et al.*, 2010; Schlimme *et al.*, 1993; Rout and Field, 2001; Figueiredo *et al.*, 2009; Maree and Patterton, 2014; Siegel *et al.*, 2009), and gene expression (Siegel *et al.*, 2010; Jensen *et al.*, 2009; Jensen *et al.*, 2014; Brems *et al.*, 2005; Urbaniak *et al.*, 2012; Urbaniak *et al.*, 2013).

To date, most of the work investigating nuclear DNA replication in *T. brucei* has been performed using PCF cells (Tiengwe *et al.*, 2012b; Tiengwe *et al.*, 2012a; Benmerzouga *et al.*, 2013; Godoy *et al.*, 2009). Studies in BSF cells have been restricted to analysis of the effects that result from RNAi depletion of TbORC1/CDC6, TbORC4 or Tb7980 (Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013). Perhaps reflecting differences in cell cycle checkpoint between PCF and BSF cells (Hammarton, 2007; McKean, 2003; Li, 2012), the RNAi phenotypes described in the two stages were quite different (see below, section 4.2). Presently, and reflecting the results obtained in PCF cells, only TbORC1/CDC6 was shown to be involved in DNA replication in BSF cells (Benmerzouga *et al.*, 2013). In addition, TbORC1/CDC6 depletion by RNAi resulted in a reduced expression silencing of the variant surface glycoproteins (VSGs) in BSF cells (Benmerzouga *et al.*, 2013), an effect also reported in PCF cells (Tiengwe *et al.*, 2012a). Intriguingly, the results of TbORC1/CDC6 downregulation in BSF cells appeared to be different between the two studies in which it was analysed (Benmerzouga *et al.*, 2013; Tiengwe *et al.*, 2012a) (discussed more fully below, section 4.2). To elucidate the discrepancies between these two studies, as well

as to investigate the differences between RNAi phenotypes in BSF and PCF cells, and to assess whether TbORC4, Tb7980, Tb3120 and TbORC1B are also involved in DNA replication in BSF cells, RNAi cell lines targeting individually each of these factors were created and analysed in detail for growth, cell cycle, and DNA replication defects.

To date, none of the putative ORC factors have been localised in BSF cells. Considering that the two life cycle stages differ in so many aspects, it is relevant to investigate these factors' subcellular localisation and cell cycle dynamics in BSF cells. By applying the strategies described in Chapter 3 for PCF cells, these parameters were also investigated in BSF cells.

Origins of replication have also only been mapped in PCF cells (Tiengwe *et al.*, 2012a). Because PCF and BSF cells are distinct, developmentally differentiated cell types, and in other eukaryotes it has been shown that different cell types of the same organism activate different origins of replication, or use a different replication (origin activation) timing programme (Ryba *et al.*, 2010), replication origins were mapped in BSF cells (Tiengwe *et al.*, 2012a). Because of the complexity of origin function, this area is introduced more fully in the relevant results section (section 4.5).

4.2 TbORC1/CDC6, TbORC4 and TbORC1B are involved in DNA replication

Since *T. brucei* was confirmed to possess a functional RNA interference (RNAi) machinery, RNAi has become the method of choice to examine essential gene function in this parasite, as has been already explored in Chapter 3. In order to understand the role of TbORC1/CDC6 and interacting partners in *T. brucei* biology, and to complement the data from studies in PCF cells, inducible RNAi cell lines targeting individually TbORC1/CDC6, TbORC4 and Tb7980, have been generated in BSF cells (Benmerzouga *et al.*, 2013; Tiengwe *et al.*, 2012b). Interestingly, the phenotypes reported in BSF cells after RNAi of these factors differed from the consistently observed effect seen in PCF cells (Tiengwe *et al.*, 2012b; Godoy *et al.*, 2009; Benmerzouga *et al.*, 2013), in which the formation and accumulation of enucleated cells (0N1K or zoids) predominates (Chapter 3). Moreover, the phenotype described following TbORC1/CDC6 downregulation in BSF cells was different in the two studies published (Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013). In one study, induction of RNAi individually targeting TbORC1/CDC6, TbORC4 or Tb7980, resulted in strikingly similar, rapid phenotypes: growth defects were detectable as early as 18 h post-induction, with a parallel increase in the number of multi-nuclei and multi-kinetoplast cells (Tiengwe *et al.*, 2012b). Surprisingly, flow cytometry analysis of these cells suggested that they had replicated their DNA, with a population containing double the DNA content (8N) present in G2/M cells (4N) appearing, perhaps suggesting re-replication of the nuclear DNA (Tiengwe *et al.*, 2012b). Although it was suggested that this might result from the inhibition of non-replication roles of TbORC1/CDC6, TbORC4 or Tb7980, what such roles might be is unclear. In another study, RNAi against TbORC1/CDC6 in BSF cells also resulted in a growth defect, although at a later time point (48 h) post-induction (Benmerzouga *et al.*, 2013). Here, growth perturbation was accompanied by the appearance of zoid cells (although not to the same levels as seen following RNAi in PCF cells), and the accumulation of multi-nuclei and multi-kinetoplast cells (Benmerzouga *et al.*, 2013). In contrast with (Tiengwe *et al.*, 2012b), an accumulation of cells in G2/M phase (4N) was observed by flow cytometry analysis, while cells with higher DNA content were not detected (Benmerzouga *et al.*, 2013). Still, this would suggest that the cells progressed though S phase, but arrested in G2/M

phase. Nevertheless, BrdU incorporation (thymidine analogue, mentioned and discussed in Chapter 3), suggested that DNA replication was impaired, confirming that TbORC1/CDC6 is involved in DNA replication in BSF cells. The somewhat contradictory phenotypes in the two studies may be accounted for by suggesting that in BSF cells TbORC1/CDC6 performs other roles not directly related to DNA replication, or that such roles assume greater prominence than in PCF cells. Indeed, like Orc1 in other eukaryotes (reviewed in Costa *et al.*, 2013), TbORC1/CDC6 has been shown to be involved in gene silencing: RNAi induction resulted in increased levels of metacyclic *VSG* mRNAs in PCF cells (Tiengwe *et al.*, 2012a) and, in one study, led to the expression of elevated levels of *VSG* mRNAs derived from the silent VSG expression sites in BSF cells (Benmerzouga *et al.*, 2013). However, whether this function accounts for the differences in RNAi phenotypes between BSF and PCF cells is not known, in particular because TbORC1/CDC6 RNAi induction in PCF cells was shown to result in genome-wide changes of the mRNA levels (Tiengwe *et al.*, 2012a), meaning that the *VSG* mRNA changes may not represent a specific effect.

4.2.1 RNAi system used and generation of the RNAi BSF cell lines

To allow individual inducible RNAi silencing of TbORC1/CDC6, TbORC4, Tb7980, Tb3120 and TbORC1B in *T. brucei* BSF cells, the 2T1 cell line (Alsford *et al.*, 2005) was used. This cell line expresses constitutively the tetracycline repressor (TetR), and has been modified to have an incomplete fragment of the hygromycin (*HYG*) resistance gene (the 3' region of the gene's ORF), followed by a complete puromycin (*PUR*) resistance marker gene (the cells are resistant to puromycin but not to hygromycin), integrated into one of the ribosomal RNA (*RRNA*) locus (Alsford *et al.*, 2005) (Figure 4.4). Several plasmids have been designed to allow integration of an RNAi construct, by homologous recombination, into the 2T1 cells genome (e.g. pRPA^{iSL} (Alsford and Horn, 2008) and pGL2084 (Jones *et al.*, 2014; Alsford and Horn, 2008). These plasmids contain the remaining part of *HYG* (5' region of the ORF), plus an overlapping region of the *HYG* fragment present in the parasite's genome, downstream of the *EP procyclin* promoter. In addition, the plasmids contain regions to allow cloning to generate gene-specific stem-loop dsRNA downstream of a RRNA promoter and a tetracycline operator (TetO). Finally, they also possess a fragment of the *RRNA* spacer (Alsford and Horn, 2008; Jones *et al.*, 2014).

Integration of the construct after transformation is thus locus-specific, based on the overlapping *HYG* sequence and *RRNA* spacer, leading to the excision of *PUR* and generation of a complete *HYG* ORF; this means that transformants should be hygromycin-resistant, but puromycin sensitive, cells (Alsford and Horn, 2008; Alsford *et al.*, 2005). The use of these plasmids and 2T1 cells allows locus-specific integration of a tetracycline inducible stem-loop RNAi construct (Alsford and Horn, 2008), with demonstrated higher transfection efficiency and, it is thought, greater clone-to-clone reproducibility (Alsford *et al.*, 2005). Furthermore, dsRNA expression is driven by a RRNA promoter, circumventing the need to have a cell line expressing the bacteriophage T7 RNA polymerase, as other systems (Wirtz *et al.*, 1999), including the ones described in Chapter 3.

Here, the construct pGL2084 (Jones *et al.*, 2014) was used (Figure 4.1). This construct was originally modified from pRPa^{iSL} (Alsford and Horn, 2008) to include two inverted regions containing Gateway® cloning recombination attachment (*att*) sites, *att*P1 and *att*P2, flanking a *ccd*B gene (a bacterial negative selective marker) separated by a 150 bp *lac*Z gene fragment ("stuffer") (Figure 4.1), in order to allow high throughput recombinational cloning of PCR products using the Gateway® system (Hartley *et al.*, 2000). For this, a gene-specific PCR product flanked by *att*B sites is generated using a forward primer containing the *att*B1 site (5' end) and a reverse primer harbouring an *att*B2 site (in the 5' end) (Figure 4.2). The *att*B1-PCR-*att*B2 product is then incubated with the pGL2084 vector in the presence of BP clonase™ enzyme mix (containing the bacteriophage lambda integrase (Int) and the *E. coli* integration host factor (IHF) enzymes), allowing the BP recombination between the *att*B sites-containing substrate (PCR product) and the *att*P-containing substrate (pGL2084 vector). This results in an *att*L-containing construct and excision of the *ccd*B gene flanked by *att*R sites (thus allowing the growth of the bacteria transformed with the recombined plasmid, as expression of CcdB inhibits the growth of most of the *E. coli* strains used in laboratory) (Hartley *et al.*, 2000) (Figure 4.2). The pGL2084 construct was designed so that the *att*B1 site only recombines with the *att*P1 site, and the *att*B2 site only recombines with the *att*P2 site, providing orientation-specific insertion of the PCR product (Hartley *et al.*, 2000). Furthermore, because the pGL2084 vector possesses two inverted regions containing the *att*P1 and *att*P2 sites, it allows the integration, in a single cloning

step, of a single PCR product, into two places in the plasmid, generating a final construct with the gene-specific PCR fragment inserted in a tail-to-tail orientation separated by the “stuffer”, and thus allowing the expression of a dsRNA stem-loop once integrated into the parasite’s genome (Jones *et al.*, 2014) (Figure 4.2). In order to confirm the correct integration of the PCR product, pGL2084 was designed to allow diagnostic enzymatic digestion, using several restriction enzymes (Jones *et al.*, 2014) (Figure 4.3, F). Moreover, the presence of two AscI restriction sites adjacent to the incomplete *HYG* gene and the *RRNA* fragment, allows excision of the construct prior to transfection into 2T1 cells (Jones *et al.*, 2014).

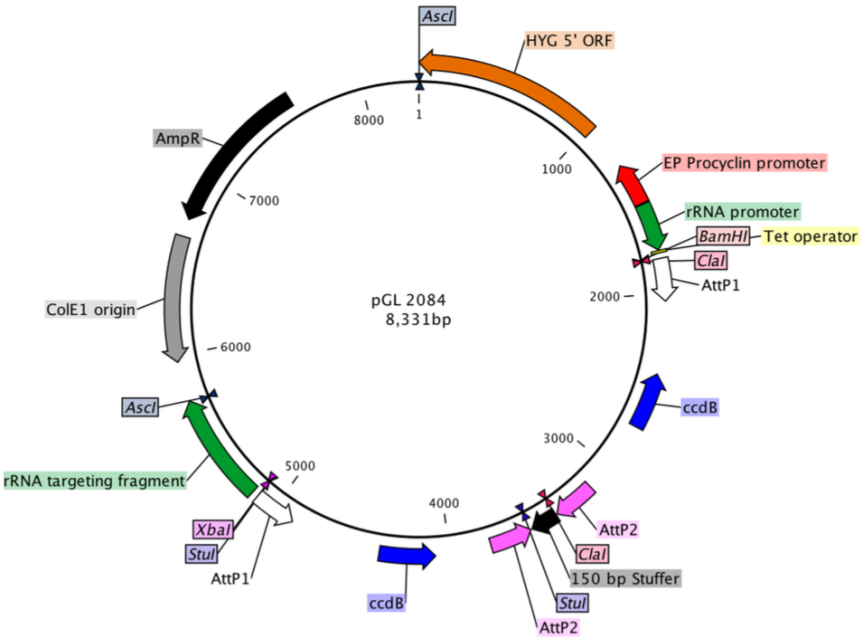


Figure 4.1. Plasmid map of pGL2084.
Detailed map of the pGL2084 vector (Jones *et al.*, 2014) generated via modification of the pRPa^{iSL} construct (Alsford and Horn, 2008), as detailed in the main text. The *att*P1 (white arrows) and *att*P2 (pink arrows) sites have been inserted in two regions of the plasmid, inverted, separated by a 150 bp “stuffer” fragment of the bacterial *lac*Z gene (black arrow). Each of these *att*P sites pairs flank one copy of the *ccd*B gene (blue arrows), which will be excised upon recombination with the PCR product flanked by *att*B sites. Features from the pRPa^{iSL} construct such as the *RRNA* promoter (green arrow) and tetracycline operator (yellow) are placed upstream of the first *att*P1 site, to allow inducible expression of the fragments cloned downstream. Also, the incomplete hygromycin resistance (*HYG*) gene, comprising the 5’ region of the ORF (orange arrow), is localised downstream of the *EP procyclin* promoter (red arrow), and allows construction of the complete gene once the construct is integrated into the parasite genome. The *RRNA* spacer is shown downstream of the second *att*P1 site, which also allows homologous recombination with the parasite genome upon transfection. The bacterial ampicillin resistance gene (*amp*R, large black arrow) and the bacterial origin (ColE1, grey arrow) are also shown. The relevant restriction sites are shown. Some features, such as the mRNA processing regions upstream of the incomplete *HYG*, are not shown for simplicity.

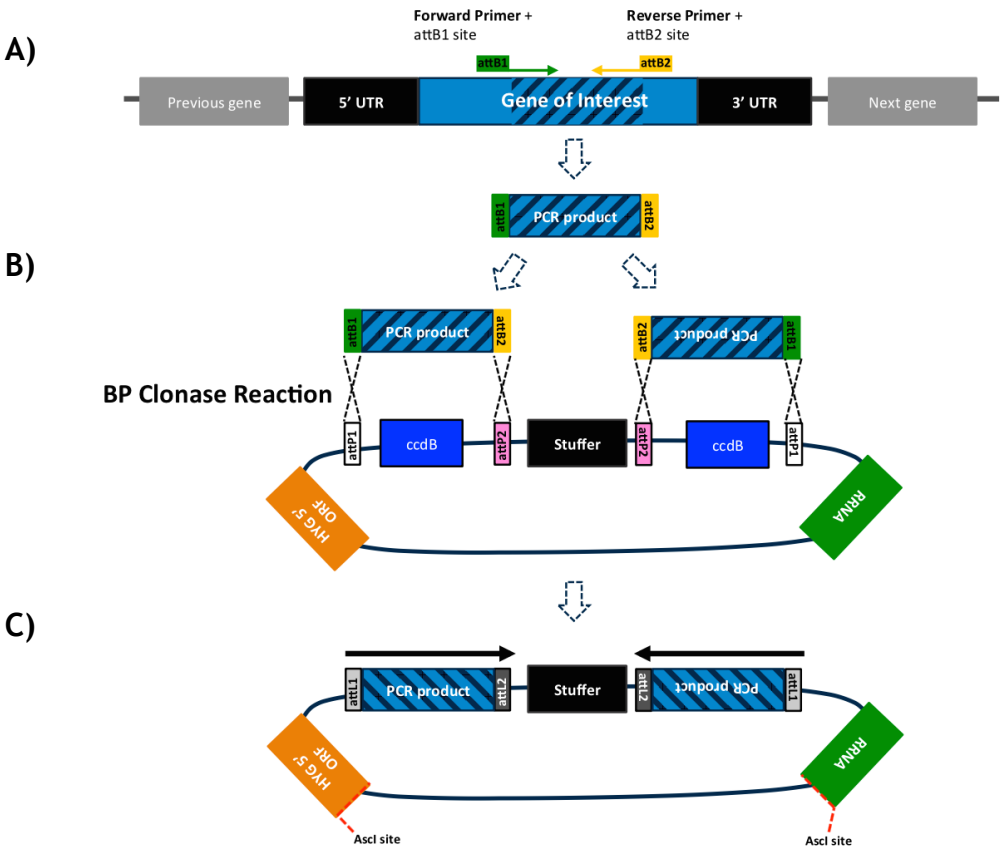
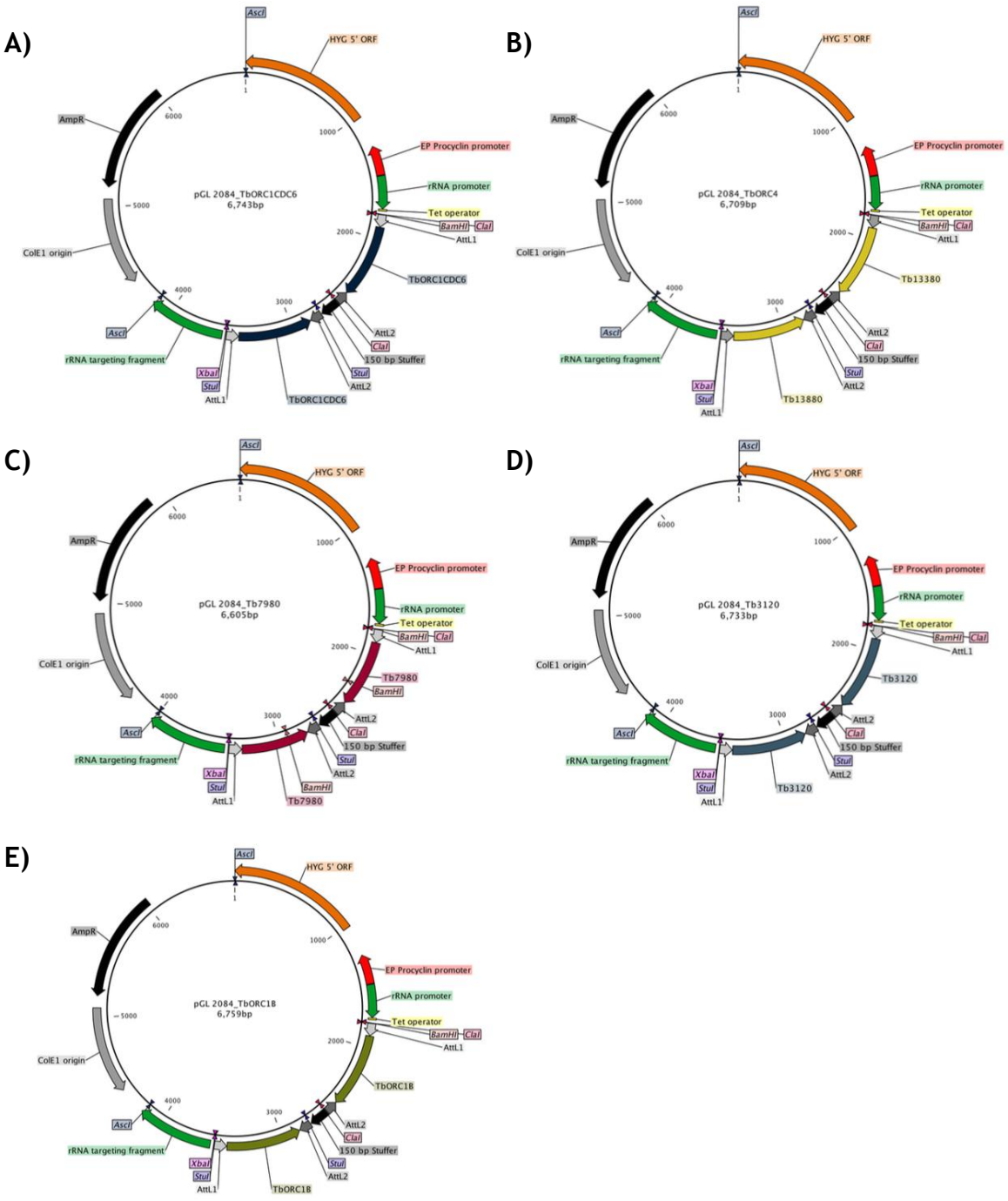


Figure 4.2. Recombinational cloning of PCR products into pGL2084.
Graphic representation of the cloning process used to generate the RNAi stem-loop constructs in the pGL2084 vector (Jones *et al.*, 2014). A) A single PCR product is amplified from parasite gDNA using a pair of primers containing an *att*B1 site in the forward primer (green primer), and an *att*B2 site in the reverse primer (yellow primer). B) The PCR product is then incubated with the pGL2084 vector in the presence of the BP clonase™ mix, allowing recombination between the *att*B1 and *att*P1 (white) sites, as well as the *att*B2 and *att*P2 (pink) sites. Because pGL2084 has two pairs of *att*P sites, in opposite directions, two PCR products are inserted into the vector, one in the sense orientation (→), and the other in the antisense orientation (←), both flanked by newly generated *att*L1 and *att*L2 sites, and separated by the "stuffer" (black) (C). The insertion of the PCR fragments results in the excision of the *ccd*B genes, flanked by *att*R sites (not represented). Not all features of the pGL2084 plasmid are shown. The diagram was designed based on descriptions in (Jones *et al.*, 2014) and (Hartley *et al.*, 2000), and is not represented to scale.

The 2T1 cells/pGL2084 construct strategy for RNAi targeting of TbORC1/CDC6, TbORC4, Tb7980, Tb3120 and TbORC1B was chosen because we considered that using a stem-loop RNAi approach in BSF cells, as was used in PCF cells (Chapter 3), could provide the best chance of comparing results between the life cycle stages. The region of each gene to be PCR-amplified (between 400-600 bp), as well as the best pair of primers to be used, was chosen using RNAit (http://trypanofan.bioc.cam.ac.uk/ software/RNAit.html) (Redmond *et al.*, 2003), as described in Chapter 3, section 3.3.1. To the primer sequences, *att*B sites were added, as described in (Jones *et al.*, 2014). The PCR reactions, using primers C 108 and C 109 (*TbORC1/CDC6*), C 114 and C 115 (*TbORC4*), C 116 and C 117 (*Tb7980*), C 110 and C 111 (*Tb3120*), and C 112 and C 113 (*TbORC1B*), as

well as the cloning and transfection into *E. coli* DH5α bacteria, were performed as detailed in (Jones *et al.*, 2014), and are explained in the materials and methods Chapter 2, section 2.2.1.10. The plasmids were then purified and confirmed by restriction enzyme digestion as described in (Jones *et al.*, 2014), and are shown in Figure 4.3, F: double digestion with BamHI and XbaI results in the excision of the two opposing PCR fragments and *att*L sites; digestion with ClaI excises one of the PCR fragments and respective *att*L sites; digestion with StuI excises the other PCR fragment and *att*L sites; and digestion with AscI excises the entire insert fragment for transfection into the 2T1 cells. All constructs produced the expected restriction digestion patterns, except the Tb7980 construct, where digestion with BamHI and XbaI resulted in an extra band, since the PCR product included a BamHI restriction site (Figure 4.3, F). All constructs were then digested with AscI and transfected into 2T1 cells, as described in the materials and methods Chapter 2, section 2.3.5. Transformants were selected in the presence of 5 $\mu g.ml^{-1}$ of hygromycin, and isolated antibiotic resistant clones were further tested for puromycin sensitivity (in the presence of 0.2 $\mu g.ml^{-1}$ of puromycin, as integration should remove *PUR*) (Jones *et al.*, 2014; Alsford *et al.*, 2005). Clones (Cl) sensitive to puromycin were selected for each gene, and used for further experiments: TbORC1/CDC6 RNAi Cl9a, TbORC4 Cl12a, Tb7980 Cl3a, Tb3120 Cl5a, and TbORC1B Cl2a. In the case of TbORC1/CDC6, TbORC1B, Tb7980 and Tb3120, the RNAi constructs were transfected into 2T1 cells that had been previously endogenously tagged in one allele with 12myc, and confirmed to express the fused version of the respective protein, as described in section 4.3, and using the tagging-constructs described previously in Chapter 3. This allowed the assessment of the RNAi induction efficiency by the detection of the 12myc-tagged protein levels. For reasons that are unclear, it was not possible to generate a 2T1 cell line with TbORC4 endogenously tagged with 12myc.



1. BamHI x XbaI	1740 bp + 5009 bp	1756 bp + 5009 bp	1706 bp + 5009 bp	471 bp + 660 bp + 5009 bp	1734 bp + 5013 bp
2. ClaI	786 bp + 5963 bp	794 bp + 5971 bp	786 bp + 5946 bp	717 bp + 5894 bp	783 bp + 5960 bp
3. StuI	786 bp + 5963 bp	794 bp + 5971 bp	786 bp + 5946 bp	717 bp + 5894 bp	781 bp + 5958 bp
4. AscI	2606 bp + 4143 bp	2606 bp + 4159 bp	2606 bp + 4109 bp	2606 bp + 4005 bp	2610 bp + 4137 bp

Figure 4.3. Plasmid maps and confirmatory enzymatic digestions of the constructs used for RNAi of the different factors in BSF cells.
Detailed plasmid maps of the constructs used for RNAi. (A) plasmid for targeting of *TbORC1/CDC6*; (B) *TbORC4*; (C) *Tb7980*; (D) *Tb3120*; and (E) *TbORC1B*. The gene-specific products, flanked by an *att*L1 and an *att*L2 sites, are shown integrated in opposing directions,

flanking a 150 bp *lac*Z gene fragment ("stuffer", black arrow). The gene-specific products are expressed by the *RRNA* promoter (green arrow), regulated by tetracycline (Tet operator, TetO, localised downstream of the promoter shown in yellow), while the incomplete hygromycin (*HYG*) ORF (orange arrow) is shown downstream of the *EP procyclin* promoter (red arrow), localised upstream of the RRNA promoter. The bacterial ampicillin resistance gene is shown (*amp*R, large black arrow), as well as the bacterial origin (ColE1 origin, grey arrow). All relevant restriction sites are shown. (G) All constructs were confirmed through restriction enzyme digestions, as described in (Jones *et al.*, 2014). In each case, the 1 Kb plus DNA ladder (Invitrogen™) is shown as a size reference. The expected sizes of the products resulting from the different enzymatic digestions are shown below each gel image. Note that in some cases the same gel image has been cropped so to display the results from the same plasmid side-by-side, and are separated by white space.

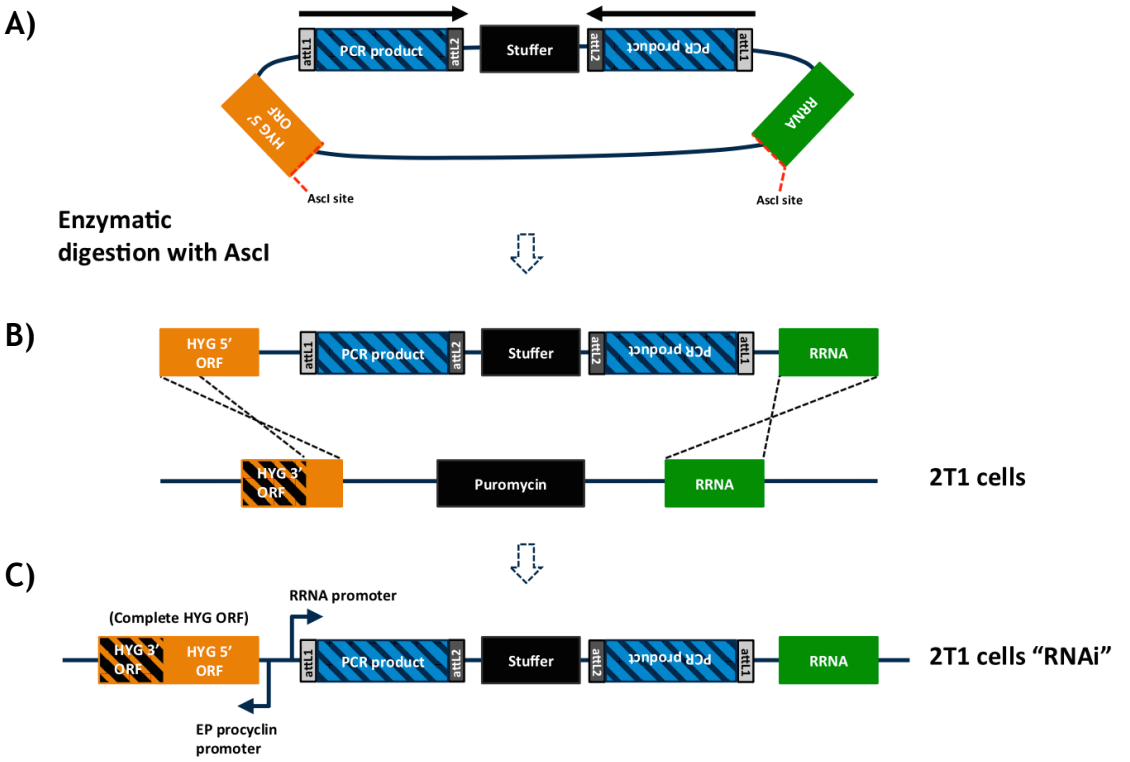


Figure 4.4. Graphic representation of the integration of the RNAi constructs into the genome of 2T1 cells by homologous recombination.
Simplified schematic representation of the process followed to integrate the RNAi construct into the parasite's genome. A) The construct is digested with AscI, excising a fragment containing the incomplete hygromycin (*HYG*) ORF, the two gene-specific PCR products, and the *RRNA* spacer. This is then transfected into the 2T1 cells. B) The excised fragment of the plasmid is then integrated into the parasite genome at the *RRNA* locus that has been tagged with the remaining part of *HYG* (3' region of the gene's ORF). Homologous recombination between the two *HYG* fragments, as well as between the *RRNA* spacer regions takes place, resulting in the integration of the construct fragment into the cell's genome, thus generating a complete *HYG* ORF expressed from the *EP procyclin* promoter (therefore, hygromycin-resistant cells), as well as deletion of the puromycin resistance gene (rendering the cells susceptible to puromycin) (C). Once integrated, the PCR products and "stuffer" are expressed from the inducible *RRNA* promoter, controlled by tetracycline as described in the main text. The diagram was designed based on the descriptions in (Alsford and Horn, 2008; Alsford *et al.*, 2005), and is not represented to scale.

4.2.2 Effect on growth, cell cycle progression, and DNA replication of TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, and Tb3120 expression downregulation by RNAi

In order to investigate the effect of *TbORC1/CDC6*, *TbORC4*, *Tb7980*, *Tb312*0 and *TbORC1B* expression silencing in BSF cells, assays were performed as described previously for PCF cell in Chapter 3, with some adaptations to BSF cells. To assess effects on cell growth, mid-log phase cell cultures (of ~1 x 10^6 cells.ml^{-1}) were diluted to a concentration of 1 x 10^4 cells.ml^{-1}, and divided into two cell culture flasks. To one, 1 µg.ml^{-1} of tetracycline was added (Tet +) in order to induce expression of the stem-loop dsRNA targeting the gene of interest, while the other was left as the non-induced control without tetracycline (Tet -). Cell concentration was then assessed every 24 hours, at which times samples were taken for cell cycle analysis by DAPI staining (as described in Chapter 3), as well as for assessment of EdU incorporation (see below). Due to the low concentration of the cultures and severity of some phenotypes, it was not possible to analyse the samples for DNA content by flow cytometry, as described for PCF cells in Chapter 3 and as performed in both (Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013). In contrast to the cell culture media in which PCF cells are grown (SDM-79), the media routinely used to grow BSF cells, either HMI-9 (Hirumi and Hirumi, 1989) or its variations such as HMI-11 used to grow 2T1 cells (Alsford *et al.*, 2005), contains a considerable amount of thymidine. This rendered EdU-incorporation assays very unreliable (Dr Daniel Paape, personal communication), as the thymidine in the HMI-9 and -11 media appears to compete with EdU for incorporation into the DNA. For this reason, cells were cultured in thymidine-free HMI-11 media (Dr Gloria Rudenko lab, personal communication; recipe in materials and methods Chapter 2, section 2.3.3) for 24 h prior to the assay, as well as for the full course of the experiment. For cloning and routine maintenance, however, the cells were cultured in regular HMI-11 medium. Furthermore, although the use of a thymidine-free media improved EdU uptake by the 2T1 cells, it was necessary to increase both the concentration of EdU used and the incubation period relative to PCF cells to generate a reproducible and reliable experimental protocol: whereas PCF cells were incubated for 3 h with 50 µM of EdU, here BSF 2T1 cells were incubated for 4 h with 150 µM of EdU (optimal conditions established by Dr Daniel Paape). BSF cells have a faster generation time in culture than PCF,

though measurements of replication rate suggest this does not differ dramatically between the two life cycle stages (Calderano *et al.*, 2015). It seems likely, then, that the need for increased levels of EdU and incubation time might be a consequence of non-cell cycle factors, such as EdU uptake. It should be noted, however, that EdU (like BrdU) is somehow, though not to the same extent as in *Leishmania*, toxic to *T. brucei* (Reynolds *et al.*, 2014; Borst and Sabatini, 2008), as well as to other cells (Zhao *et al.*, 2013; Fujii *et al.*, 2002; Duque and Rakic, 2011), leading to pronounced S phase arrest at the concentrations used if incubated for long periods of time (Dr Daniel Paape, unpublished). Therefore, limiting the incubation time was felt necessary to exclude (as much as possible) confounding effects on DNA replication and cell cycle. Efficiency of the induction of the gene-specific targeting RNAi was assessed either by RT-qPCR, for the TbORC4 RNAi cell line, or western blot, for the remaining cell lines, both as described in Chapter 3 and in the materials and methods Chapter 2, sections 2.4.2.6 and 2.2.5.3. Overall, effects of individual downregulation of *TbORC1/CDC6* and *TbORC1B* were assessed in two independently performed experiments, while the effects of *Tb7980*, *Tb3120* and *TbORC4* silencing were only investigated once. In each experiment, the parental cell line, referred to here simply as 2T1 wt, was used as a negative control.

4.2.2.1 Cellular growth, cell cycle progression and DNA replication, are not affected in 2T1 cells by the presence of tetracycline

As expected, no clear growth or DNA replication defects were observed between the Tet – and Tet + cultures of the parental 2T1 wt cells (Figure 4.5, A-C), thus showing that the presence of tetracycline alone did not cause detectable effects on any of the analysed parameters. Nevertheless, there was some variability in the cell cycle profile obtained by analysis of the nuclear (N) and kinetoplast (K) ratios of DAPI-stained cells. The profile appeared to change between 24 h and 48 h, even in the absence of tetracycline, perhaps reflecting the change in cell density from mid-log phase at the 24 h time point, to a high density by 48 h (Figure 4.5, A). Therefore, the cell cycle profile of the non-induced cultures (Tet -) is shown for all cell lines (below) at both 24 h and 48 h post RNAi induction (Figure 4.5, B). Addition of tetracycline also did not cause any impairment in EdU uptake by the cells, suggesting no effects on DNA replication: with or

without tetracycline, ~95% of cells displayed a detectable nuclear signal at both time points analysed (Figure 4.5, C).

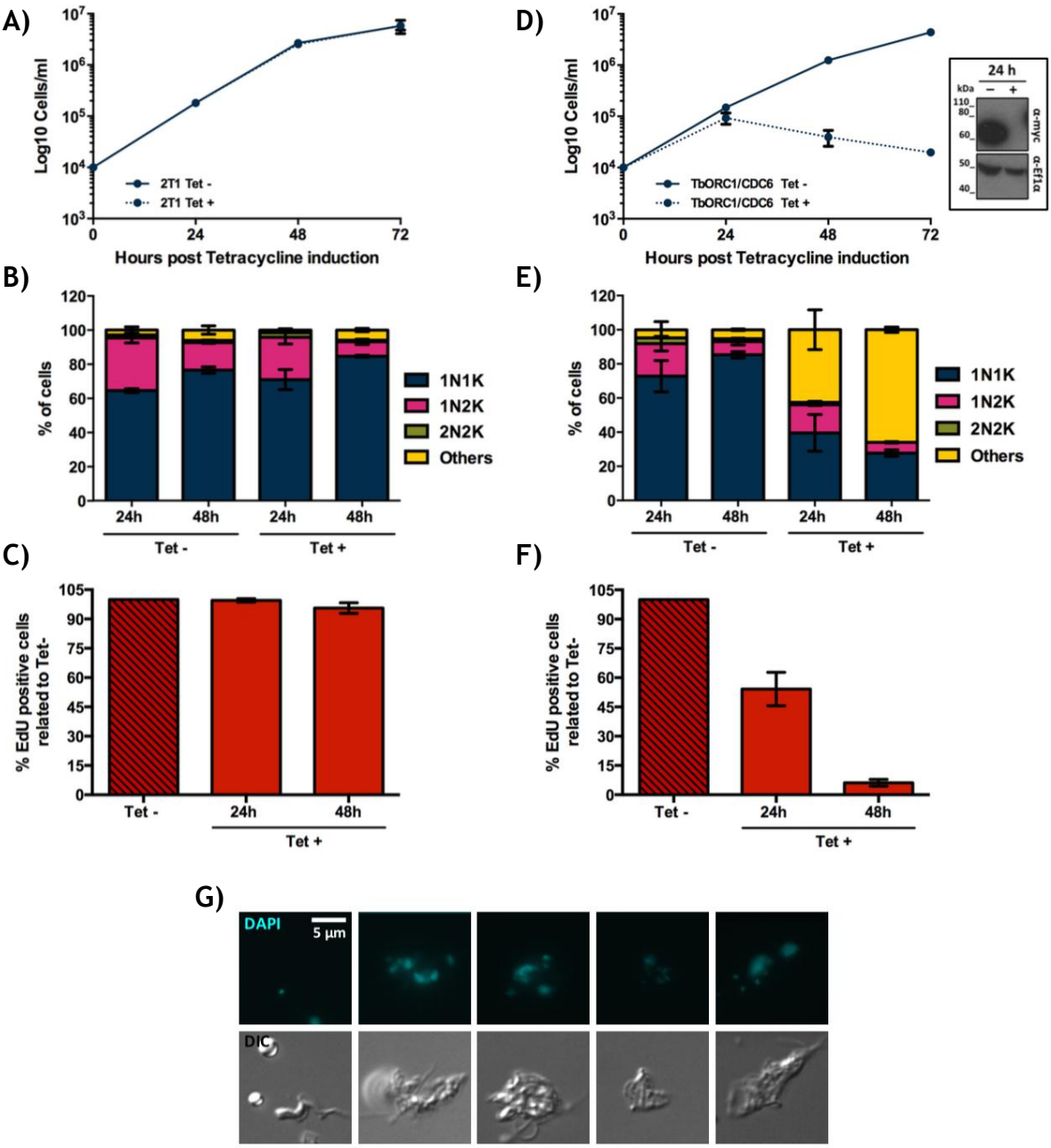


Figure 4.5. Effect of tetracycline on the parental 2T1 cell line, and effect of TbORC1/CDC6 expression silencing by RNAi.
A) to C) 2T1 cell line, and D) to F) TbORC1/CDC6 RNAi cell line. A) and D) show growth curves comparing cultures containing tetracycline (Tet +) or not (Tet -). Cell concentration was assessed every 24 h, and plotted as Log_{10} values (y-axis). The individual points represent the mean concentration calculated from two independent experiments (n = 2), while the error bars depict the standard error of the mean (SEM). In D) (insert box), TbORC1/CDC6^{12myc} protein levels in the Tet - and Tet + samples at 24 h post-induction are shown by a western blot of cell extracts probed with anti-myc antiserum. The transcription elongation factor Ef1α (α-Ef1α) was used as a loading control. B) and E) show the quantification of cells in the different cell cycle stages at 24 h and 48 h after RNAi induction, based on the nuclear (N) and kinetoplast (K) configuration of individual cells stained with DAPI. A minimum of 125 cells was counted per time point and experimental group (Tet - and Tet +), and percentages of each stage (1N1K, 1N2K, 2N2K, and "others") were calculated relative to the total amount of analysed cells. The graph represents the mean of each cellular population obtained in two independent experiments (n = 2), while the error bars refer to the SEM.

C) and F) show the percentage of EdU positive cells in the Tet + sample relative to the percentage of EdU positive cells in the Tet - culture from the respective time point. A minimum of 125 cells were analysed per time point and group (Tet - and Tet +). The mean from two independent experiments is shown (n = 2), with error bars representing the SEM. G) representative DAPI-stained and DIC images of cells categorised as 'others', seen at 24 h and 48 h after TbORC1/CDC6 RNAi induction.

4.2.2.2 RNAi of TbORC1/CDC6

Induction of RNAi against TbORC1/CDC6 resulted in a growth defect that was noticeable at 24 h post-induction and, subsequently, apparent cell death (Figure 4.5, D). This growth impairment was accompanied by a severe change in the cell cycle profile, with an increase, to ~40% of the population at 24 h, in the number of cells displaying an abnormal nucleus and kinetoplast ratio (distinct from the expected 1N1K, 1N2K and 2N2K configurations) (Figure 4.5, E). This was further aggravated at 48 h, with abnormal cells constituting more than 60% of the population, with a parallel decrease in the number of 1N1K and 1N2K cells, and the virtual abolishment of 2N2K cells (Figure 4.5, E). The abnormal cells displayed a variety of morphologies (Figure 4.5, G): large cells with multiple nuclei and kinetoplasts, cells with only multiple kinetoplasts, cells with multiple flagella, and zoids (0N1K, although these were rare). Thus, there was not a clear common morphology amongst the aberrant cells, which were therefore simply categorised as "others" (Figure 4.5, E). Analysis of EdU incorporation showed that loss of EdU signal was concomitant with the emergence of the growth and cell cycle defects: the number of EdU positive cells was reduced by ~50% in the induced cells when compared to un-induced (Tet –) at 24 h, and further reduced to only ~5% by 48 h post-induction (Figure 4.5, F). Analysis, by western blot, of the levels of TbORC1/CDC6^{12myc} showed that virtually no protein could be detected 24 h after RNAi induction (Figure 4.5, D, insert box). Together, these results, in particular the abrogation of EdU incorporation, corroborate previous evidence claiming the involvement of TbORC1/CDC6 in DNA replication in BSF cells (Benmerzouga *et al.*, 2013). However, these phenotypes observed as a whole do not entirely overlap with the previous reports, as discussed below (section 4.6.1).

4.2.2.3 RNAi of TbORC1B

Prior to this work, the effects of TbORC1B expression silencing have not been reported. Like in PCF cells, efficient downregulation of TbORC1B (no

TbORC1B^{12myc} protein was detectable 24 h post-induction; Figure 4.6, A, insert box) resulted in the most severe phenotype amongst the analysed Orc-like factors. A clear growth impairment was observed 24 h after RNAi induction, with pronounced cell death by 48 h (Figure 4.6, A). Thereafter, growth appeared to recover, as cell numbers increased by 72 h (Figure 4.6, A), and were apparently growing normally 96 h after tetracycline addition (data not shown). Other clones analysed showed the same results (data not shown), and it is possible that this represents selection for revertants in which RNAi knockdown is no longer effective. This however, was not confirmed by western blot detection of protein levels at such time points. Nevertheless, during the period of growth impairment (24 h and 48 h time points) a reduction in more than 90% in the number of EdU positive cells (Figure 4.6, C) was observed, suggesting a strong impairment of DNA replication. In addition, abnormal cells ("others") accumulated during this period, constituting more than 60% of the population at 24 h post-induction, and ~90% of the population at 48 h, with a severe reduction in 1N1K and 2N2K cells, and a mild reduction in 1N2K cells (Figure 4.6, B). These aberrant cells shared the same, mixed nucleus and kinetoplast ratios as seen following TbORC1/CDC6 RNAi (data not shown). Taken together, these data show that, like in PCF cells, the effects of TbORC1B silencing were more severe than the ones observed with TbORC1/CDC6 downregulation, appearing to show the same phenotypes earlier after RNAi induction (24 h versus 48 h). Overall, these results confirm that TbORC1B, like in PCF cells, has a role in DNA replication in BSF cells, and its loss affects this process more rapidly than the loss of TbORC1/CDC6, perhaps supporting the hypothesis of TbORC1B as a regulatory factor, suggested in Chapter 3.

4.2.2.4 RNAi of TbORC4

In PCF cells, TbORC4 silencing resulted in an identical phenotype to that seen after TbORC1/CDC6 downregulation, suggesting that the two factors act in the same biological context, potentially in a complex (Chapter 3). Previous results in BSF cells suggested the same overlapping response (Tiengwe *et al.*, 2012b). Surprisingly, this phenocopying was not observed here (Figure 4.6, D). No clear effects on growth, cell cycle or EdU incorporation were observed 24 h post-induction of TbORC4 RNAi (Figure 4.6, D-F). Nevertheless, a small growth defect was noticeable at 48 h post-induction, in parallel with a reduction in ~35% of

EdU positive cells and an increase in the number of abnormal cells ("others") in the population (to ~40% of the population), with concomitant reduction in the number of 1N1K cells (Figure 4.6, D-F). Unfortunately, it was not possible to endogenously tag TbORC4 in order to monitor protein levels. Instead, RT-qPCR was performed, suggesting only a small reduction in ~20% relative *TbORC4* mRNA levels at 24 h after RNAi-induction. It is possible that such a small reduction indicates inefficient degradation of *TbORC4* mRNA after RNAi-induction, and might explain the differences in phenotype relative to TbORC1/CDC6. However, it is worth noting that this is a similar level of *TbORC4* mRNA loss observed in the PCF TbORC4 RNAi cells, which resulted in a severe phenotype (Chapter 3). This experiment was only conducted once, and therefore requires repeating (and perhaps analysis of other clones), but the results suggest a role for TbORC4 in DNA replication in BSF cells, though at present, it is not clear how comparable this function is to that of TbORC1/CDC6.

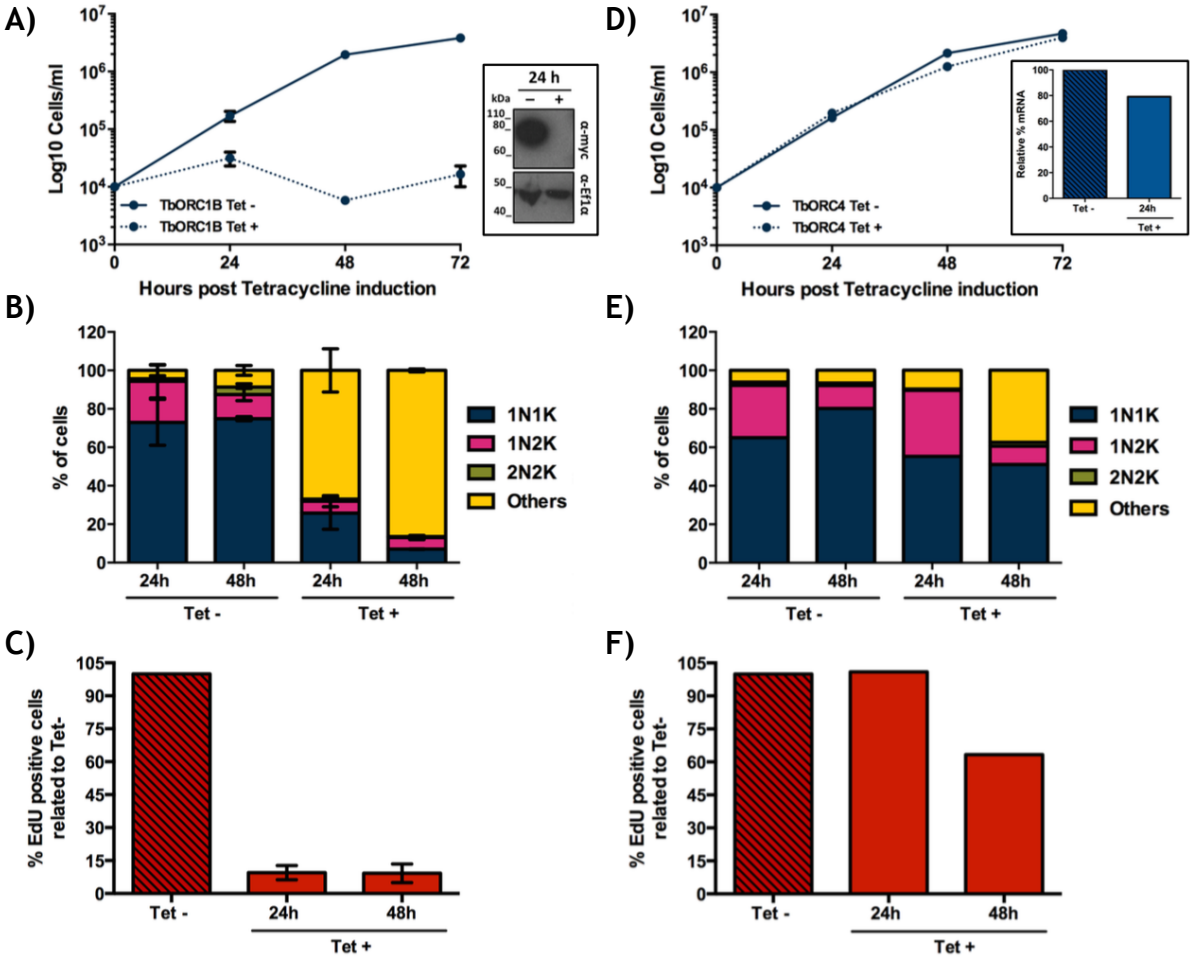


Figure 4.6. Effect of TbORC1B and TbORC4 expression silencing by RNAi.
A) to C) TbORC1B RNAi cell line, and D) to F) TbORC4 RNAi cell line. A) and D) shows growth curves comparing cultures containing tetracycline (Tet +) or not (Tet -). Cell concentration was assessed every 24 h, and plotted as Log_{10} values (y-axis). In A), the individual points represent the

mean concentration calculated from two independent experiments (n = 2), while the error bars depict the SEM. In D) the results refer to a single experiment, thus no error bars are represented. In A) (insert box), TbORC1B^{12myc} samples at 24 h post-induction are shown by a western blot of cell extracts probed with anti-myc antiserum. The transcription elongation factor Ef1α (α-Ef1α) was used as a loading control. In D), *TbORC4* relative mRNA levels assessed by RT-qPCR are shown within the insert box. The results refer to the relative amount of mRNA levels in the Tet + sample to the non-induced sample (Tet -), calculated using the $\Delta\Delta C_t$ method. Results refer to a single experiment (n = 1) using three technical replicates (the mean is represented), and therefore no error bars are shown. B) and E) show the quantification of cells in the different cell cycle stages at 24 h and 48 h after RNAi induction, based on the nuclear (N) and kinetoplast (K) configuration of individual cells stained with DAPI. A minimum of 125 cells was counted per time point and experimental group (Tet - and Tet +), and percentages of each stage (1N1K, 1N2K, 2N2K, and "others") were calculated relative to the total amount of analysed cells. In B), the graph represents the mean of each cellular population obtained in two independent experiments (n = 2), while the error bars refer to the SEM. In E) results refer to a single experiment so error bars are not shown. C) and F) show the percentage of EdU positive cells in the Tet + sample relative to the percentage of EdU positive cells in the Tet - culture from the respective time point. A minimum of 125 cells were analysed per time point and group (Tet - and Tet +). In C) the mean from two independent experiments (n = 2) is shown, with error bars representing the SEM. F) Shows the results of a single experiment and therefore no error bars are shown.

4.2.2.5 RNAi of Tb7980

In a previous study (Tiengwe *et al.*, 2012b), downregulation of Tb7980 in BSF cells resulted in a strikingly similar phenotype to the one detected for TbORC1/CDC6 silencing, but it was not confirmed to be a result of impaired DNA replication. Unfortunately, none of the RNAi clones obtained here for this gene (one representative is shown) showed any growth phenotype (Figure 4.7, A), or evidence for reduction in protein levels after RNAi induction (Figure 4.7, A, insert box). No differences in cell cycle or EdU incorporation were observed at the 24 h and 48 h time points (Figure 4.7, B-C), although the cell line had, intrinsically, a considerable number of abnormal cells (Figure 4.7, B, "others"). It appears unlikely that effective RNAi against Tb7980 occurred, precluding discussion of its function in the BSF. It will be therefore, necessary to generate new cell lines and test more clones for an efficient reduction in Tb7980 protein levels in order to take elations on Tb7980's role in DNA replication.

4.2.2.6 RNAi of Tb3120

The effects of Tb3120 silencing have never been investigated in BSF cells, while in PCF cells, the effects were only noticeable at later time points than those observed for TbORC1/CDC6, TbORC4 or TbORC1B downregulation (Chapter 3). Although protein levels appeared to decrease 48 h after RNAi induction (Figure 4.7, D, box insert), no growth, cell cycle or EdU incorporation effects were detected at that time point (Figure 4.7, D-F). Like in PCF cells, it is possible that

any effects of Tb3120 RNAi induction will only emerge after a certain number of generations, and it may then be important to repeat the experiment for a longer period of time. Presently, however, nothing can be inferred about Tb3120's role in DNA replication in BSF cells.

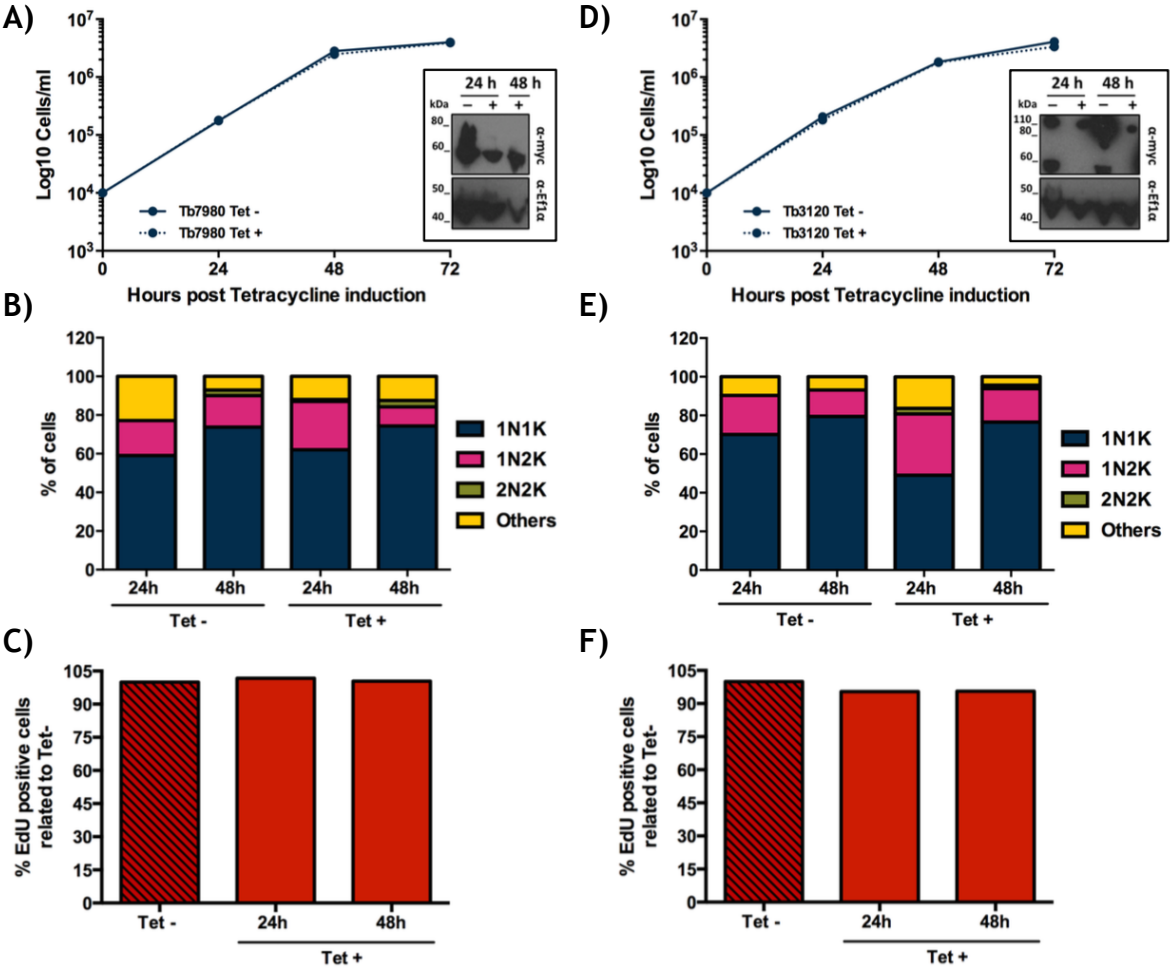


Figure 4.7. Effect of Tb7980 and Tb3120 expression silencing by RNAi.
A) to C) Tb7980 RNAi cell line, and D) to F) Tb3120 RNAi cell line. A) and C) show the growth curves comparing cultures containing tetracycline (Tet +) or not (Tet -). Cell concentration was assessed every 24 h, and plotted as Log_{10} values (y-axis). The results refer to a single experiment, thus no error bars are shown. In A) and D) (insert boxes), Tb7980^{12myc} and Tb3120^{12myc} (respectively) protein levels in the Tet - and Tet + samples at 24 h and 48 h post-induction are shown by a western blot of cell extracts probed with anti-myc antiserum. The transcription elongation factor Ef1α (α-Ef1α) was used as a loading control. B) and E) show the quantification of cells in the different cell cycle stages at 24 h and 48 h after RNAi induction, based on the nuclear (N) and kinetoplast (K) configuration of individual cells stained with DAPI. A minimum of 125 cells was counted per time point and experimental group (Tet - and Tet +), and percentages of each stage (1N1K, 1N2K, 2N2K, and "others") were calculated relative to the total amount of analysed cells. The results refer to a single experiment, so error bars are not shown. C) and F) show the percentage of EdU positive cells in the Tet + sample relative to the percentage of EdU positive cells in the Tet - culture from the respective time point. A minimum of 125 cells were analysed per time point and group (Tet - and Tet +). Both represent the results of a single experiment and therefore no error bars are shown.

Together, the results of these RNAi studies confirm roles for TbORC1/CDC6 and TbORC1B, and to some extent, TbORC4 as well, in DNA replication in BSF cells, while no conclusions could be drawn about Tb7980 or Tb3120 function in this life cycle stage of the parasite. Complementing these experiments with DNA content analysis by flow cytometry would be valuable, as well as generating a cell line targeting Tb1120 in order to test the function of this factor in BSF cells.

4.3 Generation of TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120 *in situ* – tagged cell lines

In Chapter 3, the subcellular localisation of TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120 was investigated in PCF cells after tagging each factor endogenously with 12myc. With the exception of TbORC1B, this revealed that all these factors localise to the nucleus of the cell throughout the cell cycle, while TbORC1B expression and localisation appeared to be cell cycle dependent. To date, no information is available on these factors' localisation and dynamics in BSF cells. As shown in Chapter 3 and in the previous section, some of the phenotypes resulting from TbORC1/CDC6 expression silencing are different in PCF and BSF cells. Thus, although it might be predicted that these factors would display the same subcellular behaviour in both PCF and BSF cells, in the case that each of these acts as a core component of the replication initiation process, it was considered that analysis of the proteins subcellular localisation and dynamics in BSF cells could be informative.

4.3.1 Cloning the constructs for endogenous tagging with 12myc

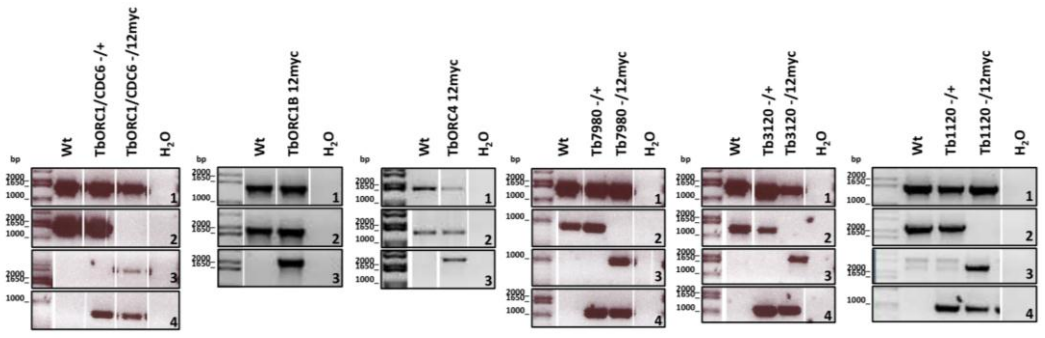
The constructs described in Chapter 3, section 3.4.1, formed the basis for the experiments below. These plasmids were designed and constructed using the DNA sequences and gDNA from *T. brucei* cells of strain TREU 927, the strain in which the PCF localisation studies described in Chapter 3 were conducted. The cells used for protein localisation in the BSF stage of the parasite were, however, strain Lister 427. Alignments of the gene sequences available on TriTrypDB for TREU 927 and Lister 427 revealed that, while *TbORC1/CDC6* and *TbORC1B* were identical in the two strains, all other gene sequences possessed

minor nucleotide differences that resulted in amino acid changes. Some of these were in the 3' region of the gene ORFs cloned into the pNAT^{x12M} plasmid vector (alignments of the 3' regions cloned into the vector are shown in appendices section 7.4.1). For this reason, the constructs for the endogenous tagging of TbORC4, Tb3120 and Tb1120, with 12myc at the C-terminus, and Tb7980, with 12myc at the N-terminus, were re-made using Lister 427 gDNA, as detailed in Chapter 3, section 3.4.1. The resulting plasmids maps are in all identical to the ones shown in Chapter 3, section 3.4.1, and therefore are not shown here. The new constructs were confirmed by enzymatic digestion (shown in the appendices, Figure 7.33), as well as sequencing, as detailed in Chapter 3, section 3.4.1, and described in the materials and methods Chapter 2, sections 2.2.1.9.

4.3.2 Generation of the 12myc endogenously tagged TbORC1/CDC6, TbORC1B, TbORC4, Tb7980, Tb3120 and Tb1120 cell lines, and confirmation by PCR and western blot

As explained in Chapter 3, it is important, whenever possible, to ask if the tagged version of a protein is functional. One way to address this relies on deleting the non-tagged allele, thus ensuring that only the tagged protein is expressed in the cell and allowing checking for relevant phenotypes. As described in Chapter 3, constructs were generated to delete one of the alleles of each factor's gene by replacing it with an antibiotic resistance marker gene (Chapter 3, section 3.4.3). To attempt this in the BSF cells, one of the two alleles was first deleted, prior to the endogenous tagging of the remaining allele. For this, BSF Lister 427 cells (here referred to as 427 wt), were transfected with the KO constructs TbORC1/CDC6-KO-PURO, TbORC1B-KO-NEO, TbORC4-KO-NEO, Tb7980-KO-NEO, Tb3120-KO-NEO and Tb1120-KO-NEO (maps shown in Chapter 3, Figure 3.16), as described in the materials and methods Chapter 2, section 2.3.5. One of the gene alleles was successfully deleted for *TbORC1/CDC6*, *Tb7980*, *Tb3120* and *Tb1120*, as confirmed by PCR (detailed in the materials and methods Chapter 2, section 2.2.2): in each of these cell lines, here referred to as TbORC1/CDC6 -/+, Tb7980 -/+, Tb3120 -/+, and Tb1120 -/+, the KO construct integrated into the right locus (Figure 4.8, row 4, KO PCR), and the remaining wild type allele was still detectable (Figure 4.8, row 2, Wt allele PCR). To date, although attempted multiple times, it was not possible to disrupt

one of the alleles of either *TbORC1B* and *TbORC4,* and it is not clear whether the parasite requires the presence of the two copies of each gene for survival. The TbORC1/CDC6 -/+, Tb7980 -/+, Tb3120 -/+, and Tb1120 -/+ cell lines were further transfected with the respective 12myc-tag constructs, whose integration was confirmed by PCR (Figure 4.8), as described previously in Chapter 3 (Figure 4.8, row 3, myc PCR). These cell lines, named TbORC1/CDC6 -/12myc, Tb7980 -/12myc, Tb3120 -/12myc and Tb1120 -/12myc, were also confirmed to not possess a copy of the wild type allele (Figure 4.8, row 2, Wt allele PCR), and to have retained the deleted allele (Figure 4.8, row 4, KO PCR). Parental 427 wt cells were transfected with the 12myc-tag constructs for the endogenous tagging of *TbORC1B* and *TbORC4*. The resulting cell lines, TbORC1B 12myc and TbORC4 12myc, were then confirmed by PCR (Figure 4.8, row 3, myc PCR), and shown to still possess the wild type allele (Figure 4.8, row 2, Wt allele PCR). Like for the PCF cell lines (Chapter 3), the expression of each 12myc-tagged protein was confirmed by western blot, as described in the materials and methods Chapter 2, section 2.2.5.3. All cell lines were confirmed to express a 12myc-tagged protein of the expected size, as shown in Figure 4.9, and were therefore used for further analysis.



1. Control PCR	1543 bp	1543 bp	1543 bp	1543 bp	1543 bp	1543 bp
2. Wt allele PCR	1508 bp	1847 bp	1234 bp	870 bp	1485 bp	2300 bp
3. myc PCR	2281 bp	1941 bp	1993 bp	885 bp	1997 bp	2393 bp
4. KO PCR	817 bp	----------	----------	903 bp	1023 bp	960 bp

Figure 4.8. Confirmation of construct integration by PCR.
Panels 2-4 show the products resulting from PCR reactions using different pairs of primers designed specifically to assess the presence of the wild type allele (row 2), to confirm integration of the C- or N-terminal endogenous tagging constructs (12myc, row 3), and successful allele deletion (KO, row 4). Panels in row 1 show a PCR-amplification from *TbMCM10*, an unrelated gene, as a control for the presence of gDNA. In all PCR reactions, the parental wild type cell line, 427 wt (represented as Wt), was used both as a negative (12myc PCR and KO PCR) or positive control (Wt allele PCR). The heterozygote TbORC1/CDC6 -/+, Tb7980 -/+, Tb3120 -/+, and Tb1120 -/+ cell lines, generated prior to 12myc tagging, provide an extra negative control of the 12myc PCR, and a positive control of the KO PCR. In each case, the 1 Kb plus DNA ladder (Invitrogen™) is shown as a size reference, and the expected PCR product sizes, resulting from the different PCR reactions, are shown below each gel image. Note that in some cases the same gel image has been cropped so to display the results from the same PCR reaction, and are separated by a white space.

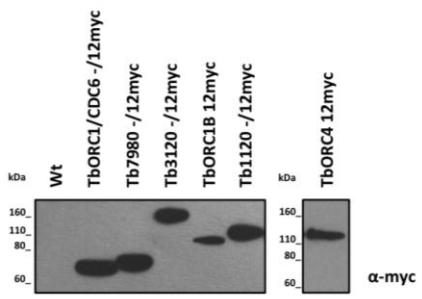


Figure 4.9. Confirmation of the expression of 12myc tagged proteins in BSF cells.
Total protein extracts from 427 wt, TbORC1/CDC6 -/12myc, Tb7980 -/12myc, Tb320 -/12myc, TbORC1B 12myc, Tb1120 -/12myc and TbORC4 12myc cell lines were separated by SDS PAGE and analysed by western blot. The 12myc-tagged proteins were detected with an anti-myc (represented as α-myc) antiserum. Only the cell lines used for further experiments are represented. The Novex® Sharp Protein Standard is shown as a size reference.

4.4 Cellular localisation and cell cycle dynamics of TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7890, Tb3120^{12myc}, Tb1120^{12myc} and TbORC1B^{12myc} in bloodstream form cells

4.4.1 Cell lines

In order to investigate whether the endogenously 12myc-tagged cells, generated as described in the previous sections, behaved comparably to the parental cell line, 427 wt, all cells were set up in culture at a concentration of 1 x 10^4 cells.ml^{-1}, and their cell density monitored every 24 h for a total period of 72 h. Like what was observed for the endogenously tagged PCF cells (Chapter 3, section 3.5.1), no growth defects were observed for any of the BSF cells, each of which grew in culture at a rate indistinguishable from the untagged 427 wt (Figure 4.10). Growth curve analysis was not performed for the TbORC4 12myc cell line, but routine maintenance in cell culture did not suggest any defects (not shown). In contrast to the PCF cells (shown in Chapter 3, section 3.5.1), no cell cycle or DNA replication analyses were performed, but from their normal growth rate, it has been assumed that these 12myc-tagged cell lines behave similarly to the wild type parental cell line, 427 wt, and therefore, that at least TbORC1/CDC6^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc} and Tb1120^{12myc} versions of the proteins are functional. As shown in section 4.2.2.3, depletion of TbORC1B by RNAi results in a severe phenotype, possibly suggesting that if the 12myc-tagged version of the protein were not functional, a defect in growth would be observed, which is not the case.

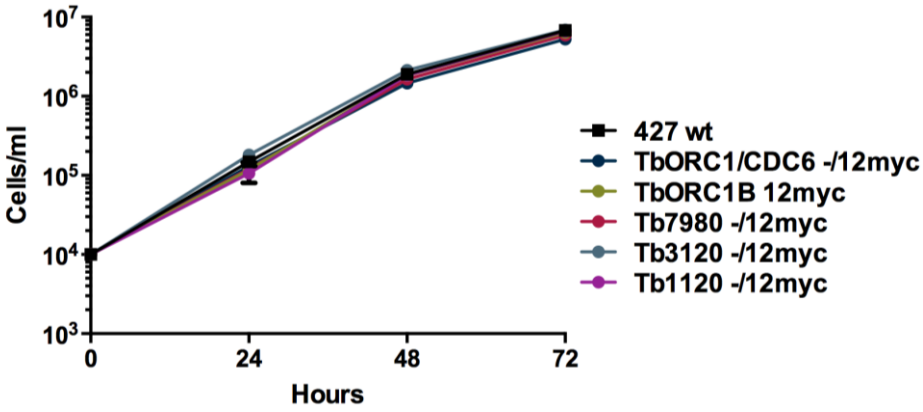


Figure 4.10. Analysis of the growth of endogenously 12myc tagged cells in culture. The different cell lines were transferred to a new culture flask with fresh HMI-9 culture medium, without drugs, at a starting concentration of 1 x 10^4 cells.ml^{-1}. Cell density was assessed every 24 h, and plotted as Log_{10} values (y-axis). The individual points on the graph represent the mean of three independent experiments (n = 3), with the error bars representing the standard error of the mean.

4.4.2 Subcellular localisation and cell cycle dynamics of TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc}, Tb1120$^{12\ myc}$, and TbORC1B^{12myc}

TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc} and Tb1120^{12myc} were detected by immunofluorescence assay (IFA) as described in detail in Chapter 3, section 3.5.2, with slight changes to the overall protocol as a consequence of the morphological differences between the PCF and BSF cells (detailed in the materials and methods Chapter 2, section 2.5.4). All images shown below were acquired with a DeltaVision imaging system.

As for PCF cells, the anti-myc monoclonal antibody conjugated with AlexaFluor® 488 used in these assays appears to be specific to the 12myc tagged proteins, as IFA of non-tagged 427 wt cells resulted in no discernible signal (Figure 4.11, A). Nevertheless, it appears that in BSF cells there is a greater background signal than the one that was detected in PCF cells. Immunofluorescence of TbORC1/CDC6 -/12myc cells showed that TbORC1/CDC6^{12myc} localised to the nucleus of the cells, as a punctate signal, throughout the cell cycle (Figure 4.11, B). The same pattern of localisation was also observed for TbORC4^{12myc} (Figure 4.12), ^{12myc}Tb7980 (Figure 4.13), Tb3120^{12myc} (Figure 4.14) and Tb1120^{12myc} (Figure 4.15) when analysing TbORC4 12myc, Tb7980 -/12myc, Tb3120 -/12myc and Tb1120 -/12myc cell lines, respectively. Moreover, quantification of the cells displaying nuclear staining showed that TbORC1/CDC6^{12myc}, TbORC4^{12myc},

^{12myc}Tb7980, Tb3120^{12myc} and Tb1120^{12myc} were found in the nucleus of ~100% of cells (Figure 4.17, A). Thus, each of the above factors displayed constitutive nuclear localisation throughout the cell cycle – the same pattern of localisation that was seen in PCF cells (Chapter 3, section 3.5).

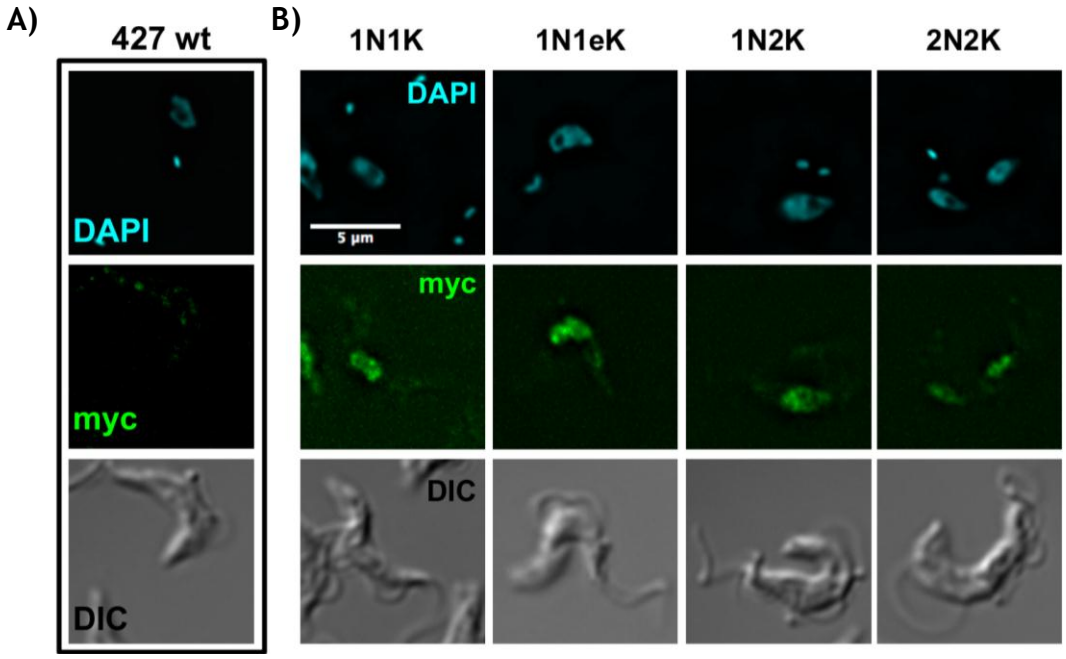


Figure 4.11. Immunofluorescence of TbORC1/CDC6^{12myc}.
A) Panels within the box show the staining of 427 wt cells with DAPI (top panel), AlexaFluor® 488-conjugated anti-myc antibody (middle panel), and the cell outline by DIC. B) Panels show the staining of TbORC1/CDC6 -/12myc cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show TbORC1/CDC6^{12myc}, recognised by the anti-myc antiserum. Lower panel row shows the cell outline by DIC. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

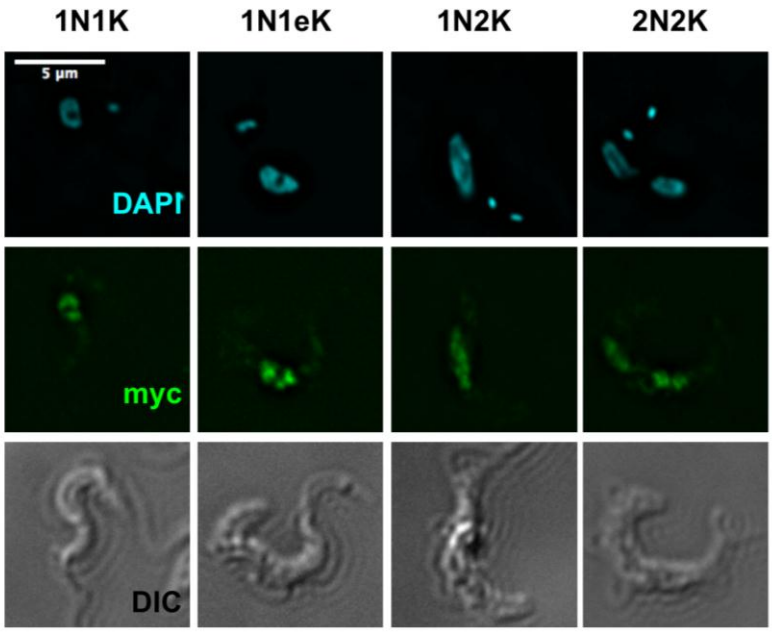


Figure 4.12. Immunofluorescence of TbORC4^{12myc}.
Panels show the staining of TbORC4 12myc cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show TbORC4^{12myc}, recognised by the anti-myc antiserum. Lower panel row shows the cell outline by DIC. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

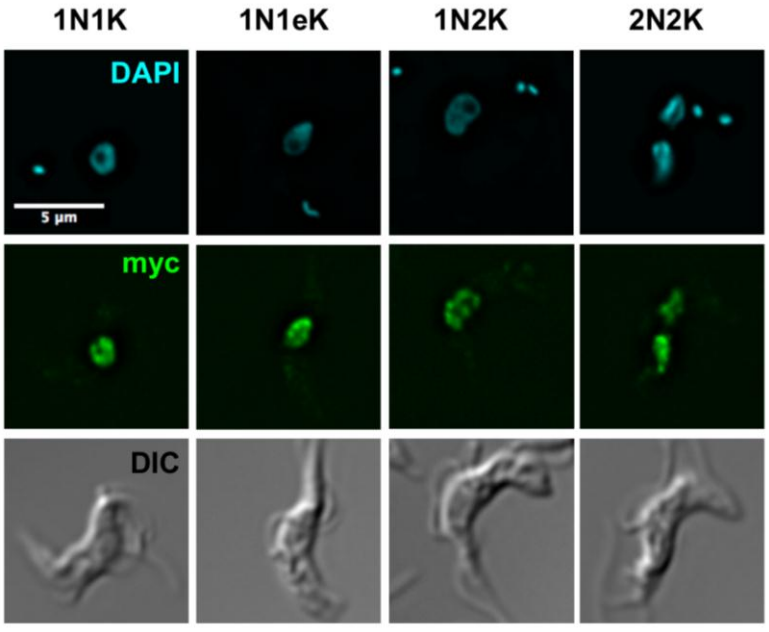


Figure 4.13. Immunofluorescence of Tb7980^{12myc}.
Panels show the staining of Tb7980 -/12myc cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show Tb7980^{12myc}, recognised by the anti-myc antiserum. Lower panel row shows the cells outline by DIC. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

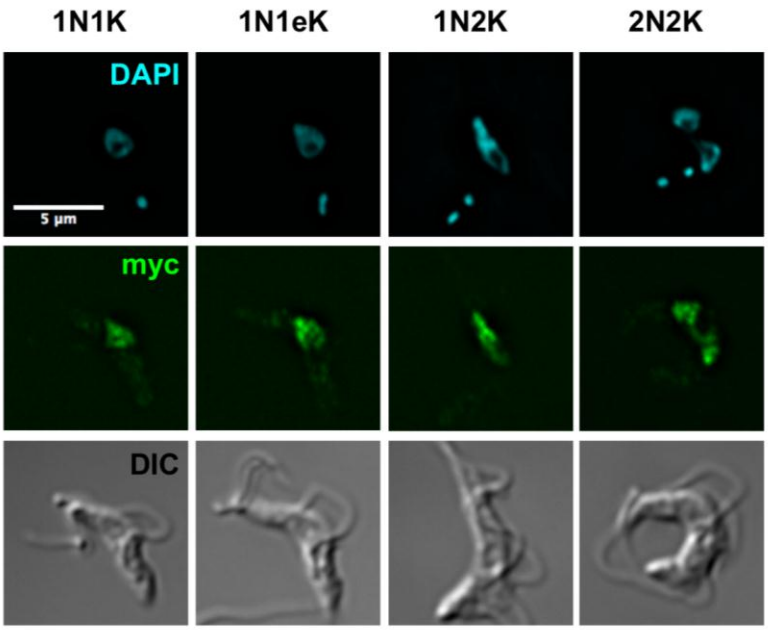


Figure 4.14. Immunofluorescence of Tb3120^{12myc}.
Panels show the staining of Tb3120 -/12myc cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show Tb3120^{12myc}, recognised by the anti-myc antiserum. Lower panel row shows the cell outline by DIC. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

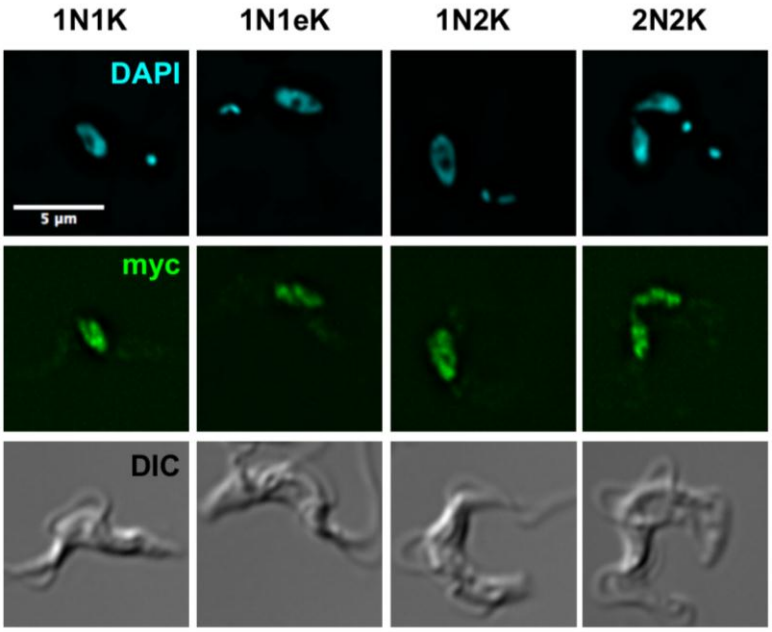


Figure 4.15. Immunofluorescence of Tb1120^{12myc}.
Panels show the staining of Tb1120 -/12myc cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show Tb1120^{12myc}, recognised by the anti-myc antiserum. Lower panel row shows the cell outline by DIC. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

In contrast, and again similar to the observations made in PCF cells (Chapter 3, section 3.5.6), TbORC1B^{12myc} localised to the nucleus of only ~45% of cells (Figure 4.17, A). Within the population that showed TbORC1B^{12myc} signal, 18% were 1N1K cells, 68% 1N1eK, and ~13% were 1N2K cells (Figure 4.17, B), which corresponded with ~21% of the total 1N1K cells, 88% of 1N1eK cells and ~45% of 1N2K cells (Figure 4.17, C). TbORC1B^{12myc} signal was detected in only one putative 2N2K cell out of 38 2N2K cells examined from two independent experiments (total of 481 analysed cells). These results show a pronounced similarity with those obtained for PCF TbORC1B 12 myc cells, suggesting that TbORC1B expression or localisation is cell cycle regulated in both life cycle stages. Again, TbORC1B 12myc localised to the nucleus as puncta (Figure 4.16).

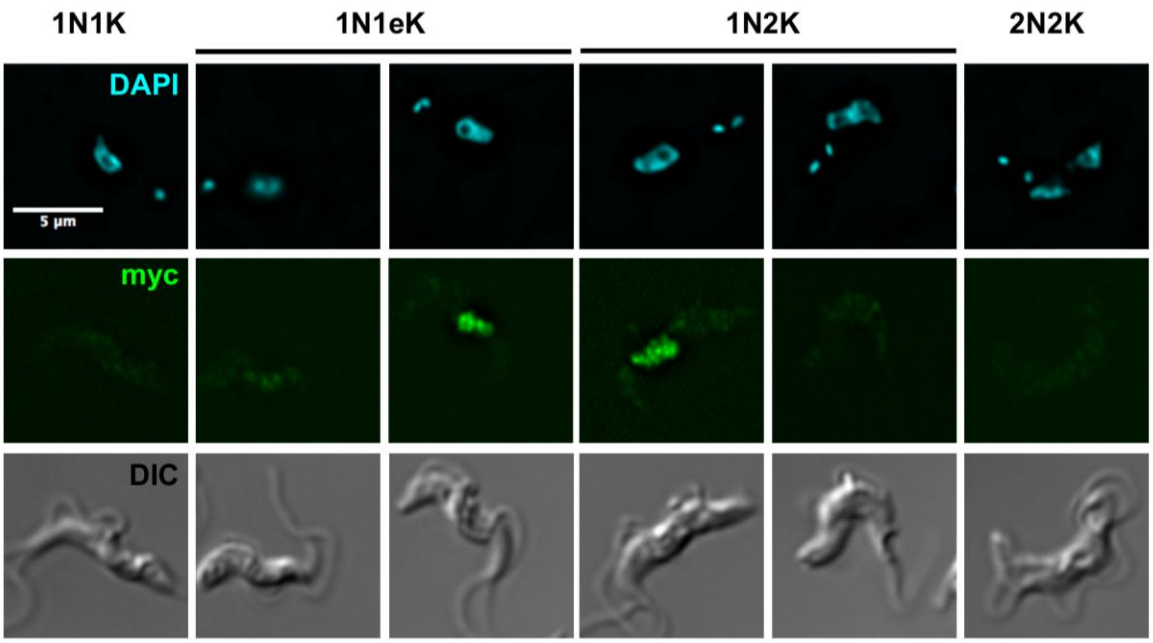


Figure 4.16. Immunofluorescence of TbORC1B^{12myc}.
Panels show the staining of TbORC1B 12myc cells with DAPI (top row panels), allowing their classification in 1N1K cells (G1 phase), 1N1eK cells (S phase), 1N2K (G2 phase), and 2N2K cells (post-mitosis). Panels in the middle row show TbORC1B^{12myc}, recognised by the anti-myc antiserum. Lower panel row shows the cell outline by DIC. Images were acquired using a DeltaVision imaging system, and deconvolved using the ratio conservative method, on SoftWoRx software. Scale bar represents 5 µm.

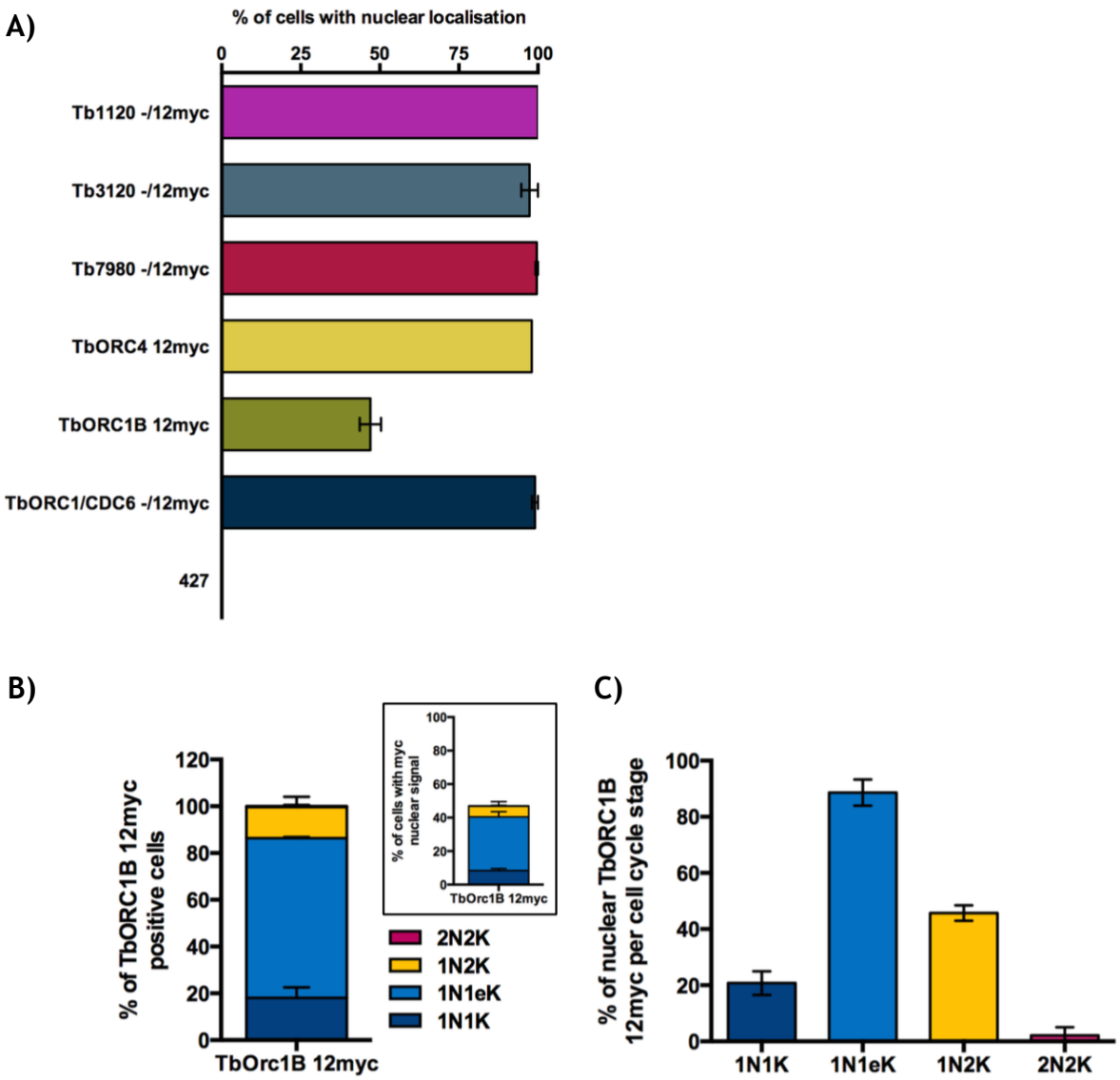


Figure 4.17. Analysis of the percentage of cells containing nuclear myc signal.
A) Representation of the percentage of cells, per cell line, containing signal for the respective 12myc-tagged protein in the nucleus. The mean of two independent experiments is shown (n = 2), with the error bars representing the standard error of the mean (SEM). In each experiment, at least 125 cells were counted per cell line. Percentage is calculated to the total number of cells counted per cell line in the two experiments. B) Quantification of the proportion of cells containing TbORC1B^{12myc} nuclear signal. Only approximately 45% of the cells showed TbORC1B^{12myc} signal in the nucleus as shown in A), and here separated (inside the insert box) into the different cell types (dark blue, 1N1K cells; light blue, 1N1eK cells; yellow, 1N2K cells; pink, 2N2K cells). The main graph represents the total of cells (45% of the total cell population) with TbORC1B^{12myc} nuclear signal, with the percentage of each different cell type individually represented; the mean of two independent experiments (n = 2) is represented, and the error bars correspond to the SEM. C) percentage of TbORC1B^{12myc}-positive cells, per cell type; the mean of two independent experiments is represented (n = 2), and the error bars depict the SEM.

4.5 Origins of DNA replication in *T. brucei* bloodstream form cells

In characterised eukaryotes, by the end of G1 phase, in order to ensure complete genome replication within a cell cycle, cells must have licensed all potential origins of replication in the genome by binding the pre-replication complex (pre-RC). However, only a subset of these sites will eventually be activated during S phase, in an apparent stochastic manner (reviewed in Masai *et al.*, 2010; Mechali, 2010; Leonard and Mechali, 2013; McIntosh and Blow, 2012; Costa *et al.*, 2013). Depending on whether the pre-RC sites are activated or not, and in which circumstances, replication origins have been categorised into three types: constitutive – origins that are always used in every cell cycle or cell type; flexible – potential origins that can be used stochastically in different cells and in different conditions; and dormant – potential origins that in normal growth conditions are not used during the cell cycle, but can be activated in specific cellular programmes (e.g. during development or differentiation) or in conditions of stress (reviewed in Mechali, 2010). Once cells enter S phase, no more origins can be licensed in order to prevent re-replication of the genome (reviewed in Costa *et al.*, 2013; Masai *et al.*, 2010). Therefore, the excess of licenced origins in G1 phase provides the cell with a flexible choice of origins that can be used in the event of replicative stress, obstacles encountered by the replication machinery, poor growth conditions, environmental cues, as well as changes in the chromatin context or transcription program of the cell that might occur upon cellular differentiation (reviewed in Mechali, 2010; Leonard and Mechali, 2013; McIntosh and Blow, 2012). Indeed, it appears that there is a relationship between transcription and DNA replication, with early replicating regions being associated with actively transcribed genes, frequently at the promoter regions, and late replicating origins being associated with non- or poorly transcribed regions of the genome (reviewed in Mechali, 2010; Leonard and Mechali, 2013; McIntosh and Blow, 2012). This association may be because transcription influences chromatin structure at these sites, for instance, by increasing the accessibility of the chromatin, and indeed some histone modifications are associated with gene expression and origins of replication. However, transcription may also interfere with DNA replication, and thus cross talk between the two molecular machineries is most likely essential for the

maintenance of genome stability and cell identity during development and cellular differentiation (reviewed in Mechali, 2010; Helmrich *et al.*, 2013). Nevertheless, though there is flexibility in the choice of activated origins among pre-RC sites in a specific cell type or tissue, origin location and firing timing are reproducibly found at the same sites in a cellular population. For instance, if a different subset of origins is activated in a cell cycle, e.g. due to replication stress or cellular differentiation, this subset will be used in the next cell cycle, and thus forming the innate origin profile of the different cell types (reviewed in Mechali, 2010; Masai *et al.*, 2010; Leonard and Mechali, 2013; Jackson *et al.*, 2012). Therefore, at least in multicellular organisms, the origins of replication activated per cell cycle may differ between cell types, perhaps reflecting differences in the transcription programs, nuclear organisation or chromatin structure, and epigenetic characteristics of the differently differentiated cells (reviewed in Rhind and Gilbert, 2013; Jackson *et al.*, 2012; Mechali, 2010). Indeed, it was shown that the timing programme of origin activation is highly stable and conserved between several cell lines of a single cell type, but differ between cell types, to the point that the replication timing profiles can be used for the identification of the different cell types (Ryba *et al.*, 2010).

Recently, both TbORC1/CDC6-binding sites and active origins of replication have been mapped in *T. brucei* PCF cells (Tiengwe *et al.*, 2012a). This genome-wide analysis revealed that, like in other eukaryotes, the amount of initiator (TbORC1/CDC6) binding sites, assumed to be potential origins, outnumbered the detectable active origins. TbORC1/CDC6-binding sites localised with high density to the subtelomeric, *VSG*-rich regions of each chromosome, where no origins were mapped. In the chromosome cores, TbORC1/CDC6 was more dispersed and localised virtually exclusively to the boundaries of the polycistronic transcription units, at the strand-switch regions (SSRs) (Tiengwe *et al.*, 2012a). Of the sites detected in the chromosome core, around 92% of TbORC1/CDC6-binding sites localised to transcription start sites, which had been previously inferred by localisation of histone H4 acetylated at lysine 10 (H4K10ac) (Siegel *et al.*, 2009), while the remaining localised to transcription termination sites. This, perhaps even more than in other eukaryotes, suggested that there is a pronounced association between DNA replication and transcription (Tiengwe *et al.*, 2012a), possibly functional, since RNAi of TbORC1/CDC6 altered mRNA abundance at

many of the mapped binding sites. Active origins were mapped by marker frequency analysis (MFA, detailed in the next section) coupled with next generation sequencing (MFA-seq) using cells in early-mid S phase and G2 phase (Tiengwe *et al.*, 2012a). This analysis identified 42 well-spaced, broad peaks of varying height, representing one origin per approximately 600 Kbp, all within the chromosomes cores and localising to TbORC1/CDC6-binding sites at the SSRs (Tiengwe *et al.*, 2012a). The different heights of the active origins suggested an activation timing programme of replication origins in *T. brucei*, with higher and lower peaks denoting early and late firing origins, respectively. Indeed, the MFA-seq peaks with highest amplitude, and thus earliest replicating, co-localised with mapped centromeres (chromosomes 1 to 8) (Obado *et al.*, 2007), suggesting these features are amongst the earliest replicating, as reported in other eukaryotes, like fission yeast (Kim *et al.*, 2003). Overall, *T. brucei* seems to possess a very low density of active origins when compared to other eukaryotes, with origin number correlating to chromosome size (Tiengwe *et al.*, 2012a).

In light of the differing pattern of origin usage in different cell types and during development in other eukaryotes, it is reasonable to ask if the putative timing programme of origin activation seen in PCF *T. brucei* cells is conserved in BSF cells. These distinct life cycle stages of *T. brucei* display major changes in gene expression associated with, for instance, their distinct cell morphology, cell surface protein expression, and metabolic activity (as mentioned in the introduction to this chapter). Many of these changes may be directed by alterations in nuclear architecture and chromatin structure (reviewed in Daniels *et al.*, 2010; Schlimme *et al.*, 1993; Rout and Field, 2001), by changes in histone modifications (reviewed in Figueiredo *et al.*, 2009; Siegel *et al.*, 2009; Maree and Patterton, 2014), and by the presence (BSF cells) or absence (PCF cells) of base J (Borst and Sabatini, 2008; van Leeuwen *et al.*, 1997; Reynolds *et al.*, 2014). Nonetheless, differences in gene expression between BSF and PCF cells (Siegel *et al.*, 2010; Jensen *et al.*, 2009; Jensen *et al.*, 2014; Brems *et al.*, 2005; Urbaniak *et al.*, 2012; Urbaniak *et al.*, 2013), unlike most of other eukaryotes, appear to not be a result of changes in the transcription regulation *per se* (reviewed in Clayton, 2002), but instead, a consequence of differential regulation at the post-transcriptional level (Clayton, 2013; Kramer, 2012; Siegel *et al.*, 2011), as a consequence of their unusual genome organisation in large

polycistronic transcription units (El-Sayed *et al.*, 2005a; El-Sayed *et al.*, 2005b; Berriman *et al.*, 2005; Jackson, 2014; Daniels *et al.*, 2010). Moreover, it has been recently shown that the DNA replication rate in BSF cells is 15% faster than PCF cells (4.4 Kbp/min versus 3.7 Kbp/min, respectively) (Calderano *et al.*, 2015), and whether this has an effect on, or results from, the origins that are activated or their activation timing during the S phase, it is not known.

In consideration of these differences between the two life cycle stages, and in order to investigate origin flexibility in *T. brucei*, the origins of replication in BSF cells were mapped and compared against the ones in PCF cells, as detailed in the next few sections.

4.5.1 Mapping origins of DNA replication: the concept

Origins of replication and TbORC1/CDC6 localisation have to date only been mapped in *T. brucei* PCF cells of strain TREU 927 (Tiengwe *et al.*, 2012a), the reference genome strain (Berriman *et al.*, 2005). Here, BSF cells from strain Lister 427 were used. In order to avoid potential differences in origin usage between the TREU 927 and Lister 427 strains, and allow direct comparison of origin usage between the two life cycle stages, replication origins were mapped in both PCF and BSF Lister 427 cells. Briefly, both cell types were cultured as described in the materials and methods Chapter 2, section 2.3, and collected for analysis when at a concentration corresponding to mid-late log phase (~1 x 10^7 cells.ml^{-1} for PCF cells, and ~1 x 10^6 cells.ml^{-1} for BSF cells). A total of 1 x 10^9 PCF cells and 3 x 10^8 BSF cells were collected and fixed, respectively, in 70% methanol and 1% formaldehyde (Tiengwe *et al.*, 2012a), as described in the materials and methods Chapter 2, section 2.7.1 and 2.7.2. The fixed cells were then stained with propidium iodide (PI), and sorted into early S, late S and G2/M phases by FACS, using a BD FACSAria I™ cell sorter (BD Biosystems), according to their DNA content. A schematic representation of the gates used is shown in Figure 4.18 A (BSF cells), and B (PCF cells). Each fraction was collected directly into lysis solution, and gDNA was purified and sent to Eurofins Genomics (Germany), where a DNA library of each sample was prepared with the TruSeq® DNA Sample Preparation kit (Illumina), and sequenced using the Illumina HiSeq paired-end 100 bp sequencing system (Illumina). For sequencing, the samples were multiplexed, with each of the early S, late S, and G2 phase libraries, both

from BSF and PCF, processed in the same run. The resulting data was then analysed by MFA using a script similar to the one described in (Tiengwe *et al.*, 2012a), and adapted by Dr Nicholas J. Dickens (WTCMP[n] Bioinformatics team leader, University of Glasgow). The general process is detailed in the materials and methods Chapter 2, section 2.8.1.2, represented in a simplified fashion in Figure 4.19, and the script used is shown in the appendices, section 7.8. Briefly, the sequence reads from each sample were aligned to the Lister 427 reference genome, retrieved from TriTrypDB v8.0. The Lister 427 genome also contains the sequences for the telomeric BSF *VSG* expression sites (BES) (Hertz-Fowler *et al.*, 2008), absent from the TREU 927 reference genome. For each sample, the frequency of reads (coverage) was assessed per base pair (each base pair acts as a marker). The genome was then "broken down" in 2.5 Kbp sections (bins) and the median of the frequency of reads per bin was assessed, and then used to calculate the ratios between the early S phase, or late S phase, versus G2 phase data that, ultimately, were used to generate the graphical representations shown in Figure 4.20. In this type of analysis, MFA coupled with deep sequencing (MFA-seq), a non-replicating sample, either G1 or G2 phases (though G2 phase is most commonly used (Muller *et al.*, 2014; Muller and Nieduszynski, 2012), and appears to confer more robust results – Dr Nick J. Dickens, personal communication), is used for normalisation, since both should have an even read coverage across the whole genome, thus excluding 'noise' in, for instance, regions of the genome where there is poor coverage (e.g. with high CG content). The same approach has been used in yeast (Muller *et al.*, 2014; Muller and Nieduszynski, 2012), where it is termed Sort-seq, and in archaea (Hawkins *et al.*, 2013), though the strategies for isolating replicating and non-replicating DNA vary. In the MFA-seq approach, the median of read depth (within a bin) of the S phase sample is compared to the general coverage (genome-wide average number of reads) of that sample, and is then compared with the median of reads (within the same bin) of the G2 phase sample (also compared to the genome-wide average number of reads). The S/G2 ratios in each bin, thus derived, are then plotted relative to the genome. Across a chromosome, peaks occur in regions where S phase reads occur more frequently than in the rest of the genome (relative to G2 phase). Because replication occurs bidirectionally, the S/G2 ratios increase as they approach an origin – hence why the plots form

[n] WTCMP – Wellcome Trust Centre for Molecular Parasitology.

peaks, where the highest point localises the origin. In contrast, “valleys” represent broad zones of termination, where two opposite replication forks converge (de Moura *et al.*, 2010; Muller *et al.*, 2014).

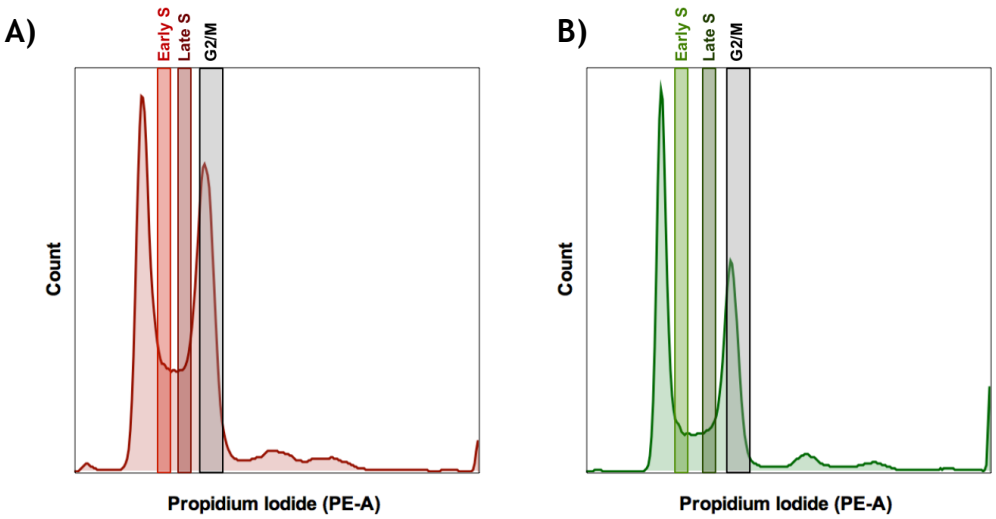


Figure 4.18. Sorting of asynchronous cultures into early S, late S and G2 phases by FACS. Histograms depicting the cell cycle profile of a mid-log phase BSF (A) and PCF (B) cell cultures, reflecting the fluorescent signal emitted by the staining of DNA with PI. Different rectangles represent the gates defined to purify the different cell cycle populations.

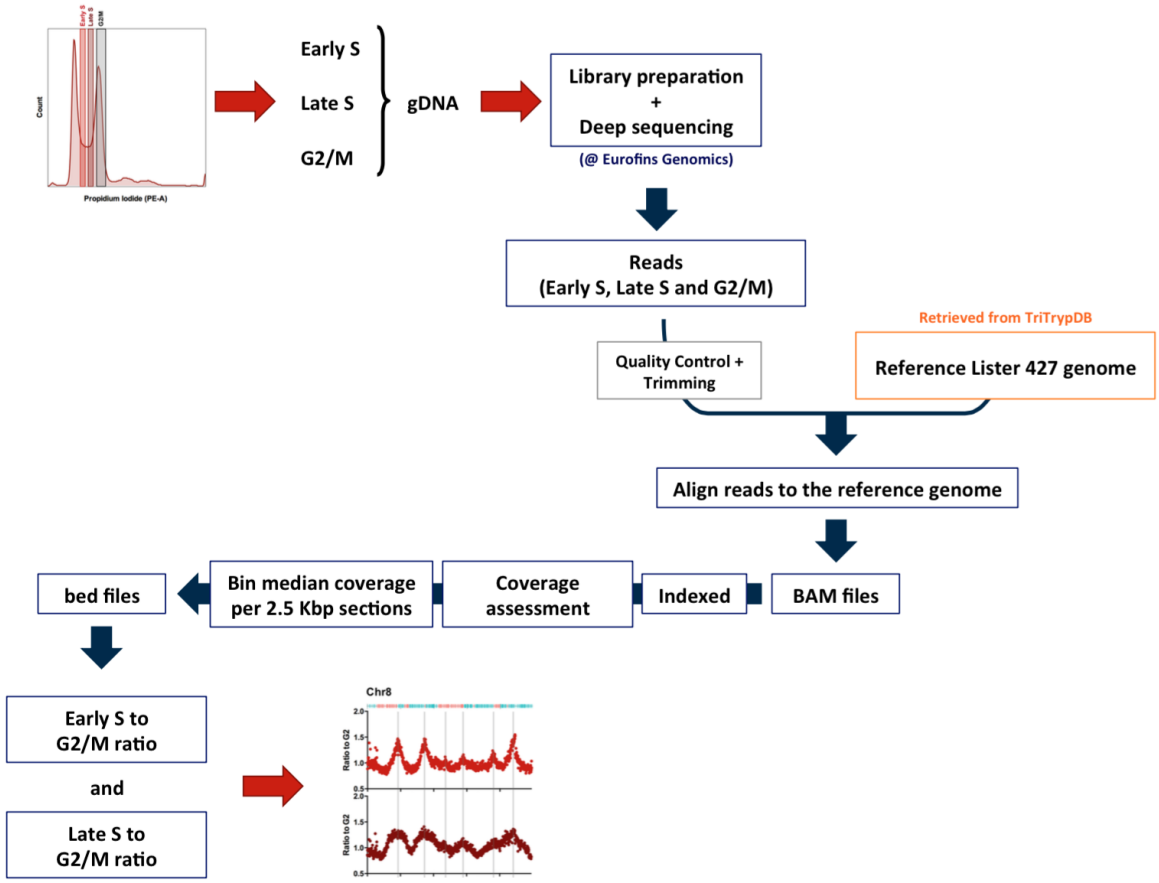


Figure 4.19. MFA-seq workflow.
Simplified schematic representation of the workflow followed to obtain the MFA-seq graphs. Details are in the main text and materials and methods, while the coding script is shown in the appendices.

4.5.2 Mapping the origins of DNA replication in *T. brucei* BSF cells

MFA-seq results for the eleven megabase chromosomes of *T. brucei* from both BSF and PCF cells, comparing early S with G2 phase (BSF cells light red, PCF cells light green), as well as late S with G2 phase (BSF cells dark red, PCF cells, dark green), are shown in Figure 4.20. It is important to bare in mind that MFA-seq analysis allows the identification of origins of replication that are activated at the population level – most likely it is able to detect mainly constitutive origins and therefore, origins that are activated less frequently at the population level (for instance, flexible origins, which may be activated differently from cell to cell in the population), might not be detected. Consequently, the analysis and conclusions here drawn relate to the detected active origins, and thus does not take in consideration potential differences between the two life cycle stages in respect to origins that might be activated differently between individual cells. The overall MFA-seq pattern of peak location was strikingly similar when comparing early S phase BSF and PCF cells, and when comparing late S phase BSF and PCF cells (Figure 4.20; chromosome 8 is shown in more detail in Figure 4.21). These data suggest that origin location within the chromosomes core regions is unaltered between the two life cycle stages, thus suggesting that following a distinct cellular differentiation programme does not lead to, at least, drastic changes in the overall subset of origins that are activated in S phase. Late S MFA-seq has not been reported for *T. brucei*, though it was inferred that the number of mapped origins (counted as 42) might be underestimated by the early-mid S phase sampling, and that the actual origin number may be at least the double of that (Tiengwe *et al.*, 2012a). However, this prediction appears to be inaccurate, as the peaks detected in the late S samples were merely wider than the early S peaks, and no clear, further peaks were seen; indeed, in several locations early S peaks appeared to have merged as replication forks converged (Figure 4.20).

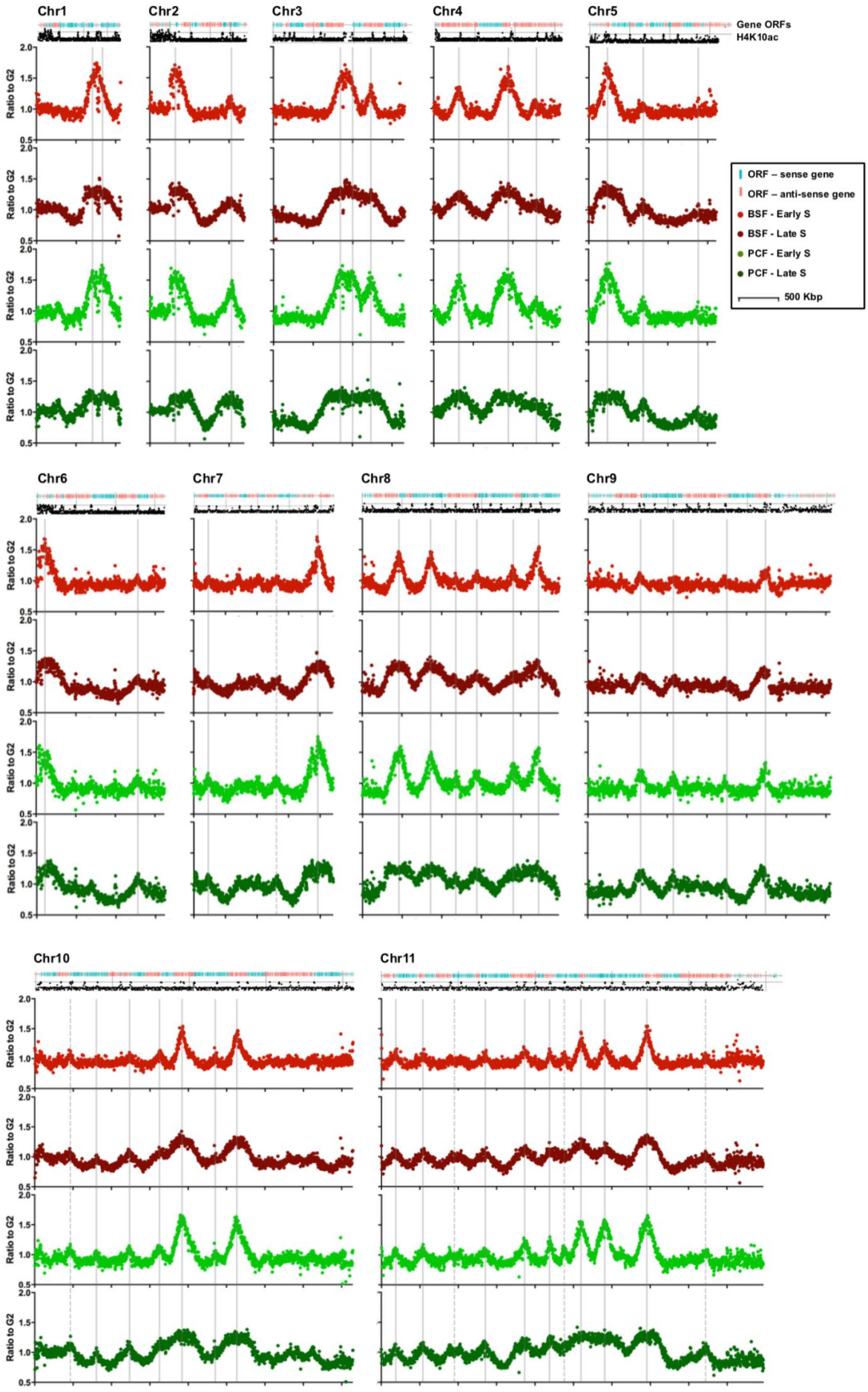


Figure 4.20. Mapping origins of replication in *T. brucei* BSF and PCF cells.

Each set of four graphs (data represented in light red, dark red, light green, and dark green dots) shows the distribution of the active origins of replication in each of the megabase chromosomes (depicted as Chr1 to Chr11), assessed by MFA-seq, as in (Tiengwe *et al.*, 2012a). At the top of each set of graphs a track represents the genes in the chromosome: in blue the open reading frames (ORFs) are transcribed from the left to the right, and in red the genes are transcribed from right to left. Below, a track is shown depicting histone H4K10ac-enriched sites, as described in (Siegel *et al.*, 2009). Dr Nicholas J. Dickens generated both tracks using the R package software and data available for the TREU 927 reference genome. Each of the four graphs shows the ratio between the coverage (read-depth) between early S phase and G2 phase samples, or late S phase and G2 phase samples, where each point (light red, light green, dark red and dark green) represents the median S/G2 ratio (y-axis) per 2.5 Kbp section across the chromosome (x-axis). Each x-axis gap represents 500 Kbp intervals. The y-axis scale is the same for all graphs, but legend is only shown in those on the far left. All graphs are scaled according to chromosome size. Light red graph shows the results for the BSF early S sample, while dark red represents the data from late S phase. PCF data is shown in light green, early S, and dark green, late S. Full grey lines across the four graphs represent the origins identified previously (Tiengwe *et al.*, 2012a), while grey dashed lines highlight extra replication origins observed in this study. Individual graphs were generated using GraphPad v6.0.

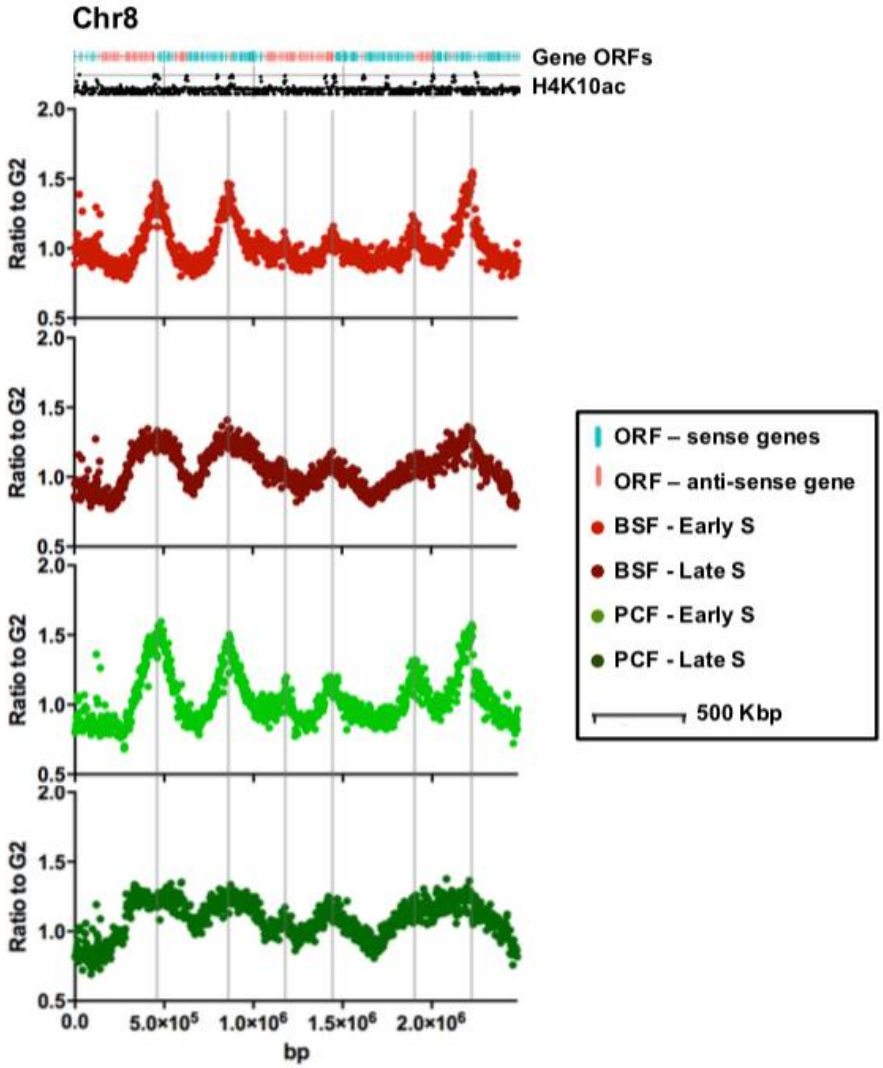


Figure 4.21. Chromosome 8 in detail.
Representation of chromosome 8 in more detail, as shown in Figure 4.20, enlarged. Description as in Figure 4.20.

The identified origins here coincided with most of the previously mapped origins (Tiengwe *et al.*, 2012a) (Figure 4.20, grey lines), with minor exceptions. In chromosome 2 (Figure 4.20, Chr2), in both the early S BSF and PCF samples, a single peak was observed around 0.325 Mbp, co-localising with the centromere, while in (Tiengwe *et al.*, 2012a) two peaks were reported in the same region. Only in the late S phase PCF samples shown here is there evidence for two peaks, corresponding with the origins predicted by (Tiengwe *et al.*, 2012a). It is possible that greater read coverage in the TREU 927 PCF analysis, or differences in S phase sampling, as well as data representation, result in slightly decreased resolution in the present work. Alternatively, this might represent a difference in centromere annotation or structure between the TREU 927 and Lister 427 genomes. Likewise, two closely located origins were previously predicted in chromosomes 1 and 3 (around 0.75 and 1.0 Mbp, respectively) (Tiengwe *et al.*, 2012a). In the present study, the same two peaks were observed in the PCF early S sample, while only a single clearly defined peak was observed in the BSF early S sample (Figure 4.20). Analysis of the late S samples appears to suggest that the PCF peaks have broadened more than the respective BSF ones, perhaps supporting the idea of PCF cells having an extra origin in chromosomes 1 and 3. Nevertheless, this may merely suggest that delineation of closely separated peaks is dependent on sample collection during the FACS sorting, and further analysis will be needed to infer whether there is a real difference between BSF and PCF cells in these chromosomes.

Five origins were observed in the present data, both in the BSF and PCF samples (Figure 4.20, grey dashed lines), which were not predicted previously (Tiengwe *et al.*, 2012a). All these were “weak” origins, with low peak heights and it is therefore possible that they were observed here due to the more compressed graphical representation of the data compared to the representation in (Tiengwe *et al.*, 2012a), rather than activation of origins in Lister 427 that are not used in strain TREU 927. Nonetheless, all these origins were present in the larger chromosomes: one in chromosomes 7 and 10, and three in chromosome 11 (Figure 4.20, grey dashed lines). Though TbORC1/CDC6 and histone H4K10Ac binding sites have not been mapped to the Lister 427 reference genome, using the mapping made to the TREU 927 genome (Tiengwe *et al.*, 2012a; Siegel *et al.*, 2009), it is clear that these five origins localise to both TbORC1/CDC6 and

histone H4K10ac enriched sites at the boundaries of the polycistronic transcription units (Figure 4.20). This adds weight to the suggestion that these are active origins, and conforms with the proposal that origins localise only at the boundaries of the transcription units: in chromosomes 7 and 10, the two origins localised to divergent strand switch regions (dSSRs), while the same was observed for one of the origins in chromosome 11; the remaining two origins localised to head-to-tail SSRs (Figure 4.20, grey dashed lines).

MFA-seq data from (Tiengwe *et al.*, 2012a) clearly showed that peak height varied across chromosomes, suggesting an activation timing programme of replication origins: peaks with the highest amplitude being origins replicating earliest in S phase, and those with smaller amplitude corresponding with origins replicating later in S phase. The overall pattern of MFA-seq peak height shown here was strikingly similar to the one shown in (Tiengwe *et al.*, 2012a) for each chromosome, as exemplified by the highest peaks in each of chromosomes 1-8 co-localising with the centromere. Moreover, comparison between the BSF and PCF MFA-seq data suggests that the replication timing programme is conserved between the BSF and PCF life cycle stages, and minor differences in peak height between BSF and PCF samples are most likely due to sampling differences during the FACS cell sorting of the two cell types.

Taken as a whole, the above MFA-seq analysis suggests striking rigidity in the coordination of nuclear DNA replication in *T. brucei*. First, the same origin locations were detected in Lister 427 PCF *T. brucei* cells that had been seen previously in TREU 927 PCF cells, with the exception of five very weak origins. Second, the same origins locations were seen in BSF and PCF cells. Third, the timing of origin firing, as assessed by MFA-seq peak height, was roughly the same in the two life cycle stages and in the two strains. Finally, all peaks were mapped to the core of the chromosomes, with the exception of the strongest origin in chromosome 6, which may be subtelomeric.

4.5.3 DNA replication at the bloodstream *VSG* expression sites

T. brucei BSF cells evade the mammalian host immune response by antigenic variation, a process through which the parasite alters ("switches") the variant surface glycoprotein (VSG) that covers the parasite's surface as a coat (reviewed in McCulloch *et al.*, 2014; Glover *et al.*, 2013b), protecting it from the immune responses targeting the parasite's previously expressed VSG. Although the parasite possesses a huge repertoire of *VSG* genes and pseudogenes (Marcello and Barry, 2007), only one VSG is expressed at a given time in the cell. In the BSF cell, the expressed VSG is found in one of ~15-25 telomeric BSF transcription sites known as the bloodstream *VSG* expression sites (BES) (Hertz-Fowler *et al.*, 2008). Although their sizes vary, the BES share a similar generic architecture (Figure 4.22), with a telomere-distal RNA polymerase I promoter, an array of expression site associated genes (ESAGs) and pseudogenes (including *VSG* pseudogenes), a block of 70 bp repeats and, finally, the functional *VSG*, localised close to the telomere repeats (Hertz-Fowler *et al.*, 2008) (reviewed in Glover *et al.*, 2013b; McCulloch *et al.*, 2014). Thus all BES, including the singular one from which the VSG coat is expressed, are located adjacent to the telomeres of the megabase and intermediate chromosomes. Unlike most of the genome, which is transcribed by RNA polymerase (RNA Pol) II, the active BES is transcribed by RNA Pol I (Gunzl *et al.*, 2003; Gunzl *et al.*, 2014), which is recruited to the BES promoter, and localises in a discrete region of the nucleus known as the expression site body (Chaves *et al.*, 1998; Navarro and Gull, 2001). In contrast, inactive BES localise to the nuclear peripheral heterochromatin (reviewed in Glover *et al.*, 2013b). Although it is present throughout the cell cycle in BSF cells, the expression site body has never been observed in PCF cells (reviewed in Glover *et al.*, 2013b), being apparently lost during differentiation, with the active BES relocating to the nuclear periphery of the cells (Landeira and Navarro, 2007). Like the VSG in BSF cells, procyclins, the surface antigens of PCF cells, are transcribed by RNA Pol I from a promoter at two *procyclin* loci, which also localise to a specific region of the nucleus: the nucleolar periphery (Landeira and Navarro, 2007). The tightly regulated expression of these life cycle specific genes by RNA Pol I most likely allows the high expression of the resultant cell surface proteins (reviewed in Maree and Patterton, 2014; McCulloch *et al.*, 2014; Gunzl *et al.*, 2014). Generally, present data suggests

that chromatin status (for instance, histone modifications and variants) (Narayanan and Rudenko, 2013; Kramer, 2012; Rudenko, 2010) and physical localisation of the BES play an important role in maintaining all BES as silent in PCF cells (Pandya *et al.*, 2013), as well as allowing the expression of only one BES at a time (active BES) in BSF cells, in addition to regulating the molecular mechanisms (not discussed here) that drive antigenic variation (reviewed in Glover *et al.*, 2013b).

The BES repertoire of *T. brucei* Lister 427 has been sequenced after directed cloning, and appears to be composed of 14 distinct BES (Figure 4.22) (Hertz-Fowler *et al.*, 2008; Becker *et al.*, 2004). The sequences of the clones used to identify the individual BES have been annotated as contigs in the available Lister 427 reference genome, and although these have not been allocated to specific chromosomes, it is known that a subset is present in intermediate chromosomes (BES 2, 4, 10, 11, 13, and 17), while the rest are located in the telomeres of the megabase chromosomes, with BES 1 localising to the right-hand side telomere of chromosome 6a (Hertz-Fowler *et al.*, 2008; Glover and Horn, 2014). BES sequences are not available for the TREU 927 reference genome, thus analysis of their replication was not possible from previous MFA-seq data (Tiengwe *et al.*, 2012a). Given that a single active BES is highly transcribed at a time in BSF cells, but all BES are silenced in PCF cells (reviewed in Glover *et al.*, 2013b; McCulloch *et al.*, 2014), we took advantage of the Lister 427 BES sequences and used the MFA-seq data to ask about replication dynamics at these loci in the two life cycle stages.

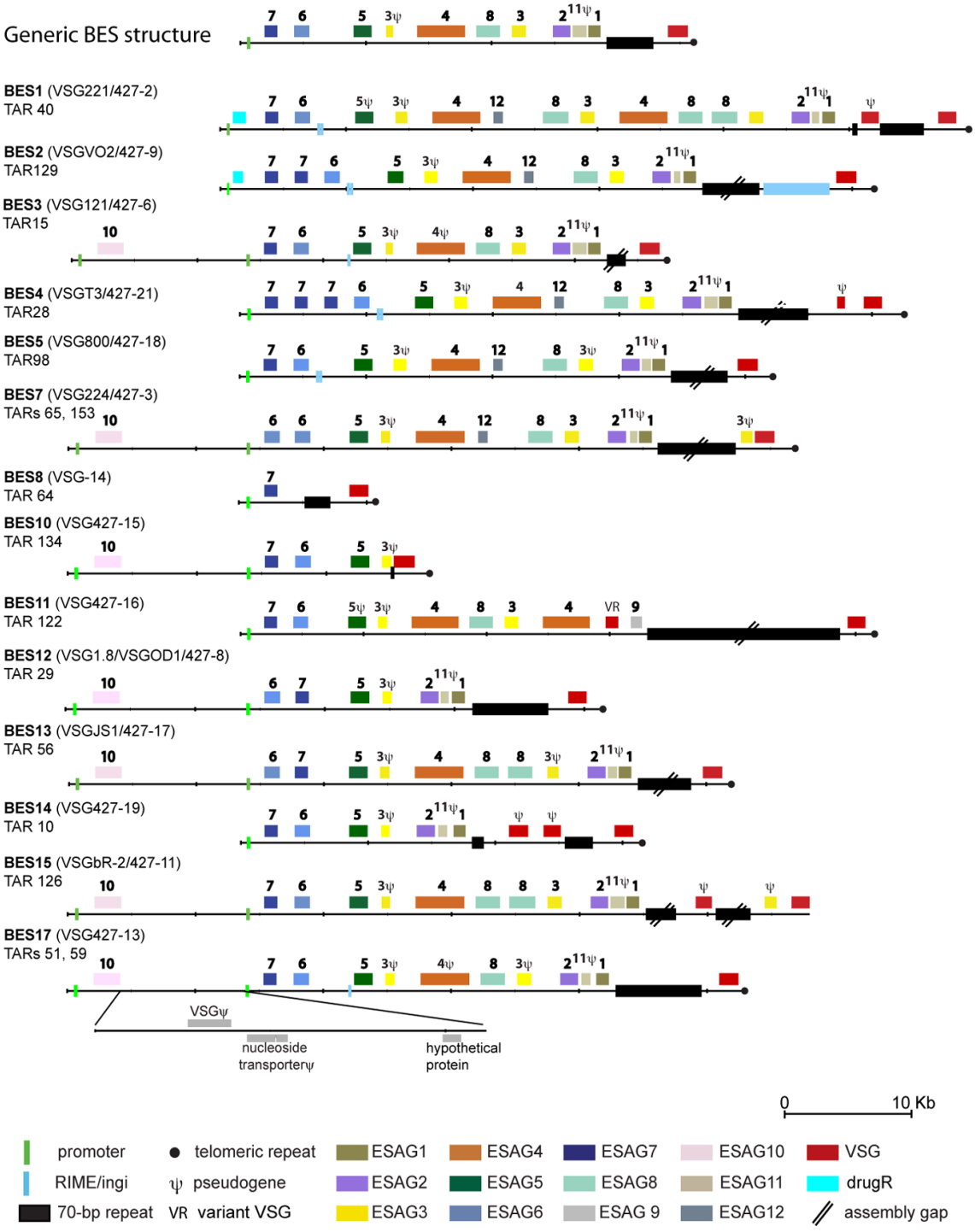


Figure 4.22. Schematic representation of the BES described in *T. brucei* strain Lister 427. Diagrams representing the different described bloodstream expression sites (BES) identified and annotated to the Lister 427 genome. The top diagram represents a model of the generic structure of a BES. TAR refers to transformation-associated recombination clones, with the numbers referring to the clones sequenced in (Hertz-Fowler *et al.*, 2008). The *VSG*s present in each BES are also highlighted. The figure is shown to facilitate the interpretation of the data shown below in Figure 4.23 only. Figure reproduced in full from (Hertz-Fowler *et al.*, 2008), © 2008 Hertz-Fowler et al., all rights reserved.

The MFA-seq data was processed as described for the compiled genome (previous section), and mapped to the 16 available contigs representing the 14 different BES (Figure 4.23). Note that there are two sequences for BES 7 (ϕ – TAR[o] 65; ϕϕ – TAR 153) and BES 17 (ϕ – TAR 51; ϕϕ – TAR 59), each of which is shown in separate (Figure 4.23). In these data, peaks cannot be discerned, as the sizes of the BESs are smaller than the distance covered by the replication forks at most origins in both early and late S phase. Thus, at the higher resolution used here, the MFA-seq mapping is seen as multiple discrete points, corresponding to the S/G2 phases ratio in each bin across the BES. For all but one of the BES, there are no clear differences between the MFA-seq mapping in the BSF and PCF cells, either for early or late S phase (Figure 4.23). Indeed, there was no evidence that the BES had been replicated, even in the BSF and PCF late S samples (Figure 4.23, dark red and dark green, respectively), since there was no consistent enrichment in S phase reads relative to G2 (S/G2 ratios were all ~1.0). These data suggest that these BES sites, which localise to the telomeres of the chromosomes, are replicated very late in S phase, similar to the telomeres in other eukaryotes (reviewed in Masai *et al.*, 2010; Rhind and Gilbert, 2013).

BES 1 appears to be a striking exception to the above trend. In this contig (Figure 4.23, within red dashed box), BSF S phase reads (both early and late) were markedly enriched relative to G2 (S/G2 ratio >1). In contrast, the MFA-seq for BES 1 did not show such enrichment in the PCF cells samples (again in either early or late S), with the S/G2 ratios being comparable with the ones in all other BES, at ~1.0. This suggests that BES 1, unlike all others, was replicated early in S phase in BSF cells. In contrast, there was no evidence for BES1 replication in PCF cells, suggesting it may only be replicated in this life cycle stage very late in S phase, like the other BES (Figure 4.23). Interestingly, BES 1 differs from all other BES because it is the actively transcribed one, encoding *VSG 221* in the BSF parasites used in this study: strain Lister 427, derived from MITat 1.2 clone 221a (the same used in Hertz-Fowler *et al.*, 2008). To verify this, expression of VSG 221 on the surface of the BSF cells, but not in PCF cells, was confirmed by immunofluorescence using anti-VSG 221 antiserum (Figure 4.24, A), as described previously (Glover *et al.*, 2013a). In parallel, expression of EP procyclin in the

[o] Transformation-associated recombination (TAR) clones, numbered as in Hertz-Fowler et al., 2008.

PCF cells, and its absence from the BSF cells surface, was confirmed with anti-EP procyclin antiserum (Figure 4.24, B), as described before (Richardson *et al.*, 1988). Both protocols are detailed in the materials and methods Chapter 2, sections 2.5.2 and 2.5.5. The observation that the only clear difference in replication timing between BSF and PCF cells may lay in the only locus that is transcriptionally active in BSF cells and is silenced in PCF cells, reinforces the idea that DNA replication and transcription are functionally related in *T. brucei* (Tiengwe *et al.*, 2012a).

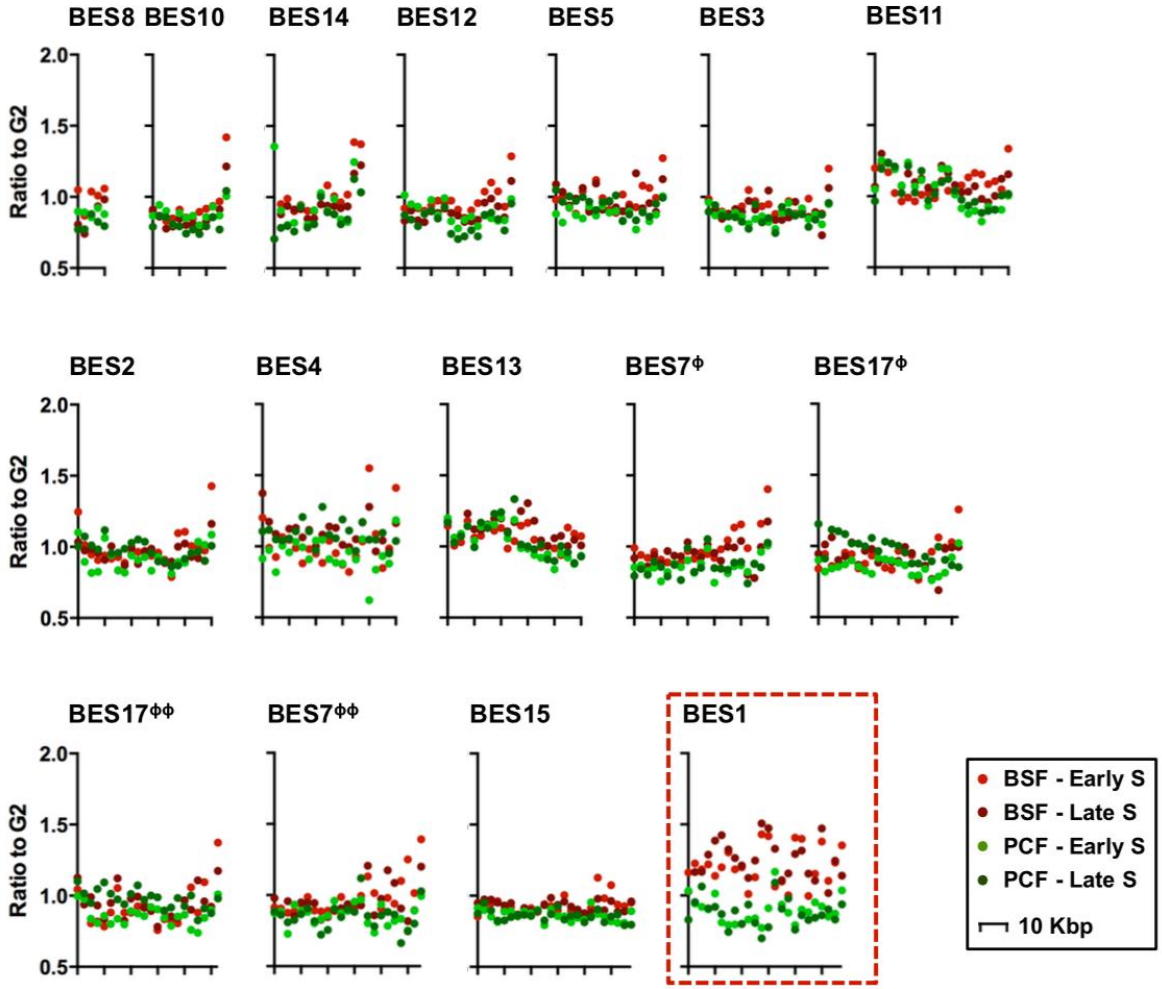


Figure 4.23. DNA replication mapping at the bloodstream expression sites (BES).
The BES sequenced by (Hertz-Fowler *et al.*, 2008) and annotated as contigs to the Lister 427 genome (available on TriTrypDB), were used to map the MFA-seq data. Each BES is numbered according to (Hertz-Fowler *et al.*, 2008), and as shown in Figure 4.22; note that two contigs, referring to two independently sequenced TAR clones (shown in Figure 4.22) (Hertz-Fowler *et al.*, 2008), are duplicates of two BES: BES 7 (ϕ – TAR 65; ϕϕ – TAR 153) and BES 17 (ϕ – TAR 51; ϕϕ – TAR 59). Like in Figure 4.20, the ratio between the coverage (read-depth) in early S phase and G2 phase samples, or late S phase and G2 phase samples, is plotted, where each point represents the median S/G2 ratio (y-axis) per 2.5 Kbp section across the BES (x-axis). The size of each BES is shown on each x-axis in 10 Kbp intervals. The y-axis scale is the same for all graphs, but the legend is only shown on the ones at the far left. All graphs are scaled according to each BES size, in which each x-axis gap represents 10 Kbp intervals. Graphs are represented by order of size. BSF early S data is represented as light red, BSF late S as dark red, PCF early S as light green, and PCF late S as dark green. The red dashed box highlights BES 1. Individual graphs were generated in GraphPad v6.0.

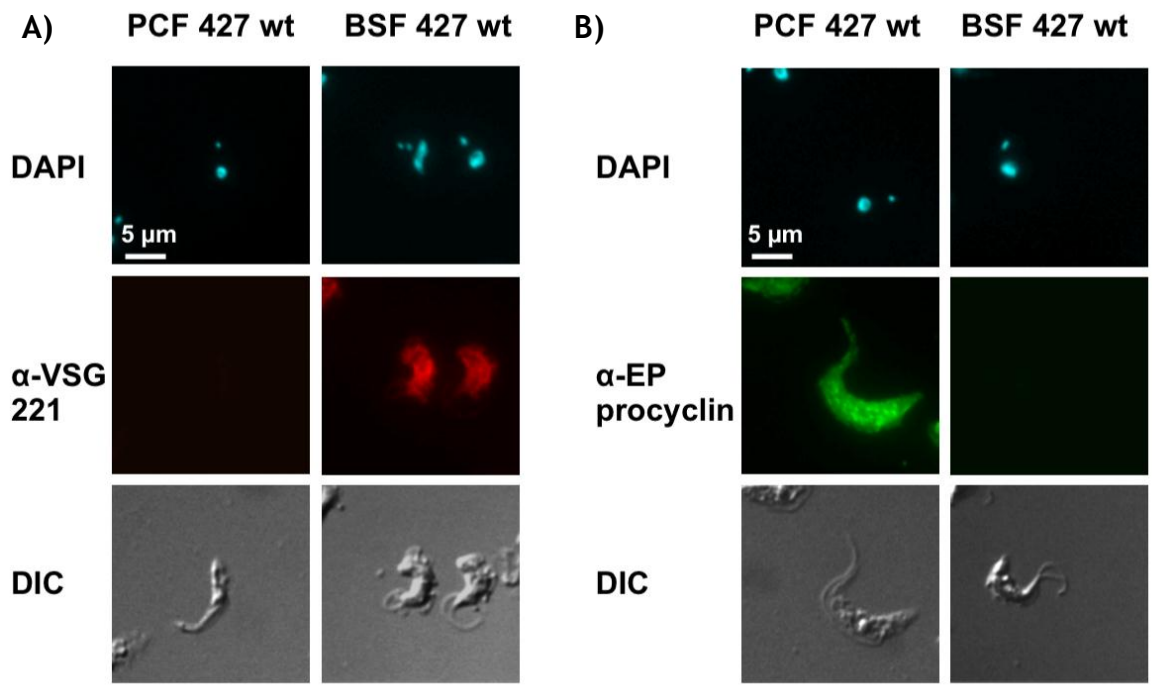


Figure 4.24. Anti-VSG 221 and anti-procyclin immunofluorescence.
A) Immunofluorescence of PCF strain Lister 427 cells (left panels) and BSF strain Lister 427 cells (right panels) with anti-VSG 221 antiserum (middle panel, red), for the detection of the expression, at the cell surface, of VSG 221 (as described in Glover *et al.*, 2013a). B) Immunofluorescence of PCF strain Lister 427 cells (left panels) and BSF strain Lister 427 cells (right panels) with anti-procyclin antiserum (central panel, green), for the detection of the expression, at the cell surface, of EP procyclin (as described in Richardson *et al.*, 1988). In both A) and B), top panels show the cells stained with DAPI, while the bottom panel shows the cells outline by DIC. Images were acquired with the Axioskop 2 imaging system. Scale bar represents 5 µm.

It is possible that the above result might be an artefact due to the small size of the BES sequences, of between 10 Kbp (BES 8) and 58 Kbp (BES1) (Figure 4.22 and Figure 4.23), with limited numbers of sequence bins mapping within them. In addition, mapping may be compromised by the relative sequence similarity between BES. In order to test this, mapping simulations are being carried out, by analysing read mapping in random sections of the parasite's genome of similar sizes to the various BES. So far, analysis performed by Samantha Campbell (PhD student, supervision by Dr Nicholas J. Dickens) suggests that the only other sites with the S/G2 ratio pattern seen for BES1 in BSF cells correspond with sections localising to origins mapped in the core regions of the chromosomes (data not shown). This supports the idea that the BES 1 region is being replicated in BSF cells, though it is not clear whether an origin is present in BES 1 or whether this region is being passively replicated from an adjacent origin in BSF cells but not in PCF cells, and further analysis is needed. In order to subject the BES mapping data to further scrutiny, the S/G2 ratio in each mapped bin was compared for each BES and in each set of MFA-seq data: BSF early S, BSF late S, PCF early S

and PCF late S (Figure 4.25). The median values and distributions were then compared between each other using the non-parametric Kruskal-Wallis test (un-matched), and p-value threshold of 0.05. A non-parametric test was preferred because the data distributions failed to pass normality tests (D'Agostino and Pearson normality test, Shapiro-Wilk normality test, and the Kolmogorov-Smirnov normality test). Although the analysis was performed by comparing the medians of all groups with one another, statistical significance is only represented in Figure 4.25 for the differences between BSF early S and PCF early S, and BSF late S and PCF late S, as the aim of this analysis was to assess differences between the two life cycle stages and not differences between early S and late S phases. These data confirm that BES1 S/G2 ratios in BSF early S significantly differ from PCF early S, as do BSF late S and PCF late S samples (both with a p-value < 0.0001; Figure 4.25, A). The median values of the combined S/G2 ratios also confirmed S-phase enrichment in BSF cells (ratios of 1.25 and 1.3 in early and late S, respectively), but no such enrichment in PCF cells (median G/2 ratios for both S phase samples <1.0). For no other BES was the same BSF S/G2 enrichment seen, and indeed in all cases the median S/G2 ratios for both BSF and PCF cells (both early and late S) were noticeably closer to 1.0 than for BES 1 BSF. Some significant differences were observed between the BSF and PCF median S/G2 values in five other BES (BES 7ϕ, BES 7$\phi\phi$, BES 15, BES 5 and BES 3), though in all cases with less significance than seen for BES1 (with the exception of BES15, and here only in the late S phase samples; Figure 4.25, A and B).

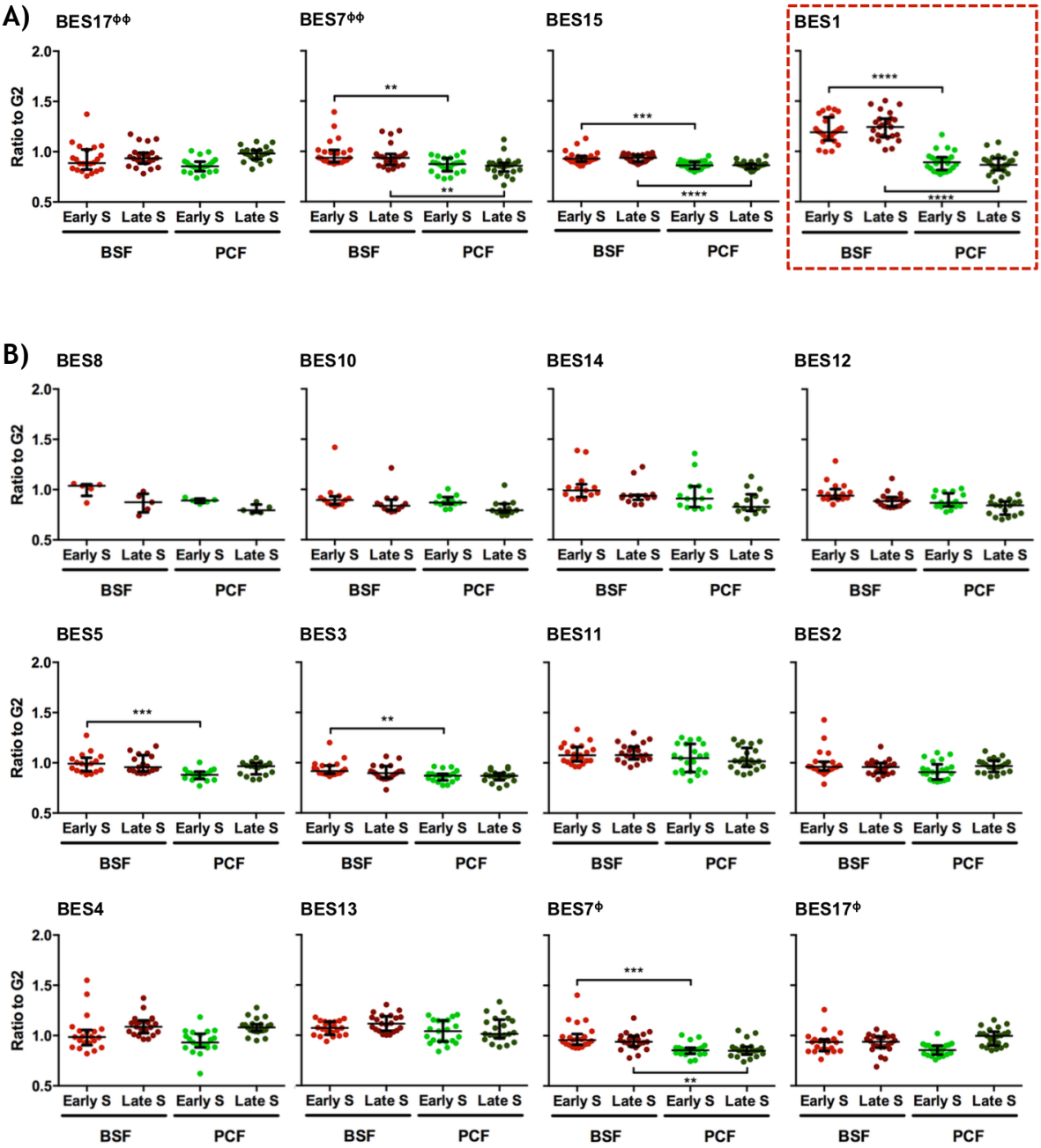


Figure 4.25. Median and scatter of MFA-seq S/G2 values in each individual BES.
The S/G2 values used to generate the graphs in Figure 4.24 were plotted per sample (BSF early S – light red, BSF late S – dark red, PCF early S – light green, and PCF late S – dark green), rather than by genomic location, for each BES (numbered as before). Horizontal bars (black) represent the median of the S/G2 values, and error bars the interquartile range. In order to infer statistical significance, the values were analysed with the non-parametric, unmatched, Kruskal-Wallis test. Statistical significance is only shown for differences between the BSF and PCF samples. Differences between early S and late S samples are not shown. (**) p-value < 0.01; (***) p-value < 0.001; (****) p-value < 0.0001. A) BES 17ϕϕ, BES 7ϕϕ and BES 15 are shown alongside BES 1, because these four BES are of roughly the same size. B) Shows all other BES.

To ask if locus-specific, stage-specific replication initiation differences are limited to the BES, the MFA-seq data around the two loci encoding the surface proteins in PCF cells, the procyclins, as well as the procyclin-associated genes (*PAG*s), was also examined. These loci are localised in chromosome 6 (from around 200 to 240 Kbp, including the EP procyclin 3 (*EP3*) and GPEET genes) and in chromosome 10 (from about 2505 to 2532 Kbp, including the *EP1* and *EP2* genes) (Haenni *et al.*, 2009; Haenni *et al.*, 2006). In both chromosomes, each locus is located in the slope of a strong origin peak (Figure 4.26, A). Zooming in on these regions did not reveal any clear differences between BSF and PCF cells, and analysis of the medians of the S/G2 ratios in the ~40 Kbp around the two loci (again comparing BSF early S, BSF late S, PCF early S and PCF late S) did not reveal any statistically significant differences (Figure 4.26, B), though visually the PCF late S/G2 ratio at the chromosome 10 locus might indicate enrichment, perhaps indicative of earlier replication in PCF cells (Figure 4.26, B). The expression of *procyclin* is apparently regulated (Acosta-Serrano *et al.*, 2001; Jensen *et al.*, 2009) via transcription attenuation and life cycle stage-dependent changes in mRNA stability due to properties of their 3' UTR sequences (Vanhamme *et al.*, 1995; Vassella *et al.*, 2001; Vassella *et al.*, 2000; Urwyler *et al.*, 2005). Whether this is comparable with the full activity of a singular BES in BSF cells, where one promoter may be more active than all others, is unclear. Nevertheless, it seems most likely that the two procyclin loci are passively replicated from an origin at a nearby SSR in PCF and BSF cells (Figure 4.26, A, and Figure 4.20), rather than facilitating DNA replication through acting as origins (Figure 4.23).

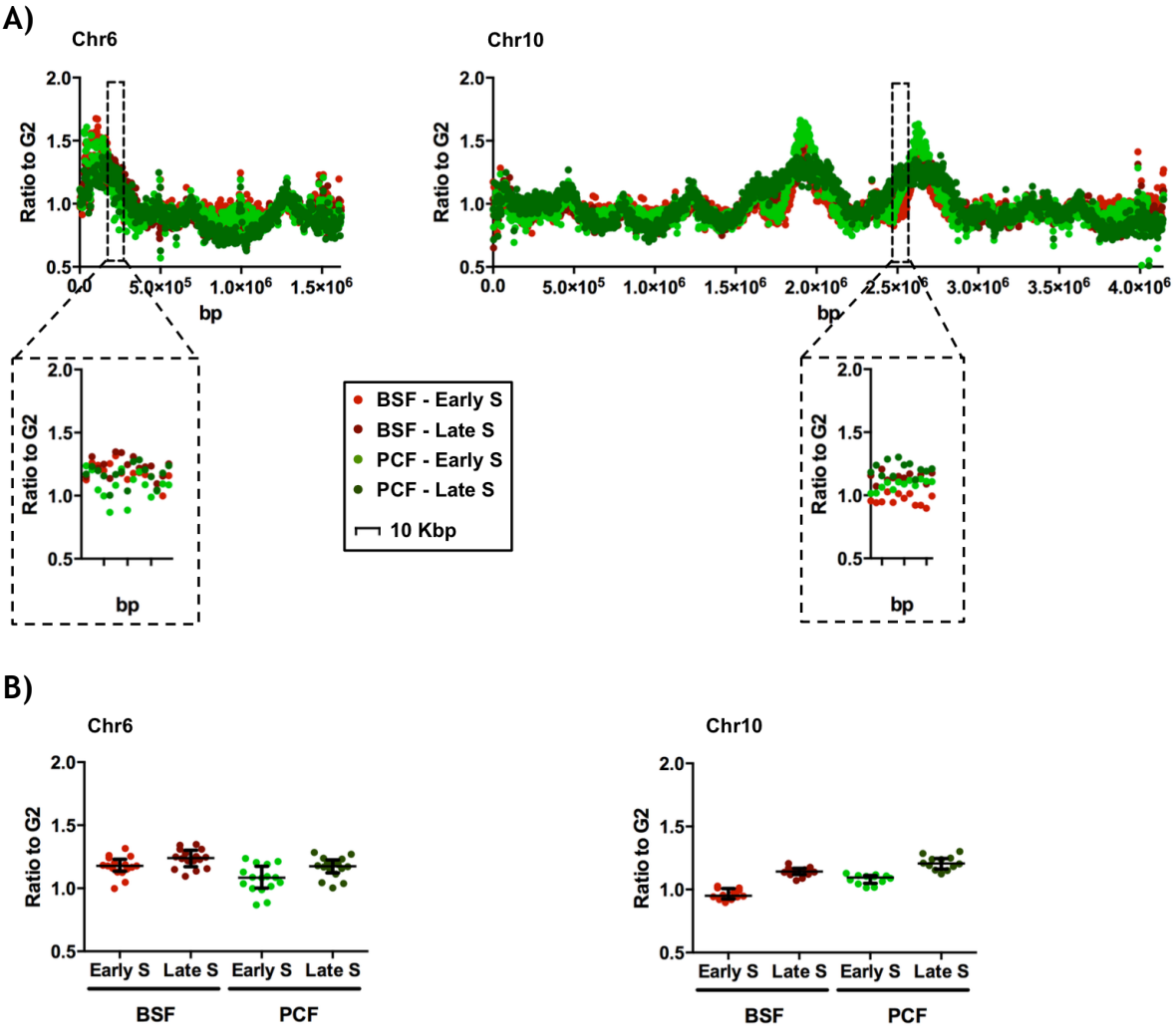


Figure 4.26. DNA replication at the *EP procyclin* loci.
The two *EP procyclin* loci are shown: chromosome 6 on the left, and chromosome 10 on the right. A) Chromosomes 6 and 10, containing the overlapped MFA-seq data from BSF early S, BSF late S, PCF early S and PCF late S. The dashed boxes mark the region of the chromosome where these loci are localised. Below, a magnified view is shown of these loci, with each x-axis interval representing 10 Kbp. B) The S/G2 values within the *EP procyclin* loci, "zoomed in" in A) were plotted per sample (BSF early S – light red, BSF late S – dark red, PCF early S – light green, and PCF late S – dark green), rather than by genomic location. Horizontal bars (black) represent the median of the values, and error bars the interquartile range. In order to infer statistical significance, the values were analysed with the non-parametric, unmatched, Kruskal-Wallis test. No statistically significant differences were observed.

4.6 Discussion

Most of the knowledge on nuclear DNA replication in *T. brucei* has been obtained from studies on the replicative midgut insect stage of the parasite's life cycle (Tiengwe *et al.*, 2012b; Tiengwe *et al.*, 2012a; Godoy *et al.*, 2009; Benmerzouga *et al.*, 2013). The few data available from studies in the replicative mammal stage of the parasite's life cycle, the BSF cells, suggest that the effects of TbORC1/CDC6 and some of its interacting factors depletion are distinct from the

ones observed in PCF cells (Benmerzouga *et al.*, 2013; Tiengwe *et al.*, 2012b). Although no differences are expected between the insect and the mammal life cycle stages, as the replication of nuclear DNA is a fundamental cellular process, it is possible that minor differences, perhaps related to non-replication roles, might exist between these two distinct life cycle stages of *T. brucei*, in which differences in cell cycle control have been described previously (reviewed in Hammarton, 2007; Li, 2012). In this chapter, the roles of TbORC1/CDC6 and interacting factors in DNA replication, as well as their cell cycle dynamics, were investigated in BSF cells. Moreover, the origins of DNA replication were mapped in BSF cells, as reports in other eukaryotes suggest that the cohort of origins activated is cell type-dependent, changing according to the cell's developmental differentiation program (reviewed in Mechali, 2010; Masai *et al.*, 2010; Leonard and Mechali, 2013; Jackson *et al.*, 2012; Rhind and Gilbert, 2013). Below, these results are discussed and compared to the findings obtained in PCF studies shown in Chapter 3.

4.6.1 TbORC1/CDC6, TbORC1B and TbORC4 are involved in nuclear DNA replication in BSF cells, while all factors localisation and cell cycle dynamics in BSF cells mimic the ones observed in PCF cells

To date, the reports on the effects of TbORC1/CDC6 downregulation by RNAi in BSF cells have shown that the resulting phenotype is distinct from the one obtained in PCF cells: multi-nucleated and multi-kinetoplast cells are generated, in contrast with the clear accumulation of enucleated cells (zoids) in PCF TbORC1/CDC6 RNAi-induced cells (Godoy *et al.*, 2009; Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013). As already discussed in Chapter 3, PCF cells appear to lack the mitosis to cytokinesis cell cycle checkpoint, which allows cell cycle progression to cytokinesis in the event of DNA replication or mitosis impairment, resulting in the generation of zoid cells (Ploubidou *et al.*, 1999; Hammarton *et al.*, 2003). As a result, it has been assumed that it is the lack of this checkpoint that results in the generation and accumulation of zoids when silencing of TbORC1/CDC6 is induced in PCF cells (Godoy *et al.*, 2009; Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013). In BSF cells, however, it has been shown that defects in mitosis, induced for example, through the depletion, by RNAi, of the mitotic cyclin TbCYC6 (Hammarton *et al.*, 2003), the mitotic CDK TbCRK3 (Tu and Wang,

2004), or aurora-like kinase TbAUK1 (Li and Wang, 2006), results in cytokinesis inhibition, with a consequent generation and accumulation of cells with an enlarged nucleus and multiple kinetoplasts that re-replicate their nuclear DNA without nuclear division, suggesting that this life cycle stage of the parasite possesses the mitosis to cytokinesis checkpoint (reviewed in Hammarton, 2007; Li, 2012).

Published reports on the effect of TbORC1/CDC6 silencing in BSF cells have not been corroborative (Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013). For instance, the timing of phenotype appearance and progression was different, although both studies reported the appearance and accumulation of the same type of cells with multiple nucleus and kinetoplasts. More strikingly, analysis of these cells by flow cytometry revealed very different results: while in (Tiengwe *et al.*, 2012b) these cells appeared to have the double of the amount of DNA than G2/M cells, in (Benmerzouga *et al.*, 2013) this was not observed, with these cells apparently arresting in G2/M phase instead. Therefore, these results suggest that somehow, the cells are still able to replicate their nuclear DNA in the absence of TbORC1/CDC6. Unfortunately, in (Tiengwe *et al.*, 2012b) it was not possible to assess the effect of TbORC1/CDC6 RNAi on DNA replication levels, but data from (Benmerzouga *et al.*, 2013) confirmed that, like in PCF cells, TbORC1/CDC6 RNAi in BSF cells resulted in a decrease of replicating cells. It is perhaps worth mentioning that in (Benmerzouga *et al.*, 2013), in order to assess the impact of TbORC1/CDC6 on DNA replication, the cells were incubated with BrdU, a thymidine analogue (already detailed in Chapter 3), for 24 h before analysis. It has been reported in mammalian cells that prolonged exposure to thymidine analogues results in cytotoxic effects, induction of DNA damage signals and perturbation of cell cycle progression, with an accumulation of cells in G2 phase (Zhao *et al.*, 2013; Fujii *et al.*, 2002; Duque and Rakic, 2011). Especially in kinetoplastids, assays involving thymidine analogues must be interpreted with care, as these parasites possess the hypermodified base J (Borst and Sabatini, 2008), which has been shown to be involved in transcription termination in *Leishmania* promastigotes (insect stage) and *T. brucei* BSF cells (van Luenen *et al.*, 2012; Reynolds *et al.*, 2014), and is known to be replaced by thymidine analogues during DNA replication. However, it was recently shown that incorporation of the thymidine analogue BrdU *per se* into the genome, and

not just the replacement of base J, is toxic for *Leishmania* cells (Reynolds *et al.*, 2014), while prolonged, as well as high dose, incubation of *T. brucei* BSF cells with EdU appears to result in a pronounced S phase arrest (Dr Daniel Paape, unpublished). It is therefore possible that the extended exposure (24 h) of the cells to BrdU in (Benmerzouga *et al.*, 2013) might have resulted in side effects of BrdU incorporation into the genome, and thus possibly obfuscates the interpretation of the effects of TbORC1/CDC6 depletion.

In an attempt to clarify the phenotype of TbORC1/CDC6 depletion by RNAi in BSF cells, a new cell line was generated and analysed here for cell growth, cell cycle effects and DNA replication. RNAi induction resulted in a similar overall phenotype to the ones reported (Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013), and analysis of EdU incorporation, with minimal incubation period, clearly showed that the loss of TbORC1/CDC6 impairs DNA replication in BSF cells. Nonetheless, the phenotypes here obtained did not mimic either of the ones previously reported (Tiengwe *et al.*, 2012b; Benmerzouga *et al.*, 2013). For instance, the growth and cell cycle defects observed were slower to arise and less severe than reported by (Tiengwe *et al.*, 2012b), although a large amount of unclassifiable cells was also detected. It is not clear whether the results here presented are more similar to the ones reported by (Benmerzouga *et al.*, 2013), as the analyses were represented differently. Nevertheless, both appear to detect a growth defect at 24 h post induction, which is aggravated in later time points, although cells continued to grow in (Benmerzouga *et al.*, 2013), while here the cells started to die. In addition, in (Benmerzouga *et al.*, 2013) no clear cell cycle defects were observed from the cell cycle analysis from DAPI-stained cells, although some abnormal cells were observed, and flow cytometry analysis showed an increase in cells with G2/M phase DNA content (Benmerzouga *et al.*, 2013). How these observations fit with the observed decrease in DNA replication is not explained (Benmerzouga *et al.*, 2013). Unfortunately, in the present study it was not possible to perform flow cytometry analysis on the cells' DNA content, and would have been of great value to attempt to unify these three studies.

In (Tiengwe *et al.*, 2012b), individual silencing of TbORC4 and Tb7980 by RNAi resulted in a strikingly similar phenotype to the one described for TbORC1/CDC6 depletion, but direct evidence for their role in DNA replication was lacking.

Silencing of TbORC4 resulted in a mild and delayed growth and cell cycle phenotype in the BSF, though ultimately similar to the one obtained for TbORC1/CDC6 RNAi. Nevertheless, a considerable reduction of EdU-incorporation at 48 h post-induction strongly supports the involvement of TbORC4 in DNA replication in BSF cells. Unfortunately, in the present study induction of Tb7980 silencing by RNAi was not successful. Indeed, it proved impossible to generate stem-loop RNAi cells in PCF *T. brucei* also (Chapter 3). Whether these experimental difficulties are related, perhaps indicating considerable importance for this factor, and selection against even a small amount of 'leaky' RNAi, though this is unclear. Therefore, any role for Tb7980 in *T. brucei* replication currently remains limited to its interaction with TbORC1/CDC6 in PCF cells (Tiengwe *et al.*, 2012b).

The effects of Tb3120 depletion have not been investigated in BSF cells before (Tiengwe *et al.*, 2012b). Although here a decrease in protein levels was detected 48 h post RNAi induction, no clear effects were observed on cell growth, the cell cycle, or levels of DNA replication. Tb3120 RNAi phenotypes emerged in PCF cells only at late time points (72 h onwards, Chapter 3) after induction, and so it is possible that a similar delayed phenotype is seen in BSF cells, and remains to be tested. The basis for this delay remains unclear, though the relatively low levels of Tb3120^{12myc} protein loss seen by western blot in BSF cells (when compared with the apparent abrogation of TbORC1/CDC6^{12myc} and TbORC1B^{12myc} 24 h after induction of TbORC1/CDC6 and TbORC1B RNAi, respectively) may suggest that this protein is relatively stable, possibly accounting for the delay in phenotype appearance, as had been suggested in Chapter 3.

Taken together, though the PCF and BSF RNAi data for Tb3120 and TbORC4 do not entirely overlap and are less compelling than for TbORC1/CDC6 and TbORC1B, it seems likely that these factors can be considered components of the replication initiation machinery.

Like in PCF cells, silencing of TbORC1B by RNAi had not been examined in BSF cells. Here we show that TbORC1B RNAi induction in BSF cells resulted in a more severe phenotype than RNAi of any of the other factors, mirroring the data obtaining in PCF cells (Chapter 3, section 3.3.2.5). The growth and cell cycle consequences of TbORC1B depletion were very comparable with those of

TbORC1/CDC6 RNAi-induction, though the reduction in the number of replicating cells (by 90% after 24 h) was more rapid. Two independent experiments have shown that the rapidity and nature of these phenotypes is highly reproducible. Thus, TbORC1B appears to be a central factor in *T. brucei* DNA replication in both BSF and PCF cells, despite the difficulties we have reported in detecting TbORC1B's interaction with TbORC1/CDC6, or other putative ORC factors, as discussed in Chapter 3. Indeed, localisation of myc-tagged TbORC1B suggests that its role in BSF cells is the same as in PCF cells: here it was shown that TbORC1B^{12myc}, uniquely amongst all the other 12myc-tagged putative Orc-like proteins, displays cell cycle dependent sub-nuclear localisation in BSF cells. The timing and punctate nature of this localisation, which is limited to late G1, S phase and early G2 phase cells, suggests that in both life cycle stages TbORC1B performs a regulatory role. Likewise, localisation and dynamics of TbORC1/CDC6 and interacting factors mimic the results obtained in PCF cells (Chapter 3, section 3.5.2): TbORC1/CDC6^{12myc}, TbORC4^{12myc}, ^{12myc}Tb7980, Tb3120^{12myc} and Tb1120^{12myc} localise to the nucleus of the BSF cell throughout the cell cycle, as punctate.

4.6.2 DNA replication dynamics *T. brucei* BSF cells

Origins of DNA replication in the eleven megabase chromosomes have been mapped to date only in *T. brucei* PCF cells of strain TREU 927 (Tiengwe *et al.*, 2012a). As in other eukaryotes, the number of initiator binding sites (TbORC1/CDC6 in this case) outnumbered the origins that were activated, which co-localised with a subset of TbORC1/CDC6 binding sites in the highly transcribed core of the chromosomes (Tiengwe *et al.*, 2012a). No consensus sequence was identified at the origins or TbORC1/CDC6-binding sites, although these localised to the boundaries of the polycistronic transcriptions units, most frequently transcription initiation sites enriched for histone H4K10ac (Siegel *et al.*, 2009), thus suggesting an intimate relationship between DNA replication and transcription (Tiengwe *et al.*, 2012a), as seen in other eukaryotes (reviewed in Leonard and Mechali, 2013; Mechali, 2010; Barlow and Nussenzweig, 2014; Helmrich *et al.*, 2013). Nevertheless, the estimated number of active origins (42) was smaller than it might have been predicted for a eukaryotic organism; for instance, ~280 origins have been mapped by Sort-seq in the genome of *S. cerevisiae* (Muller *et al.*, 2014), which is approximately half the size of the

megabase genome of *T. brucei*. Indeed, with a density of only about one origin per ~600 Kbp, origins in *T. brucei* appear to be much more dispersed than in other analysed eukaryotes (Tiengwe *et al.*, 2012a). Recently, the rate of nuclear DNA replication has been calculated in *T. brucei* PCF and BSF cells to be of 3.7 Kbp.min^{-1} and 4.4 Kbp.min^{-1}, respectively (Calderano *et al.*, 2015), thus being marginally faster than rates described in other eukaryotes (2-3 Kbp.min^{-1}), where replication rate, the number of activated origins and genome size are assumed to be interconnected (reviewed in Mechali, 2010; Masai *et al.*, 2010). In other eukaryotes, it has been reported that the location of origins used (reviewed in DePamphilis, 2005), or the timing of origin activation, can differ between cell types (Ryba *et al.*, 2010) (and reviewed in Jackson *et al.*, 2012; Leonard and Mechali, 2013; Masai *et al.*, 2010; Mechali, 2010). BSF and PCF cells are two distinct life cycle stages of *T. brucei*, differing in many aspects of their biology, which prompted us to compare the origins usage between these two cell types. To allow direct comparison between life cycle stages, BSF and PCF cells of strain Lister 427 were used, therefore also providing a further strain comparison with the analysis made in (Tiengwe *et al.*, 2012a).

Mapping origins in Lister 427 *T. brucei* PCF and BSF cells revealed considerable inflexibility in replication dynamics. Virtually all origins localised to the chromosome core regions, and there were no clear differences between BSF and PCF cells: the locations of all origins appeared the same in both life cycle stages. Although at two locations two origins could be detected in PCF cells, while only a single MFA-seq peak was seen in BSF cells, the two PCF origins were closely spaced. These minor differences may therefore merely reflect the limitations of FACS sorting and mapping resolution by MFA-seq. Even more striking than origin location conservation was the finding that the pattern of replication timing, as judged by variable MFA-seq peak height and width, was strongly comparable between BSF and PCF cells. The only, minor differences were where some origins mapped in BSF early S samples had lower amplitude peaks than the same peaks in the PCF early S data. Though it is possible that this difference reflects some origins firing later in BSF cell S phase, which may be possible if the replication rate in BSF cells is 15% faster than in PCF cells (Calderano *et al.*, 2015), it is equally likely that these differences may just reflect FACS sampling differences. Taken together, these data suggest that in

both life cycle stages not only are the same SSRs chosen to act as origins, but the timing of their firing during S phase is maintained. This is remarkable, for two reasons. First, the active origins represent only ~12% of the mapped TbORC1/CDC6 binding sites in the core region of the chromosomes (Tiengwe *et al.*, 2012a), and no sequence feature has been described that distinguishes most *T. brucei* origin-active SSRs from inactive SSRs. Second, whereas in other eukaryotes origins might be distinguished from potential origins by levels of transcription (e.g. promoter strength) (reviewed in Helmrich *et al.*, 2013; Leonard and Mechali, 2013; Mechali, 2010; Masai *et al.*, 2010), there is no evidence to suggest any regulated activity of the RNA Pol II polycistrons, and it is very possible that transcription at each 'promoter' is constitutive and initiates at equal levels (given the most gene expression control is post-transcriptional) (Jensen *et al.*, 2009; Siegel *et al.*, 2010; Kramer, 2012; Clayton, 2013; Clayton, 2002). What, then, marks the origin-active sites for selective use and maintains this during differentiation? It is increasingly clear that centromeres are very early replicating features in most eukaryotes, and in chromosomes 1-8 of *T. brucei*, where centromeres have been mapped (Obado *et al.*, 2007), the MFA-seq data appears to confirm that this is also true in *T. brucei*, since these represent the highest mapped peaks (Tiengwe *et al.*, 2012a). However, this cannot explain all the data, since the centromeres represent only ~20% of the mapped *T. brucei* origins, and the timing of the remaining ~80% also appears to be invariant.

The origins mapped in BSF and PCF Lister 427 cells not only match each other, but also very closely match the origins mapped previously in PCF cells from strain TREU 927 (Tiengwe *et al.*, 2012a). Virtually all 42 previously predicted origins were seen, with only one origin, in chromosome 2, not being identified. However, as for the BSF to PCF differences, this may simply be because in TREU 927 this origin locus was detected as two very closely separated origins. Within the 41-42 conserved origins, MFA-seq peak height was reasonably conserved, further reinforcing the suggestion that the profile of replication timing is relatively fixed. Five new potential origins were detected in the BSF and PCF Lister 427 samples, which had not been previously noticed (Tiengwe *et al.*, 2012a). Like all other origins, these localised to TbORC1/CDC6 (Tiengwe *et al.*, 2012a) and histone H4K10ac (Siegel *et al.*, 2009) binding sites, at SSRs.

Perhaps more importantly, the work here showed mapped origin usage late in *T. brucei* S phase, which had not been attempted previously (Tiengwe *et al.*, 2012a). These data may suggest that the 41-45 origins mapped in early S phase represent the whole repertoire, as no further MFA-seq peaks were clearly discernible. If so, the megabase genome is, as suspected (Tiengwe *et al.*, 2012a), replicated from a remarkably small number of origins. Very recently, single molecule analysis of replicated DNA (SMARD), limited to chromosome 1, confirmed origin activity at the centromere and, furthermore, suggested that there may be further initiation sites towards at least one of the telomeres in this chromosome (Calderano *et al.*, 2015). MFA-seq data generated to date has not shown clear peaks in any telomeres or subtelomeres, with the exception of the main origin in chromosome 6, which co-localised with the centromere, very close to the subtelomeric region of the chromosome (Tiengwe *et al.*, 2012a). However, it is possible that this reflects the limitations of the MFA-seq strategy, and mapping in these regions is confounded by repetitive sequences and limited annotation of *VSG* location. More work will be needed to localise the origins' initiation site(s) and to ask if such telomere-proximal initiation is seen in all the chromosomes.

As discussed above, DNA replication and transcription are intimately connected, (reviewed in Helmrich *et al.*, 2013; Barlow and Nussenzweig, 2014), including in *T. brucei* (Tiengwe *et al.*, 2012a; Benmerzouga *et al.*, 2013). In other eukaryotes, it has been proposed that differences in origin usage and timing are interconnected with divergences in the transcription programmes employed by different cell types of the organism (reviewed in Jackson *et al.*, 2012; Leonard and Mechali, 2013; Masai *et al.*, 2010; Mechali, 2010). In *T. brucei* and related kinetoplastids however, gene expression is controlled mainly at the post-transcriptional level (Jensen *et al.*, 2009; Siegel *et al.*, 2010; Kramer, 2012), with few genes being regulated directly at the transcriptional level (Reynolds *et al.*, 2014; Maree and Patterton, 2014). The most prominent exceptions to this general trend are the genes encoding surface proteins, which are transcribed by RNA Pol I, and expressed in a tightly life cycle stage-regulated fashion (Gunzl *et al.*, 2003; Gunzl *et al.*, 2014): while the surface of PCF cells is marked by abundant procyclins (reviewed in Dyer *et al.*, 2013), the BSF cell surface is covered by ~10^7 copies of a single VSG (reviewed in McCulloch *et al.*, 2014;

Glover *et al.*, 2013b). Fourteen *VSG* BES have been annotated in the Lister 427 genome (Hertz-Fowler *et al.*, 2008), within which strict expression regulation is seen: only one *VSG* BES is fully transcribed in a single BSF cell at a given time (active expression site), with all other sites being transcriptionally silenced, while all *VSG* BES are silenced in the PCF cell (reviewed in McCulloch *et al.*, 2014; Glover *et al.*, 2013b). MFA-seq mapping in the 14 Lister 427 BES shows that this part of the genome provides the only clear example of life cycle stage-specific transcription-associated replication, and perhaps replication dynamism that contrasts with the rigid replication programme seen in the rest of the genome. Mapping replication in the BES suggests that virtually all of these telomeric sites are very late replicating in both BSF and PCF cells, with a single exception: BES 1, the active expression site in Lister 427 BSF cells that transcribes *VSG 221*. MFA-seq mapping in BES 1 was striking in that it too was late replicating in PCF cells, where it is not transcribed, but was the single BES that showed evidence for early replication in BSF cells. Thus, BES 1 represents the single region of the genome in which active transcription can be shown to dictate replication timing. It will be important to validate this observation further, for instance repeating the same experiment using a cell line actively expressing another BES. Nonetheless, this observation raises a number of questions. Foremost amongst these, is where is the origin that dictates BES 1-specific early replication? Most BES have not been linked to specific megabase chromosome telomeres, and therefore it remains possible that BES 1 replication emerges from within the subtelomere, perhaps as suggested by (Calderano *et al.*, 2015). In several works, it is mentioned that BES 1 is localised to the right-hand side of chromosome 6a (Glover and Horn, 2014; Hertz-Fowler *et al.*, 2008), although this would have to be confirmed in the cell lines (both BSF and PCF) used in this study. Both data presented here and in (Tiengwe *et al.*, 2012a) do not suggest the presence of an origin being activated at that end of chromosome 6, though a weak origin is detectable just upstream of that end's subtelomeric region, which could passively replicate BES 1. However, this origin is detected both in BSF and PCF cells, and thus it seems unlikely that BES 1 is being replicated via this origin, since no replication was seen in the PCF sample. A perhaps more likely explanation for early BSF replication of BES1 is that this is driven by active transcription of the site. If so, mapping the location of the origin is needed to understand the mechanism that drives this. For instance, the

origin may localise to the *VSG* RNA Pol I promoter, and perhaps TbORC1/CDC6 is recruited there by the transcription machinery or chromatin status, both which have been shown to differ between the active and silent BES (reviewed in Horn and McCulloch, 2010; McCulloch *et al.*, 2014; Glover *et al.*, 2013b). Alternatively, the origin may be at the telomere, with replication being directed towards the BES promoter. This is also possible, as TbORC1/CDC6 has been documented to interact with telomere repeats (Benmerzouga *et al.*, 2013), and origin activity has been detected at telomeres in other eukaryotes (Kurth and Gautier, 2010; Drosopoulos *et al.*, 2012). It will be valuable to map TbORC1/CDC6-binding sites at the BES to infer whether these include potential origins. Moreover, if TbORC1/CDC6 and origins can be found at the telomere, selectivity for the active BES needs to be further explored.

The observation of BSF-specific early replication at the active VSG BES raises a wider question regarding *T. brucei* biology: is it possible that DNA replication might be involved in antigenic variation, the mechanism by which the cells change their expressed VSG (reviewed in McCulloch *et al.*, 2014; Glover *et al.*, 2013b)? Some indications of this relationship can be detected in published work (Tiengwe *et al.*, 2012a; Benmerzouga *et al.*, 2013; Gassen *et al.*, 2012; Glover *et al.*, 2013a), but a direct link between replication dynamics and VSG switching has not been reported. Depletion of TbORC1/CDC6 by RNAi has been shown to result in the de-repression of silent *VSGs* and lead to an increase in VSG switching (Tiengwe *et al.*, 2012a; Benmerzouga *et al.*, 2013). However, whether this is a direct consequence of TbORC1/CDC6 function is not clear, as TbORC1/CDC6 depletion by RNAi results in a widespread effect on transcription and subsequent cell death (Tiengwe *et al.*, 2012a). In other studies it has been shown that the BES of *T. brucei* are ‘fragile’ sites, with evidence of DNA breaks in the vicinity of the *VSG* gene (Glover *et al.*, 2013a; Boothroyd *et al.*, 2009); indeed, it has been proposed that it is this telomere fragility that can generate DNA double strand breaks (DSBs) that trigger antigenic variation by DNA repair (Glover *et al.*, 2013a; Glover *et al.*, 2013b; McCulloch *et al.*, 2014). However, though it was originally suggested that such breaks are found specifically in the active BES (Boothroyd *et al.*, 2009), later studies found no difference in break distribution between silent and actively transcribed BES (Jehi *et al.*, 2014; Glover *et al.*, 2013a). Thus, whether BES telomeric fragility results from

transcription or DNA replication, or clashes between them, remains unclear. Nonetheless, telomeric DNA has a complex secondary and chromatin structure, which can render it prone to replicative stress and subsequent replication fork stall or collapse, and potentially result in DSBs (reviewed in Pfeiffer and Lingner, 2013; Giardini *et al.*, 2014). Finally, a model for replication-dependent initiation of VSG switching has been proposed (Kim and Cross, 2010; Kim and Cross, 2011). Though the data here may offer some support for this model, it is based on untested assumptions about replication direction in the BES (as discussed above).

Like with other eukaryotic ORC subunits (reviewed in Barlow and Nussenzweig, 2014), most of TbORC1/CDC6 has been shown to accumulate at the subtelomeres of the *T. brucei* megabase chromosomes (Tiengwe *et al.*, 2012a). However, it is not clear whether DNA replication is initiated from these TbORC1/CDC6-binding sites (Tiengwe *et al.*, 2012a). In addition, TbORC1/CDC6 binding at these locations was mapped in TREU 927 PCF cells, and thus it is not known whether it also localises to the BES (Tiengwe *et al.*, 2012a) or if this localisation is conserved in BSF cells. It will be of great interest to map TbORC1/CDC6 binding sites in both Lister 427 PCF and BSF cells to ask whether it binds to sites within the BES, and if there are differences between PCF and BSF cells, or specific localisation within the active BES. Moreover, it would theoretically allow the identification of potential origins within the BES, if any exist. In all cells, because DNA replication and transcription are intimately related, it is crucial that the cell is able to tightly coordinate both processes, limiting the risk of subsequent genomic instability (reviewed in Barlow and Nussenzweig, 2014; Helmrich *et al.*, 2013). It is well established that when transcription and DNA replication machineries collide, replication fork stalling occurs with consequent collapse, and generation of DNA breaks (reviewed in Helmrich *et al.*, 2013). Normally such instances are avoided, and so it would be of considerable interest to investigate whether *T. brucei* has evolved to promote such interactions in the BES of BSF cells, allowing DNA replication to trigger antigenic variation.

5 Origins of DNA replication in *Leishmania*

5.1 Introduction

5.1.1 The *Leishmania* parasites and leishmaniasis

Like *T. brucei*, *Leishmania* spp. parasites belong to the order Trypanosomatida (Moreira *et al.*, 2004; Adl *et al.*, 2012; Lukes *et al.*, 2014), and possess a complex life cycle in which the parasite has to go through drastic developmental changes (Figure 5.1), including stages in the insect vector (sand fly) and the mammalian host. Unlike *T. brucei*, however, *Leishmania* parasites are intracellular in the mammalian host, with a tropism to phagocytic cells, such as macrophages (reviewed in Kaye and Scott, 2011). The infection of the mammalian host results in leishmaniasis, whose pathology, separated into visceral leishmaniasis (VL) or cutaneous leishmaniasis (CL), appears to depend on both the *Leishmania* species and the mammalian host immune response (reviewed in Kaye and Scott, 2011). For instance, *L. major* and *L. mexicana* species are associated with CL, while *L. donovani* and *L. infantum* cause VL (reviewed in Kaye and Scott, 2011). Recently, it has been estimated that, every year, there are between 700,000 and 1,200,000 cases of CL, and between 200,000-400,000 cases of VL in 98 affected countries (Alvar *et al.*, 2012).

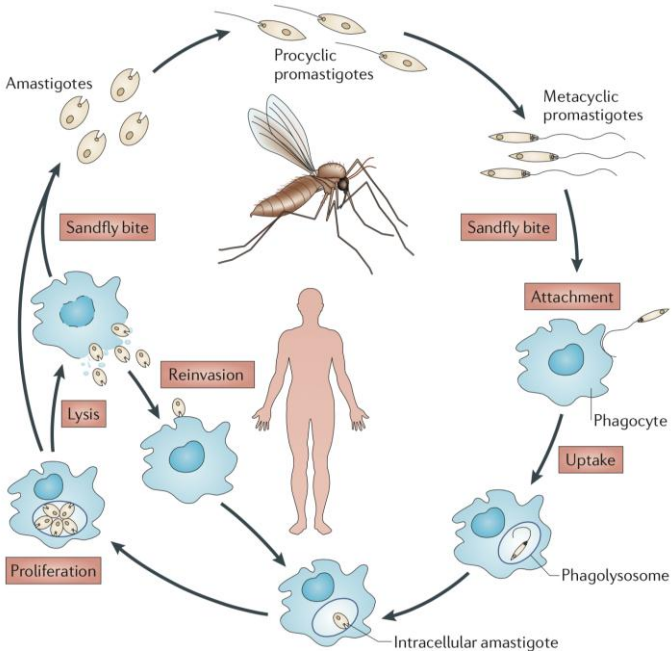


Figure 5.1. Life cycle of *Leishmania* spp. parasites.
Generic diagram of the *Leishmania* parasite life cycle. During its blood meal, the sand fly regurgitates into the mammalian host infective, non-dividing metacyclic promastigotes, which are phagocytosed by a phagocytic cell present in the bite environment. Once inside the phagocytic cell, metacyclic promastigotes differentiate into aflagellate amastigotes, which then divide within the host cell. Once the cell is full with amastigotes, it ruptures, releasing the parasites that will infect another phagocytes. Upon taking a blood meal from an infected mammal, the sand fly ingests

amastigotes, which will then develop within the fly midgut into procyclic promastigotes, and later metacyclic promastigotes, that can be transmitted to a new mammalian host. Several developmental stages occur inside the sand fly, but are not here distinguished from the procyclic promastigotes for simplicity. Reproduced from (Kaye and Scott, 2011), with permission (license number 3586541268882).

5.1.2 *Leishmania* genome organisation and ploidy

The genomes of the so-called TriTryps, *T. brucei*, *T. cruzi*, and *L. major*, were sequenced 10 years ago (Berriman *et al.*, 2005; El-Sayed *et al.*, 2005a; El-Sayed *et al.*, 2005b; Ivens *et al.*, 2005). This has allowed many genome-wide studies to be carried out, including (Tiengwe *et al.*, 2012a; Siegel *et al.*, 2009; Siegel *et al.*, 2010; Jensen *et al.*, 2009; Rogers *et al.*, 2011; Jones *et al.*, 2014; Alsford *et al.*, 2011). The main, common characteristic of the TriTryp genomes is the organisation of their genes into large polycistronic gene clusters, also known as polycistronic transcription units or directional gene clusters (DGCs) (El-Sayed *et al.*, 2005b; Jackson, 2014; Stuart and Myler, 2006), which are delimited by divergent, convergent or head-to-tail strand-switch regions (SSRs). While the *T. brucei* genome is composed of 11 diploid megabase chromosomes (0.9 to 6 Mbp), as well as variable numbers of aneuploid intermediate (150 to 900 Kbp) and mini chromosomes (50 to 150 Kbp) (Melville *et al.*, 1998; Turner *et al.*, 1997; Berriman *et al.*, 2005), the *Leishmania* genome is comprised of 34 to 36 chromosomes, depending on the species (reviewed in Stuart and Myler, 2006), which display variable ploidy (below). For instance, the old world *Leishmania* species *L. major* has 36 chromosomes (0.28 to 2.8 Mbp) (Ivens *et al.*, 2005; Ravel *et al.*, 1998), while the new world *Leishmania L. mexicana*, has only 34. This may have arisen by fusion of two pairs of *L. major* chromosomes (8 and 29, and 20 and 36) (Ivens *et al.*, 2005; Britto *et al.*, 1998) to form two larger *L. mexicana* chromosomes (8 and 20). However, it is also possible that it occurred by fission of the two *L. mexicana* chromosomes, resulting in the four *L. major* chromosomes.

Despite being estimated to have diverged around 250 million years ago (Lukes *et al.*, 2014), comparison of the *T. brucei* and *L. major* genomes revealed a high level of gene conservation and synteny, with 68% of the *T. brucei*, and 75% of the *L. major,* genes localising to the same genomic context (El-Sayed *et al.*, 2005b). Indeed, 20 of the 36 *L. major* chromosomes appear to be almost

completely syntenic within the core regions of the larger *T. brucei* chromosomes, as represented in Figure 5.2, suggesting that the *T. brucei* chromosomes might have resulted from the fusion of the smaller *L. major* ones (Figure 5.2). It was also observed that around 43% of the synteny breakpoints in *T. brucei* and *L. major* localised with, or are very close to, SSRs separating two DGCs, and it is here that species-specific gene families, retroelements and RNA genes localise (El-Sayed *et al.*, 2005b; Stuart and Myler, 2006). The main difference in gene content between the two genomes appears to lie on the telomeres and subtelomeric regions, which are short in *L. major*, but long in *T. brucei* (Berriman *et al.*, 2005; El-Sayed *et al.*, 2005b; Ivens *et al.*, 2005). The increased size of the *T. brucei* subtelomeres is due to the evolution of a huge archive of *VSG* genes, found in expression sites and *VSG* arrays, which are necessary for immune evasion via antigenic variation (reviewed in McCulloch *et al.*, 2014). Another difference between the two genomes is based on genome stability. While the 11 *T. brucei* megabase chromosomes are stably diploid, it has been shown that aneuploidy is frequent in *Leishmania*, with the parasites appearing to tolerate considerable variations in chromosome copy number (ranging from haploid to >tetraploid) (Rogers *et al.*, 2011; Sterkers *et al.*, 2011). The chromosomes that display aneuploidy vary between different species, as well as within a species and even during growth (Rogers *et al.*, 2011; Sterkers *et al.*, 2011; Lachaud *et al.*, 2014; Downing *et al.*, 2011). It is not known how this aneuploidy originates in *Leishmania* parasites, though it has been suggested that it might be related to nuclear DNA replication (Rogers *et al.*, 2011; Sterkers *et al.*, 2012; Lachaud *et al.*, 2014). In addition, it is unclear if chromosome and gene copy number variation within chromosomes may be related to the extensive gene amplification events that are found in *Leishmania*, and may relate to mechanisms for gene expression variation, such as in response to drug selection (reviewed in Leprohon *et al.*, 2014; Mannaert *et al.*, 2012; Sterkers *et al.*, 2012).

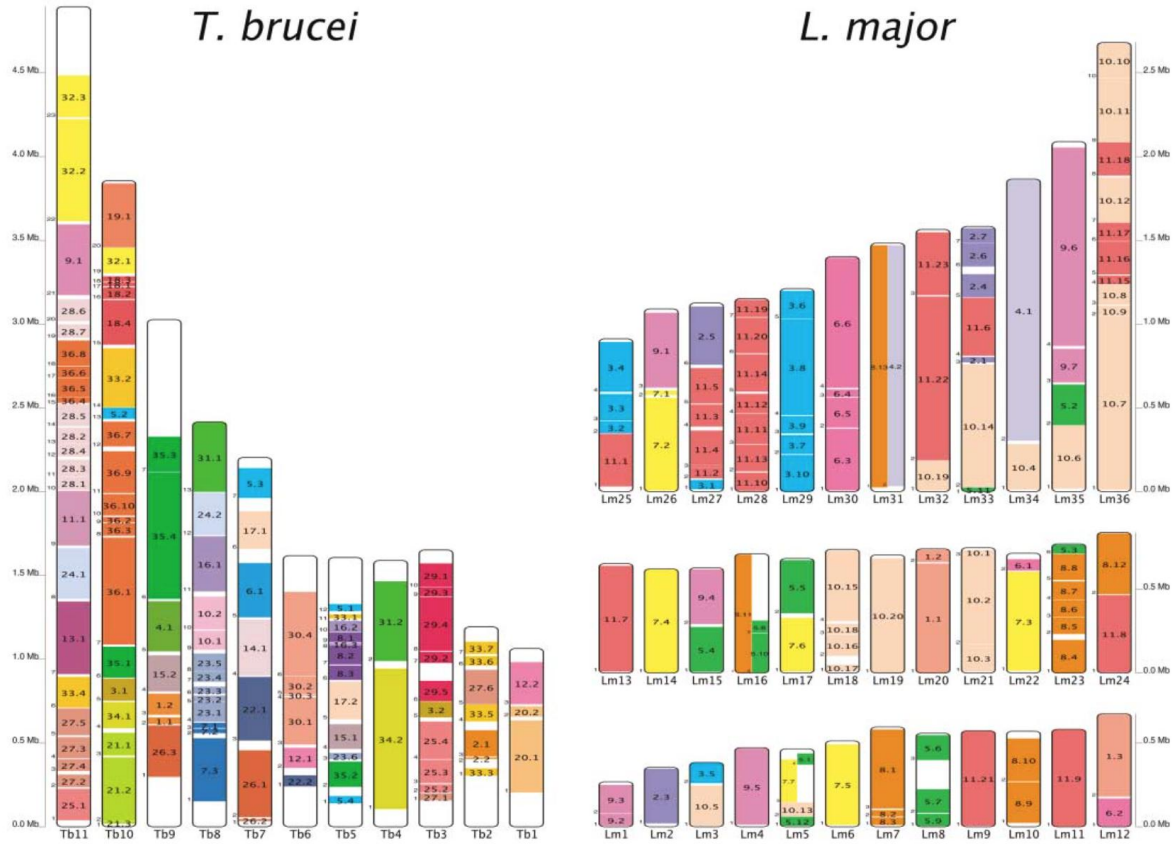


Figure 5.2. Synteny maps of *T. brucei* and *L. major* chromosomes.
The eleven megabase chromosomes of *T. brucei* are shown on the left. To each *T. brucei* chromosome, the syntenic regions in each of *L. major* chromosomes are shown, so that each colour corresponds to each of the 36 *L. major* chromosomes. The inverse is shown on the panel on the right. Here, the 36 chromosomes of *L. major* are coloured according to the *T. brucei* eleven chromosomes, each individually coloured. The syntenic regions are labelled with the chromosome number they belong to. Tb1 refers to *T. brucei* chromosome 1, while Lm1 refers to *L. major* chromosome 1, for instance. Chromosomes are designed to scale, which is shown in Mbp. Reproduced from (El-Sayed *et al.*, 2005b), with permission (license number 3586570186594).

5.1.3 Nuclear DNA replication in Leishmania

Very little is known about nuclear DNA replication in *Leishmania* parasites, with the few data available focusing on the TbORC1/CDC6 orthologues in *L. major* and *L. donovani*, LmTbORC1/CDC6 and LdTbORC1/CDC6, respectively (Kumar *et al.*, 2008; Kumar *et al.*, 2012). In both cases, the protein localised to the nucleus of the cell throughout the cell cycle, similar to TbORC1/CDC6 and TcORC1/CDC6 (Godoy *et al.*, 2009). However, both LmORC1/CDC6 and LdORC1/CDC6 were expressed from an episomal plasmid, and thus, the levels of expression might have not been the ones naturally occurring in the cell (Kumar *et al.*, 2008; Kumar *et al.*, 2012). No functional data on these proteins or analysis of replication coordination in *Leishmania* has been described, though several attempts have been made to map replication origins and centromeres by

chromosome fragmentation and directed cloning of sequences into bacterial plasmids (Dubessay *et al.*, 2001; Dubessay *et al.*, 2002; Tamar and Papadopoulou, 2001).

As already mentioned in the previous chapters, the origins of replication have been recently mapped in *T. brucei* procyclic form (PCF, insect stage) cells (Tiengwe *et al.*, 2012a). Although *T. brucei* and *L. major* are distinct organisms, in eukaryotes it is assumed that origins of replication are not defined by a consensus sequence, but by chromatin status, chromatin environment and context (reviewed in Leonard and Mechali, 2013); therefore, the high level of synteny between the two parasites' genomes may enable the identification of common characteristics or features defining origins of replication in these organisms. In addition, origins of replication appear to be highly conserved between *T. brucei* PCF and bloodstream form (BSF) cells, as well as between two strains of the parasite (Chapter 4).

Given the profound genome synteny between *Leishmania* and *T. brucei*, albeit within the context of considerable differences in genome architecture and stability, this chapter describes the mapping of the origins of replication in both *L. major* and *L. mexicana* promastigote cells (insect stage). One motivation for this was to enable comparison with the previously mapped origins of replication in *T. brucei* cells (Tiengwe *et al.*, 2012a), perhaps revealing features that define an origin of replication in kinetoplastids.

5.2 Origins of DNA replication in *Leishmania major*

5.2.1 Mapping the origins of replication in *L. major*

To investigate the origins of replication in *L. major*, promastigote cells from *L. major* strain Friedlin, used for the assembly of the reference genome (Ivens *et al.*, 2005), were used. The cell line was maintained and cell cultures were prepared by Dr Amy Goundry (Jeremy Mottram's laboratory). Around 10^9 cells were harvested (from cultures at around 5 x 10^6 cells.ml^{-1}), fixed in 70% methanol and stained with propidium iodide (PI). The cells were then sorted into early S, late S and G2 phases by FACS using a BD FACSAria I™ cell sorter system (BD Biosciences), according to their DNA content, as detailed in the materials and methods Chapter 2, section 2.7.3. The gates used for the isolation of the different cell cycle stages are represented in Figure 5.3. Each sample was collected directly into lysis solution, and the gDNA was purified and sent for sequencing at Glasgow Polyomics (University of Glasgow; samples processed by Ms Julie Galbraith). Here, the DNA libraries were generated using the Nextera® XT DNA Sample Preparation kit (Illumina), and sequenced using the Illumina MiSeq paired-end 250 bp sequencing system (Illumina). Like in Chapter 4, the samples were multiplexed, with each of the early S, late S, and G2 phase samples library DNA being processed in the same run, for direct comparison. The resulting data was analysed by marker frequency analysis (MFA) as described in (Tiengwe *et al.*, 2012a), with a few adaptations, as already detailed in Chapter 4, section 4.5.1 for the mapping of the origins of replication in *T. brucei* BSF cells. Here, the *L. major* strain Friedlin reference genome available on TriTrypDB version 6.0 was used. A simplified diagram is shown in Chapter 4, Figure 4.19, and the basic coding script can be found in the appendices, section 7.8. In this case, Dr Nicholas J. Dickens conducted all the mapping (WTCMP Bioinformatics team, University of Glasgow), and generated the output graphs for each chromosome using ggplot2 and the R package (R Development Core Team, 2010). The early S to G2 ratio for each of the 36 chromosomes is shown in Figure 5.4, while the late S to G2 ratios are shown in Figure 5.5. For better interpretation of the results, and to allow comparisons to the origins mapped in *T. brucei* PCF cells (Tiengwe *et al.*, 2012a), the S to G2 ratios are shown relative to the genes for each chromosome, as well as sites of acetylated histone H3 (H3ac) binding, which were mapped previously (Thomas *et al.*, 2009), and

demarcate the start sites of polycistronic transcription in *L. major*, similar to histone H4 acetylated at K10 (H4K10ac) in *T. brucei* (Siegel *et al.*, 2009).

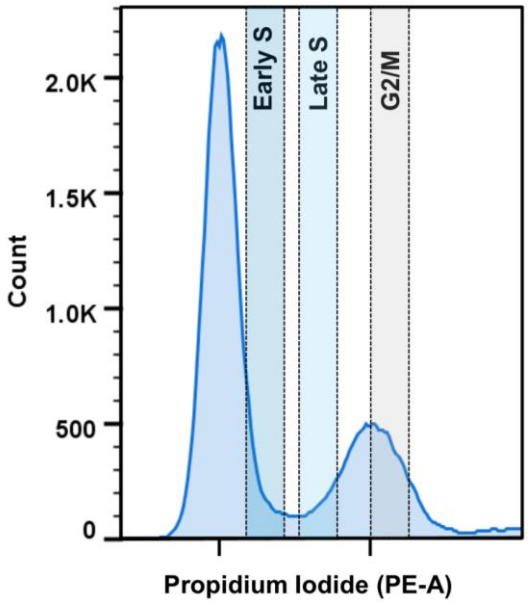


Figure 5.3. Sorting of *L. major* asynchronous cultures into early S, late S and G2/M phases by FACS.
Histogram depicting the cell cycle profile of a mid-log phase *L. major* promastigote cell culture, reflecting the fluorescent signal emitted by the staining of DNA with propidium iodide (PI). The different sections (dashed rectangles) represent the gates defined to purify the different cell cycle populations using a BD FACSAria I sorting flow cytometry machine. The same gates were used for the sorting of *L. mexicana* cells, discussed in section 5.3.

Surprisingly, only a single MFA-seq peak was detected per chromosome in early S *L. major* cells (Figure 5.4), suggesting that a single origin of replication is present in each chromosome. MFA-seq from late S cells, also compared with G2, did not reveal any further peaks per chromosome (Figure 5.5), even in the largest chromosomes, suggesting that no extra origins are activated in later stages of S phase. Nevertheless, analysis of the late S data showed that all peaks widened, suggesting that DNA replication had proceeded further from each single origin per chromosome from early S to late S (Figure 5.4 and Figure 5.5): on average the width of the early S/G2 peaks were ~0.33 to 0.46 Mbp, but were ~0.5 to 0.83 Mbp in late S/G2.

Replication from a single origin per chromosome is unprecedented in a eukaryote, and contrasts with what was observed in *T. brucei*, where multiple origins of different strength were detected per chromosome (Tiengwe *et al.*, 2012a). Analysis of the late S data suggests that a single origin may suffice for the replication of the smaller chromosomes (chromosomes 1 to 24, between 0.28

and 0.84 Mbp), as here discrete MFA-seq peaks are unclear (Figure 5.5), compared with the pronounced peak for most chromosomes in the early S/G2 data (Figure 5.4). Thus, these chromosomes appear to have completed replication in the late S sampled cells (Figure 5.5). For example, in small-mid sized chromosomes, such as 15 (0.63 Mbp), 18 (0.74 Mbp) and 24 (0.84 Mbp), early S peaks had a width of ~0.33 Mbp to 0.375 Mbp (Figure 5.4), while in late S phase these were of ~0.5 Mbp to 0.625 Mbp (Figure 5.5), the latter corresponding to 74-79% of the chromosomes' size. This was not true for the larger chromosomes (25-36), where a discrete late S/G2 MFA-seq peak could be seen.

All MFA-seq peaks in early S, as well as in lase S phase, appeared to have similar widths, independent of chromosome size. Measurement of peak width in the largest chromosomes, 30-36, suggest a width of 0.37 Mb to 0.46 Mbp in early S phase (Figure 5.4), and 0.54 Mbp to 0.83 Mbp in late S phase (Figure 5.5), very similar to the values obtained for the smaller chromosomes. These measurements suggest that each origin is of similar 'strength', in that replication has proceeded to a similar distance in the same time frame. As a result, data suggests that all origins are fired at the same time, and thus there is no order of origin activation, unlike in *T. brucei*, where some origins appear to be activated earlier than others (Tiengwe *et al.*, 2012a).

It is therefore not clear how the larger chromosomes are able to complete DNA replication successfully per cell cycle, as in many of the chromosomes the single detectable late S/G2 peak was not wide enough to suggest that these chromosomes are completely replicated from each single origin (Figure 5.5): conversely to the smaller chromosomes such as 15 and 18, the widths in late S/G2 peaks of chromosomes 30 to 36 suggest that only 28-56% of the chromosomes' length were replicated in the cells collected as late S phase.

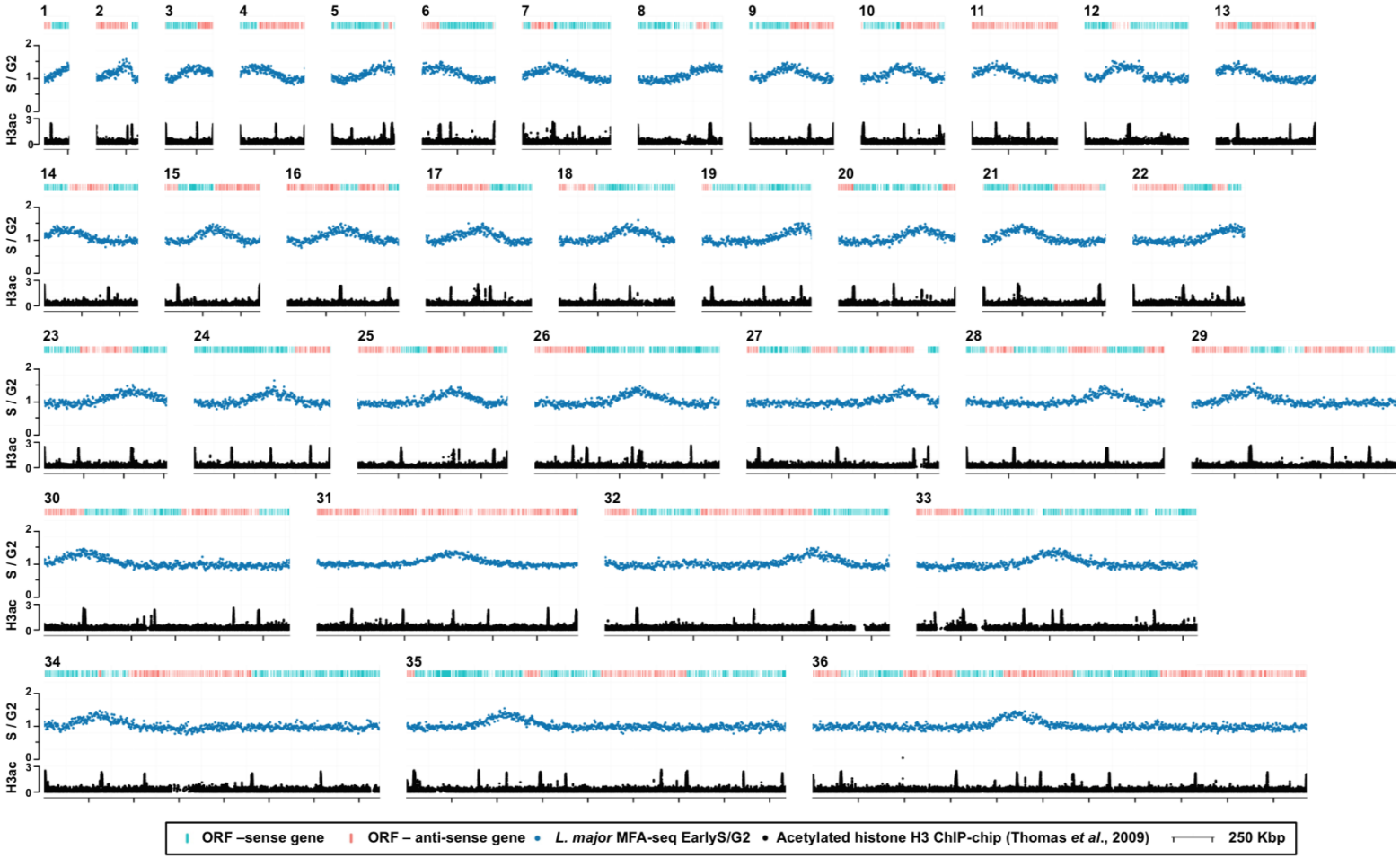


Figure 5.4. Origins of replication in early S phase *L. major* promastigote cells.
Graphs represent the distribution of active origins of replication per *L. major* chromosome (numbered 1 to 36), assessed by MFA-seq. Each graph shows the ratio of the coverage (read-depth) between early S phase and G2 phase samples, where each point (blue) represents the median S/G2 ratio (y-axis) per 2.5 Kbp section across the chromosome (x-axis). Each x-axis gap represents a 250 Kbp interval. The y-axis scale is the same for all graphs, but the legend is only shown in the ones in the far left. All graphs are made to scale according to each chromosome size. At the top of each graph a track is represented with the different genes in the chromosome, in blue the open reading frames (ORF) that are transcribed from the left to the right, and in red the genes transcribed from right to left. Below the MFA-

seq profiles, a track is shown of histone H3ac-enriched sites (Thomas *et al.*, 2009), demarcating the different polycistronic transcription start sites. Each individual graph was generated by Dr Nicholas J. Dickens as described in the materials and methods Chapter 2.

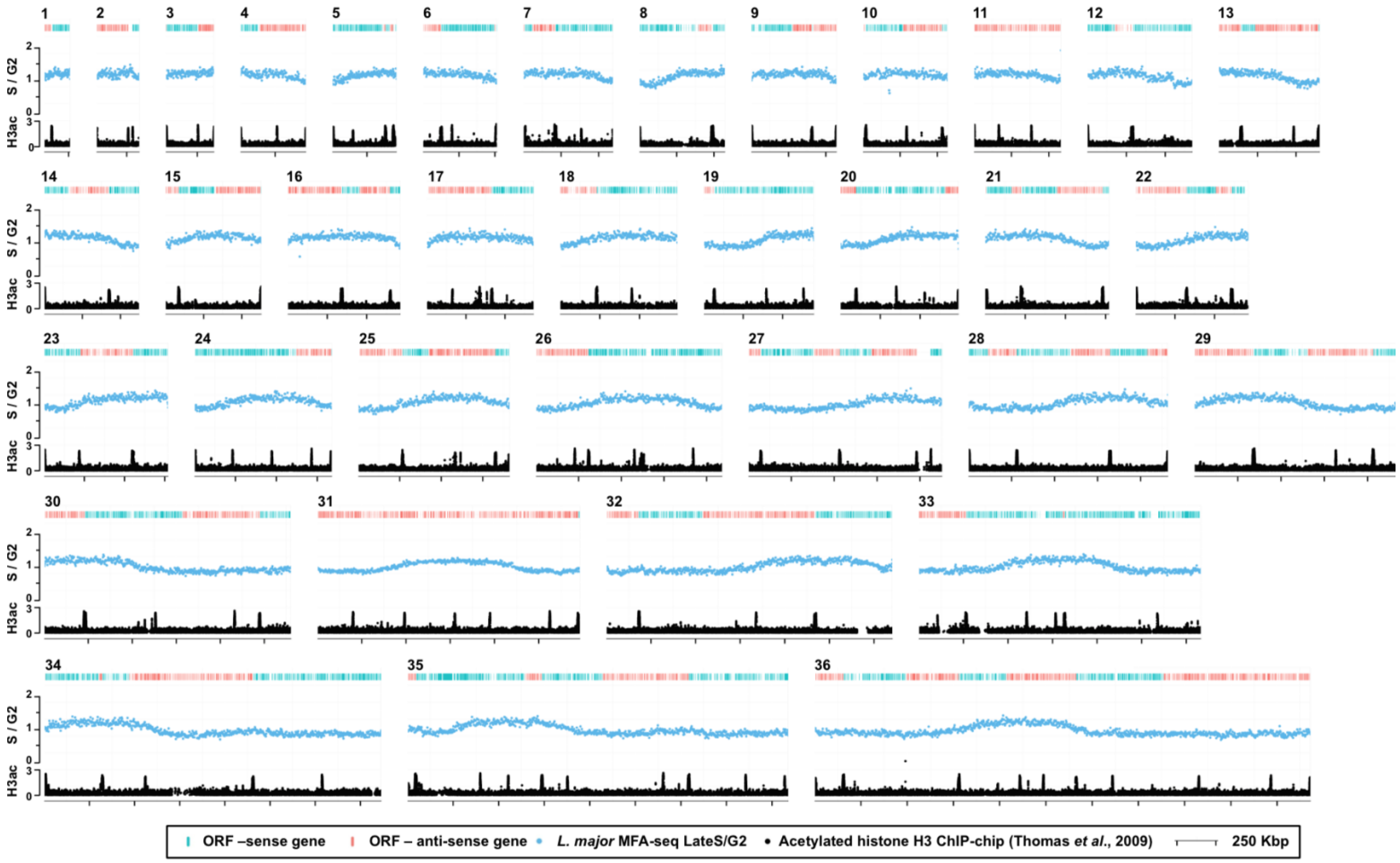


Figure 5.5. Origins of replication in late S phase *L. major* promastigote cells.
Description as in Figure 5.4, with the exception that the MFA-seq data shows the ratio between late S and G2 phases.

Recently, replication rates have been calculated for *T. brucei* PCF cells, to be of 3.7 Kbp.min^{-1} (Calderano *et al.*, 2015), relatively similar but still slightly faster than the rates reported for other eukaryotes (2-3 Kbp.min^{-1}) (reviewed in Masai *et al.*, 2010; Mechali, 2010). The replication rate in *L. major* or other *Leishmania* species has not been calculated to date. Nevertheless, in *L. major* promastigotes it was estimated that G1, S and G2 phases comprise 7.7 h of the 10.2 h cell cycle (Ambit *et al.*, 2011), while in *L. mexicana* it was possible to estimate the S phase as lasting 2.9 h (40%) of the 7.1 h cell cycle (Wheeler *et al.*, 2011). This measurements suggest a longer S phase than in *T. brucei* PCF cells, where S phase is estimated to last 1.5 h (17.6%) of a cell cycle of 8.5 h (Woodward and Gull, 1990). One possibility, then, is that *Leishmania* replication rates might be slower than in *T. brucei*, and it will be invaluable to make this measurement in the future. Assuming that in *L. major*, mirroring the data available for *L. mexicana* (Wheeler *et al.*, 2011), S phase occupies around 40% of the cell cycle (total of 10.2 h) (Ambit *et al.*, 2011), it would take about 4 h to complete. At a replication rate of 2 Kbp.min^{-1}, a bi-directional replication fork would be able to replicate about 0.96 Mbp, while at a 3.7 Kbp.min^{-1} rate it would replicate 1.77 Mbp. In either case, such rates would not be sufficient to allow complete replication of all chromosomes: for example, chromosomes 34, 35 and 36, of 1.86 Mbp, 2.09 Mbp and 2.68 Mbp, respectively, would not be replicated at the faster rate of 3.7 Kbp.min^{-1}, while chromosomes 26 to 36 would not be fully replicated at a fork rate of 2 Kbp.min^{-1}. Therefore, it seems very unlikely that a single origin drives the complete replication of the large chromosomes in *L. major*. Analysis of *L. mexicana* (below), where a firmer estimate of S phase has been made, confirms this dichotomy between the MFA-seq data and prediction.

Besides the very similar peak width (described above) between the 36 chromosomes, all of the MFA-seq peaks appeared to have the same height, of around 1.5 S/G2 ratio in the early S phase samples (Figure 5.4). This contrasts with the observations made in *T. brucei*, where MFA-seq peaks of variable heights (1.1 to 1.8) were detected within the same and between chromosomes, thus suggesting a replication timing program in *T. brucei*, with some origins firing earlier (higher peaks) than others (lower peaks) during S phase (Tiengwe *et al.*, 2012a). This further supports the absence of a temporal order of origin

usage in *L. major*, suggesting that DNA synthesis from the single origin in each chromosome is activated at the same time in S phase.

Like in *T. brucei*, almost all *L. major* origins co-localised to SSRs, at the boundaries between two polycistronic transcription units (Figure 5.4). The single exception was the origin in chromosome 1, which localises to the right end of the chromosome, where co-localisation with the telomere or the end of the 50-gene polycistron (McDonagh *et al.*, 2000; Myler *et al.*, 1999) cannot be distinguished. It is worth noting that the location of the origin predicted here by MFA-seq does not correspond with sequence-based predictions of *L. major* chromosome 1, which suggested localisation to the central SSR (McDonagh *et al.*, 2000). Of the remaining 35 origins (Figure 5.4), 30 (83.3%) co-localised with histone H3ac-binding sites, at divergent SSRs (dSSRs, 47.2%) or head-to-tail SSRs (h-t SSRs, 36.1%), while 5 localised to convergent SSRs (cSSRs, 13.9%), deemed transcription termination sites. These percentages are similar to the ones obtained for *T. brucei*: 85% of the origins co-localised with histone H4K10ac-binding sites (Siegel *et al.*, 2009), 42.9% at dSSRs, 31% at the h-t SSRs, and 4.7% at the cSSRs, while the localisation of 9 origins was either at the centromeres or unclear (21.4%) (Figure 5.7) (Tiengwe *et al.*, 2012a). To date, the centromeres in the *L. major* chromosomes have not been mapped. Consequently, it is not possible to localise the origins of replication to these regions, which in other eukaryotes (reviewed in Rhind and Gilbert, 2013; Masai *et al.*, 2010), as well as in *T. brucei* (Tiengwe *et al.*, 2012a), appear to replicate early in S phase. Nevertheless, this overlap with initiation transcription sites suggests that, like in other eukaryotes (reviewed in Barlow and Nussenzweig, 2014; Helmrich *et al.*, 2013) and *T. brucei* (Tiengwe *et al.*, 2012a), DNA replication is intimately interconnected with transcription in *L. major*. Origin conservation between *L. major* and *T. brucei* is investigated and discussed in a dedicated section below (section 5.2.3).

5.2.2 Confirmation of an origin of replication by MFA-qPCR

To complement the above MFA-seq data in *L. major*, MFA analysis by qPCR was performed for one chromosome, similar to what was done in *T. brucei* (Tiengwe *et al.*, 2012a). For this analysis, chromosome 34 was chosen, as it is one of the large chromosomes (1.86 Mbp). For this analysis, the chromosome was separated

into several sections (A-L), covering most of its length: sections A-C were designed to cover the origin, section L as an origin-proximal region that was not replicated in the early S phase sample, sections D-H covering a large early S non-replicated, central region, J-K as an early S non-replicated region towards the end of the chromosome, and I as an early S non-replicated region to serve as endogenous control ("reference gene") in the qPCR reaction and further analysis (Figure 5.6, A). Primers were designed targeting genes from each of the A-L sections, according to the MIQE guidelines for qPCR primer design (Alvarez-Fernandez, 2013; Dymond, 2013; Bustin *et al.*, 2009). All primer sequences are shown in the materials and methods Chapter 2, Table 2-10. Primer efficiency and specificity was assessed for all pairs of primers through, respectively, the analysis of calibration curves and melting profiles, which resulted in efficiencies of approximately 100%, all within a 8% interval (data not shown). The same amount of gDNA (0.01 ng) from early S and G2 samples, from two independent sorting experiments, was used for the qPCR reaction, and the resultant fluorescence intensity data was analysed by relative quantification using the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). For each pair of primers, corresponding to each section, primers in I were used as "reference gene" for normalisation, and the G2 phase sample as calibrator, as its frequency per marker (pair of primers) should be the same throughout the chromosome (fully replicated), and was therefore set at 1. The relative ratios between early S and G2 phase were then plotted on a graph relative to position in the chromosome. To allow comparison with the MFA-seq data, the qPCR data is shown overlapping with the MFA-seq profile (Figure 5.6, B). A pronounced enrichment in S phase relative to G2 was seen only for A, B, C and L, corresponding with the MFA-seq peak. Thus, the qPCR analysis also suggests that only one origin can be detected, corroborating the data obtained by sequencing (Figure 5.6, A and B).

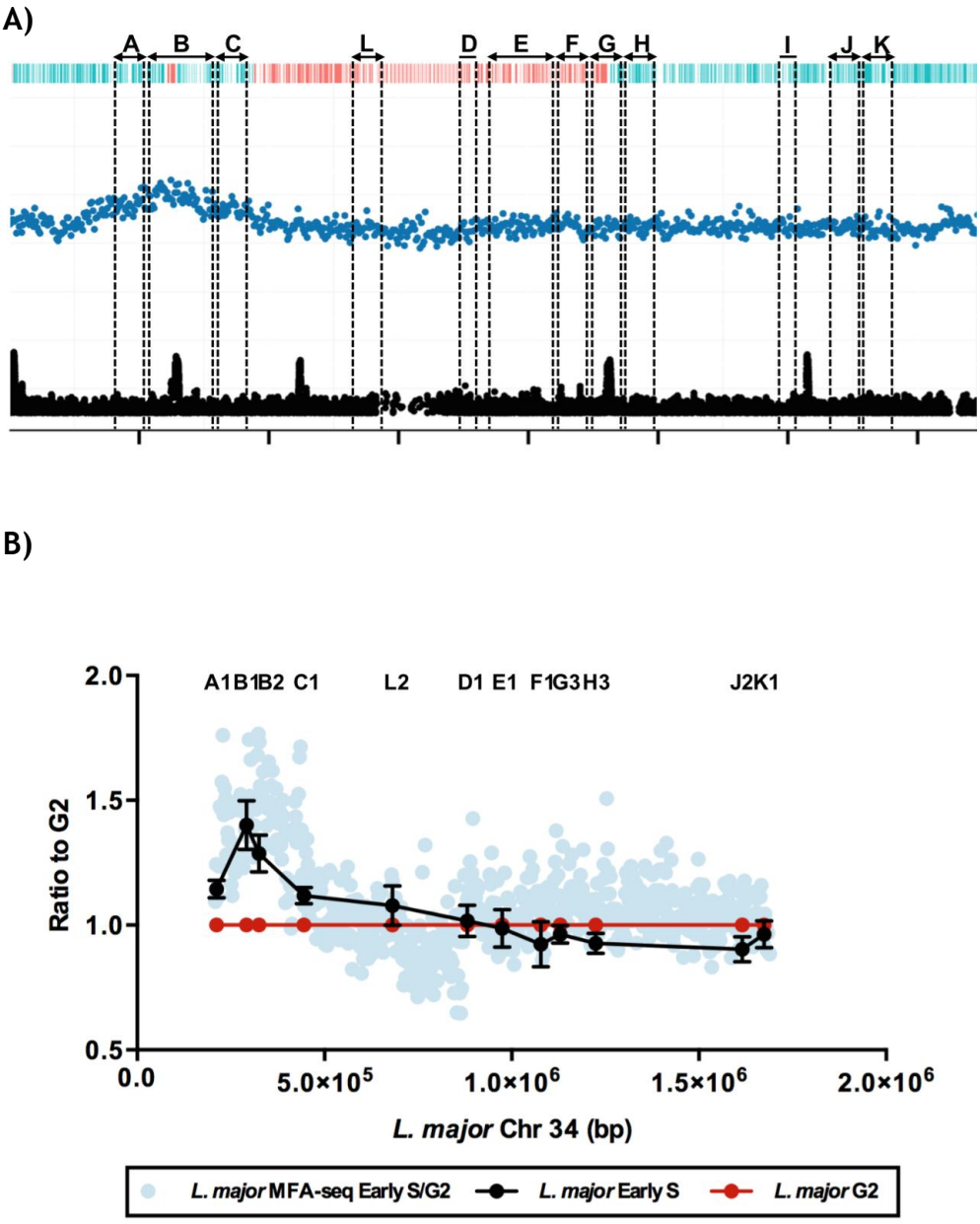


Figure 5.6. *L. major* chromosome 34 analysed by MFA-qPCR.
A) MFA-seq graph (early S/G2) of *L. major* chromosome 34, as shown in Figure 5.4. Dashed lines demark the regions (loci A-L) chosen for the design of primers for MFA analysis by qPCR. Primers pair I2 (within I locus) was used as the "endogenous gene" in the qPCR analysis by the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). B) Validation of the replication origin in *L. major* chromosome 34, by MFA-qPCR the specific loci shown in A). At each loci (A1 to K1 refer to the primers used for each loci), the values from the G2 sample (red) are set as 1 (completely replicated DNA), while the relative amount of the early S phase over G2 is shown (black). Each data point (black) represents the mean of two qPCR runs of two independently obtained samples (gDNA from two independent sorting experiments) (n = 4), with each qPCR run performed with three experimental replicates. The error bars depict the standard error of the mean of the two independent qPCR runs for the two different samples. For ease of comparison, the MFA-seq data was plotted in the background (light blue). All data is represented to their position in the chromosome, with the x-axis represented to scale, depicting the chromosome length in bp.

5.2.3 Origin conservation between *L. major* and *T. brucei*

The initial aim of the investigation of the origins of replication in *L. major* was to explore whether there are common characteristics defining origins between two distinct kinetoplastid organisms. Although the results above show that there are clear differences in replication coordination between *L. major* and *T. brucei*, they also indicate that both organisms have a similar number of origins (36 and 42, respectively), with virtually all localising to SSRs. To allow direct comparison between origins mapped in *L. major* and *T. brucei*, and to confirm the results published (Tiengwe *et al.*, 2012a), *T. brucei* PCF cells of strain TREU 927 were sorted again, sequenced (together with the *L. major* samples, as described in section 5.2.1) and the MFA-seq mapping repeated. The results of this analysis are shown in Figure 5.7, and coincide entirely with the origins previously mapped in PCF TREU 927 cells, showing that origin location does not vary as the cell line is cultured (Tiengwe *et al.*, 2012a).

As shown in Figure 5.2, the high level of synteny between the two genomes allowed the construction of a diagram in which the *T. brucei* chromosomes are represented as blocks of synteny with the *L. major* chromosomes, thus allowing direct comparison of large syntenic regions between the two species' chromosomes (Figure 5.8). Origin location conservation between *L. major* and *T. brucei* was then analysed by comparing the presence of origins in the syntenic regions between the two genomes (Figure 5.8). Of the *T. brucei* SSR loci conserved in location in *L. major*, 41% (light-dark blue) coincided with active origins in *L. major*, while 36% (orange-red) represented sites lacking origin activity in *L. major* (Figure 5.8, pie chart). Fourteen (33%) of the *T. brucei* origins localised to sites of chromosome fusion (or fission) of *L. major* chromosome synteny blocks, meaning sites of re-arrangement or synteny break points. In five of these cases (13% of the 42 *T. brucei* origins), origin activity was not detected in the *L. major* chromosomes. At the other 9 sites, origin activity was conserved in the *L. major* chromosomes; indeed, in two of these cases, one *T. brucei* origin was present in two different *L. major* chromosomes: a *T. brucei* chromosome 7 origin, at around 2 Mbp was retained in both *L. major* chromosomes 17 and 5 (Figure 5.8 and Figure 5.9), while the strongest origin in *T. brucei* chromosome 5 (around 0.25 Mbp) was conserved in both chromosomes 5 and 35 of *L. major* (Figure 5.8). In 8% of the *T. brucei* origins, conservation was

not clear. Finally, only a single origin appeared to be *T. brucei*-specific, localising to the subtelomeric region of chromosome 6, to which there is no corresponding syntenic region in the *L. major* genome (Figure 5.8).

Specific examples of origins conservation, non-conservation and complex patterns of origin conservation are shown in more detail in Figure 5.9. For instance, in Figure 5.9, A), an example is shown of an origin that is active in *L. major* chromosome 30, but is not active in the corresponding syntenic region in *T. brucei* chromosome 6. In contrast, in Figure 5.9, B), an active origin in *T. brucei* chromosome 10 is not conserved in *L. major* chromosome 33, though the same origin is, instead, conserved in *L. major* chromosome 36, as the *T. brucei* region is duplicated in the two *L. major* chromosomes. In Figure 5.9, C), an example is given in which a *T. brucei* origin, localising at a region of a synteny break point between two *L. major* chromosomes, is present in the two *L. major* chromosomes. A further example, depicting the complexity of origin conservation, is shown in Figure 5.9, D): in *T. brucei* chromosome 8, four of its 6 origins, all of various strengths, were also found in, respectively, syntenic regions of *L. major* chromosomes 7, 23, 10 and 31. However, in each case the *L. major* origins were of equal strength. Furthermore, *L. major* chromosome 31, which is >diploid in all species (Rogers *et al.*, 2011), is duplicated in the *T. brucei* genome on chromosomes 8 and 4, and the origin detected in *L. major* chromosome 31 is found in both these *T. brucei* chromosomes. However, the origin in *T. brucei* chromosome 8 is strong and co-localises with the centromere, while its counterpart in chromosome 4 is non-centromeric and weak (Figure 5.9). These data suggest that a single *L. major* origin has adapted to be activated differently in two genomic locations in *T. brucei*. Moreover, whereas in *L. major* all origins appear to have the same strength, the ones conserved, regarding location, have been adapted to differing strengths in *T. brucei*.

Like in *T. brucei* (Tiengwe *et al.*, 2012a), no specific sequence motifs (analysis by Dr Nicholas J. Dickens, not shown) were found that could define an origin in *L. major*, and no common sequence motifs or patterns were found in origins conserved between the two parasites. This is further explored in in section 5.4.

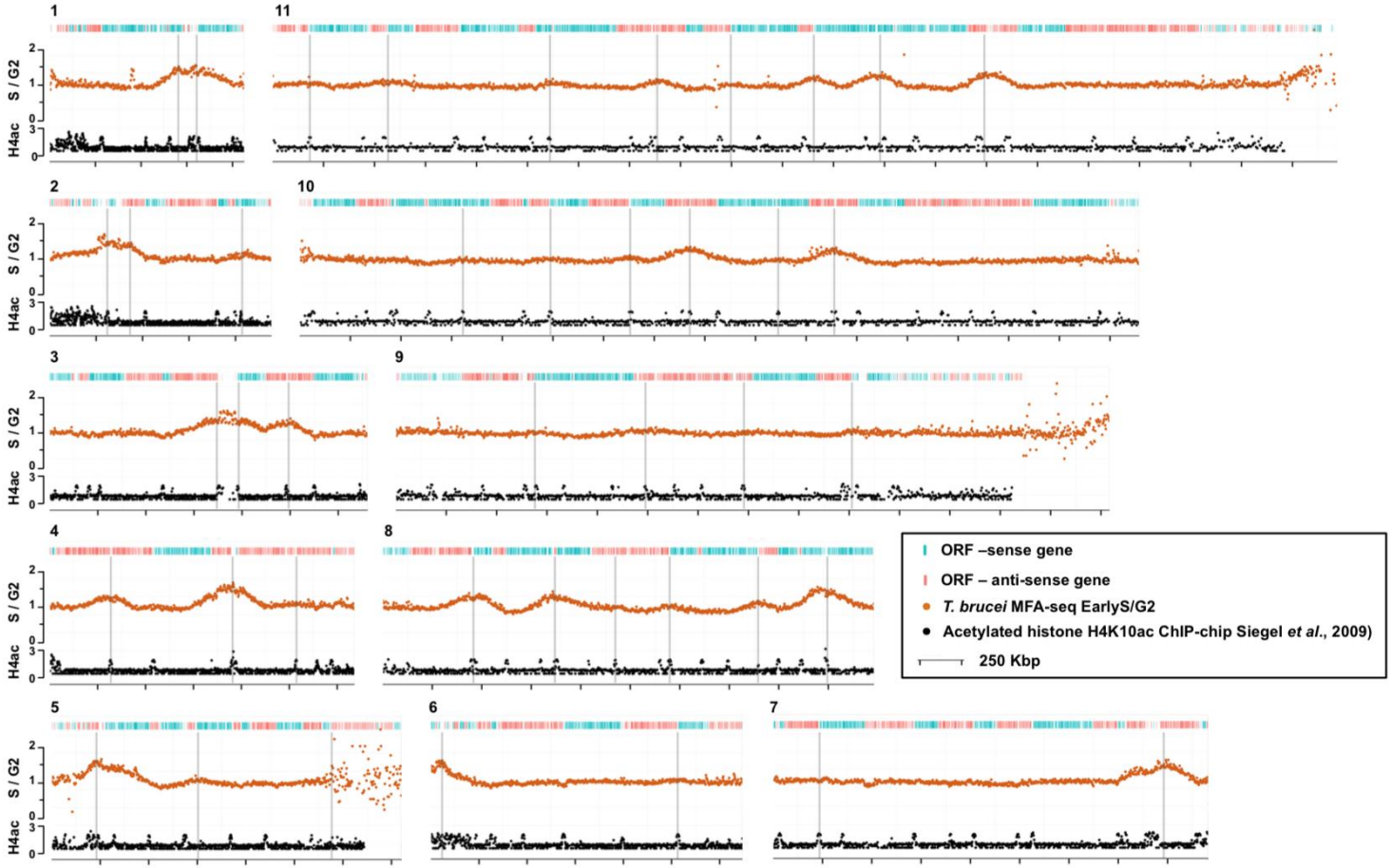


Figure 5.7. Origins of replication in early S phase *T. brucei* procyclic form cells.
Graphs represent the distribution of active origins of replication per *T. brucei* megabase chromosome (1 to 11), assessed by MFA-seq. Each graph shows the ratio of the coverage (read-depth) between early S phase and G2 phase samples, where each point (orange) represents the median S/G2 ratio (y-axis) per 2.5 Kbp section across the chromosome (x-axis). Each x-axis gap represents a 250 Kbp interval. The y-axis scale is the same for all graphs, but the legend is only shown in the ones in the far left. All graphs are made to scale according to each chromosome size. At the top of each graph a track is represented with the different genes in the chromosome, in blue the open reading frames (ORF) that are transcribed from the left to the right, and in red the genes transcribed from right to left. Below the MFA-

seq profiles, a track is shown with the histone H4K10ac-enriched sites (Siegel *et al.*, 2009) (represented simply as H4ac), demarking the different polycistronic transcription units. Each individual graph was generated by Dr Nicholas J. Dickens as described in the materials and methods Chapter 2.

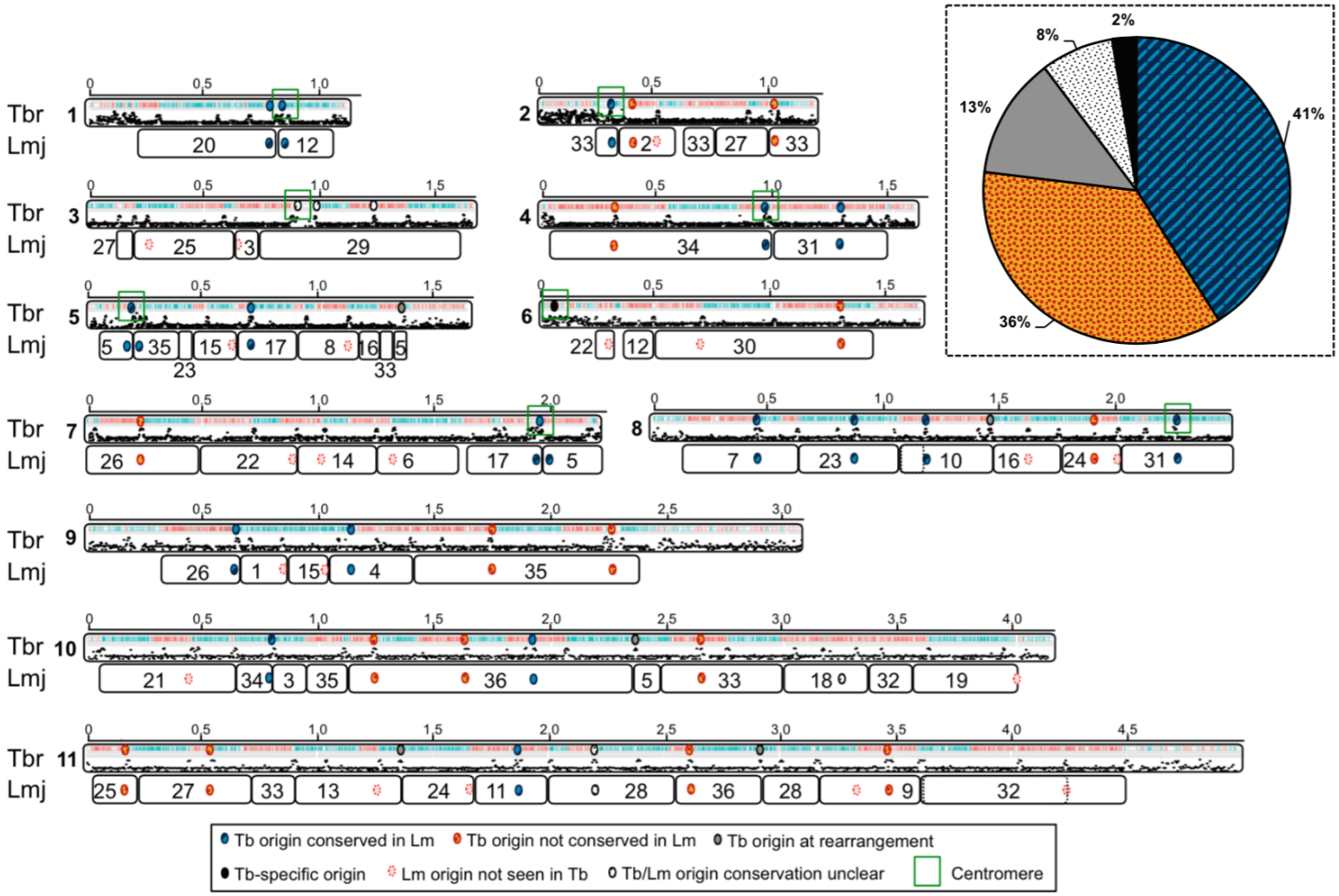


Figure 5.8. *T. brucei* origin conservation in *L. major*.

A synteny block diagram based on the one in (El-Sayed *et al.*, 2005b), and reproduced in Figure 5.2, is shown. Here, the 11 megabase chromosomes of *T. brucei* are represented by the ORF track (genes transcribed from left to right are shown in blue, and genes transcribed right to left are shown in red) and histone H4K10ac-binding sites (shown in black) (Siegel *et al.*, 2009). The sizes of the chromosomes are represented by intervals of 0.5 Mbp. *T. brucei* chromosomes are labelled as Tbr1 to Tbr11. Regions of synteny between the *T. brucei* 11 chromosomes and the 36 *L. major* chromosomes (Lmj1 to Lmj36) are represented by rectangles containing the *L. major* chromosome number below the *T. brucei* chromosome. In some cases, the whole *L. major* chromosome is syntenic with a region of a *T. brucei* chromosome, while others are fragmented. For simplicity, the orientation of the synteny blocks is not shown. Conservation of replication origins throughout syntenic regions is shown as follows (and summarised in the pie chart, inside dashed box): blue dot localises *T. brucei* origins conserved in *L. major*, while red/orange dots highlight non-conserved origins; grey dots localise *T. brucei* origins at rearrangement sites (*L. major* synteny blocks break point) that were lost in *L. major*; white/black dots depict regions where conservation is unclear; and the single black dot represents the only *T. brucei* origin that is specific to *T. brucei*, as it localises to the subtelomere of chromosome 6, that shows no synteny with *L. major*. Light pink dots show origins that appear to be *L. major* specific. The green boxes highlight the centromeres that have been mapped in *T. brucei* (Obado *et al.*, 2007).

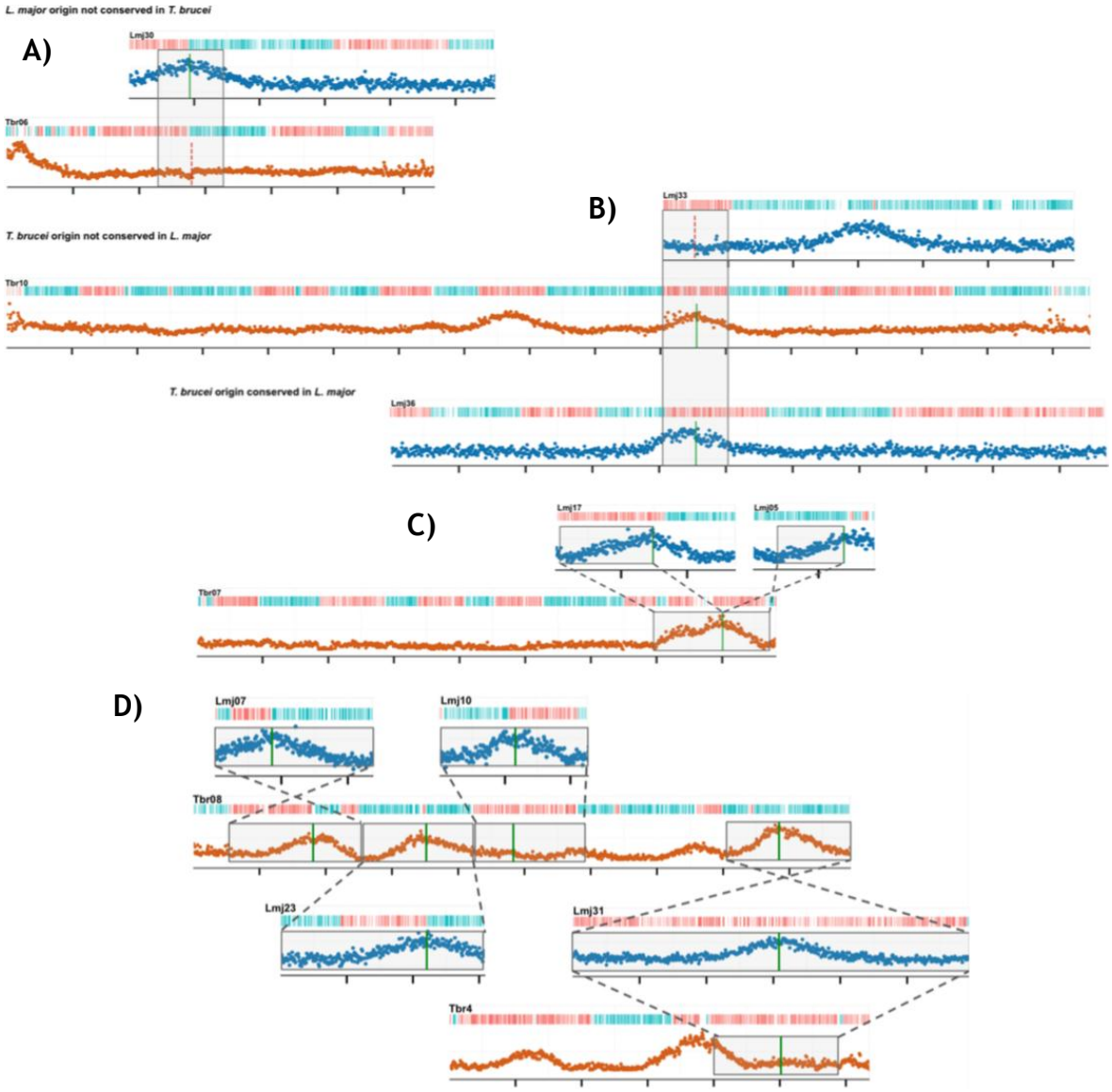


Figure 5.9. Examples of origin conservation between *T. brucei* and *L. major*.
Diagrams of some examples of origin conservation, or not, between the two parasites. All represented data is from early S samples. The MFA-seq graphs used have already been shown and described in Figure 5.4 (early S *L. major*) and Figure 5.7 (early S *T. brucei*), and the y-axes are not shown for simplicity. Syntenic regions are highlighted within the grey boxes, while conserved origins are marked with a green line, and non-conserved ones with a red dashed line. Grey dashed lines represent the direction of synteny between blocks. A) example of origin non-conservation, in which the same syntenic region displays origin activity in *L. major* chromosome 30 but not in *T. brucei* chromosome 6. B) example in which the same region in *T. brucei* (chromosome 10) is syntenic to two regions in two different *L. major* chromosomes (chromosomes 33 and 36). The origin in *T. brucei* is conserved in *L. major* chromosome 36 but not 33. C) one origin in *T. brucei* (chromosome 7), localised at a rearrangement site, is conserved in two *L. major* chromosomes (17 and 5). D) a complex example of origin conservation. Four of the origins in *T. brucei* chromosome 8 are conserved in *L. major*. In *T. brucei* these origins appear to have different strengths, suggesting that these are activated at different times during S phase; but the conserved origins in *L. major* appear to all have the same strength, suggesting activation at the same time. All x-axes graphs are to scale to chromosome size.

5.3 Origins of DNA replication in *Leishmania mexicana*

To date, no eukaryote has been reported to have a single origin of replication per chromosome, and thus the present dogma states that eukaryotic chromosomes are replicated from multiple origins that can be activated at different times during S phase (reviewed in Leonard and Mechali, 2013). Although all evidence here shown supports the existence of a single origin per chromosome in *L. major*, in contrast with *T. brucei* where multiple origins are detected, we sought to test whether origin singularity is common to other *Leishmania* species. Like *L. major*, *L. mexicana* is also a causative agent of cutaneous leishmaniasis. The two species, estimated to have diverged around 16 million years ago (Lukes *et al.*, 2014), have highly syntenic genomes of similar sizes (32-33 Mbp), though these differ in composition, since *L. mexicana* possesses 34 chromosomes (Britto *et al.*, 1998; Rogers *et al.*, 2011), in contrast with the 36 found in *L. major* (Ravel *et al.*, 1998; Ivens *et al.*, 2005). Two of the *L. mexicana* chromosomes (8 and 20) appear to have resulted from the fusion of two smaller *L. major* chromosomes (29 plus 8, and 36 plus 20, respectively), or the larger *L. mexicana* chromosomes evolved into the smaller *L. major* molecules by fission. Thus, mapping of replication origins in *L. mexicana*, in particular in chromosomes 8 and 20, provides a test of origin singularity in *Leishmania*.

5.3.1 Mapping the origins of replication in *L. mexicana*

L. mexicana strain U1103 promastigotes were used. The cell line was maintained and cell cultures were prepared by Dr Amy Goundry (Jeremy Mottram's laboratory). Cells were prepared, sorted, sequenced, and analysed as described for the *L. major* samples in section 5.2.1. In this case, the reference genome for *L. mexicana* strain U1103 (Rogers *et al.*, 2011), retrieved from TriTrypDB version 6.0, was used. The results were represented as mentioned, with the exception that no histone-binding data was included, as it is not available for this species.

As for *L. major*, a single origin was detected in each *L. mexicana* chromosome in early S phase, and no further origins were detected later in S phase (Figure 5.10 and Figure 5.11, respectively). In all cases, with two exceptions (below), the origins detected in *L. mexicana* localised to the same SSRs to which the origins

mapped in *L. major*, showing that there is an almost complete conservation in origin location between these two species (Figure 5.12).

The two exceptions were chromosomes 8 and 20. In each of these chromosomes, a single origin was detected in both early and late S phase samples (Figure 5.10 and Figure 5.11, red boxes), although each of these chromosomes is syntenic with two *L. major* chromosomes, each with a single origin (Figure 5.12, red boxes). In *L. mexicana* chromosome 8, the origin from chromosome 8 of *L. major* was maintained while the one from 29 was not, and in *L. mexicana* chromosome 20, the origin from chromosome 20 of *L. major* was kept while the one from 36 was not (Figure 5.12, red boxes). Assuming that these *L. mexicana* chromosomes result from the fusion of the *L. major* ones, it appears that origin singularity was maintained in both cases, with origin activity retained at one SSR and suppressed at the other. Alternatively, if the *L. mexicana* chromosomes split into two molecules, an inactive SSR was activated in one of the resulting *L. major* chromosomes. Further analysis of these regions is discussed in sections 5.3.2 and 5.4.

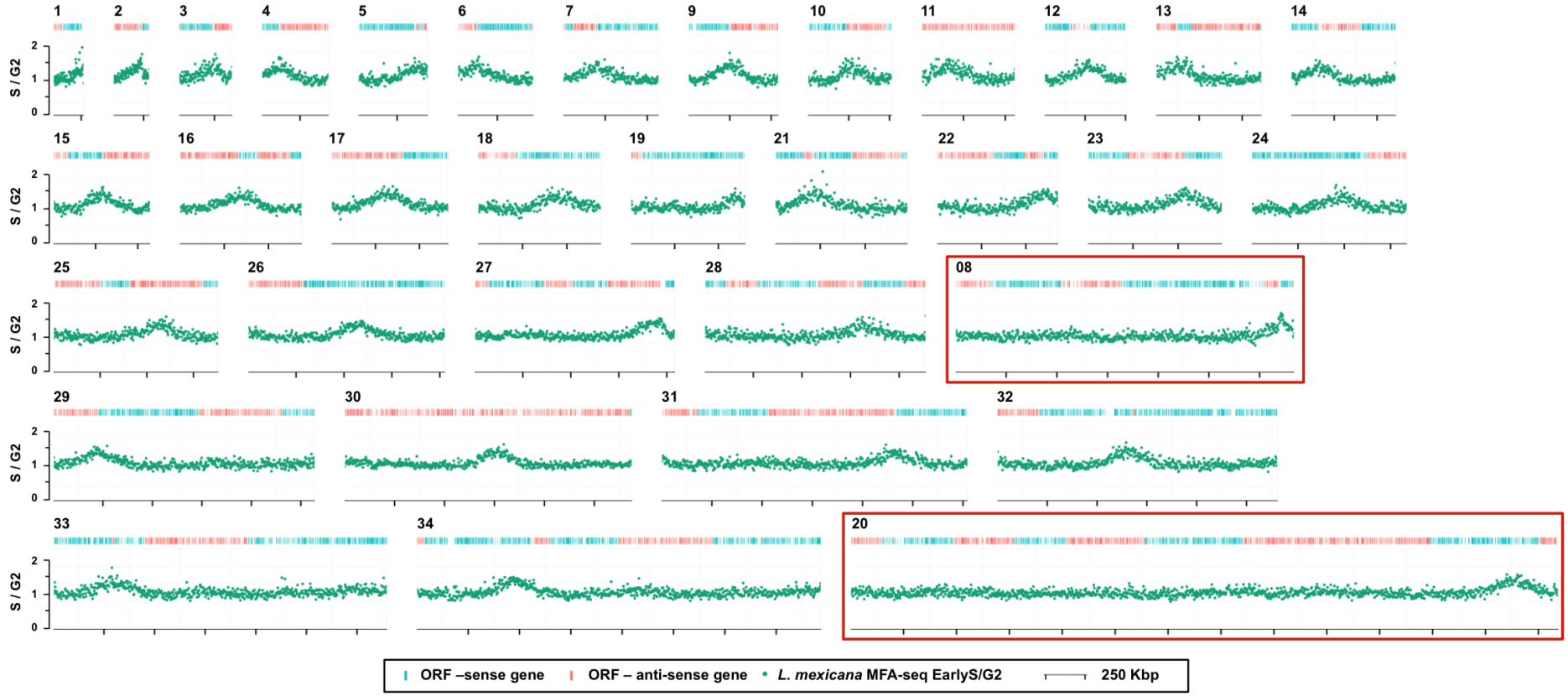


Figure 5.10. Origins of replication in early S phase *L. mexicana* promastigote cells.
Graphs represent the distribution of active origins of replication per *L. mexicana* chromosome (1 to 34), assessed by MFA-seq. Each graph shows the ratio of the coverage (read-depth) between early S phase and G2 phase samples, where each point (green) represents the median S/G2 ratio (y-axis) per 2.5 Kbp section across the chromosome (x-axis). Each x-axis gap represents a 250 Kbp interval. The y-axis scale is the same for all graphs, but the legend is only shown in the ones in the far left. All graphs are made to scale according to each chromosome size. At the top of each graph a track is represented with the different genes in the chromosome, in blue the open reading frames (ORF) that are transcribed from the left to the right, and in red the genes transcribed from right to left. Chromosomes 8 and 20 are enclosed by a red box as these are the chromosomes each syntenic to two *L. major* chromosomes (29 and 8, 36 and 20, respectively). Each individual graph was generated by Dr Nicholas J. Dickens as described in the materials and methods Chapter 2.

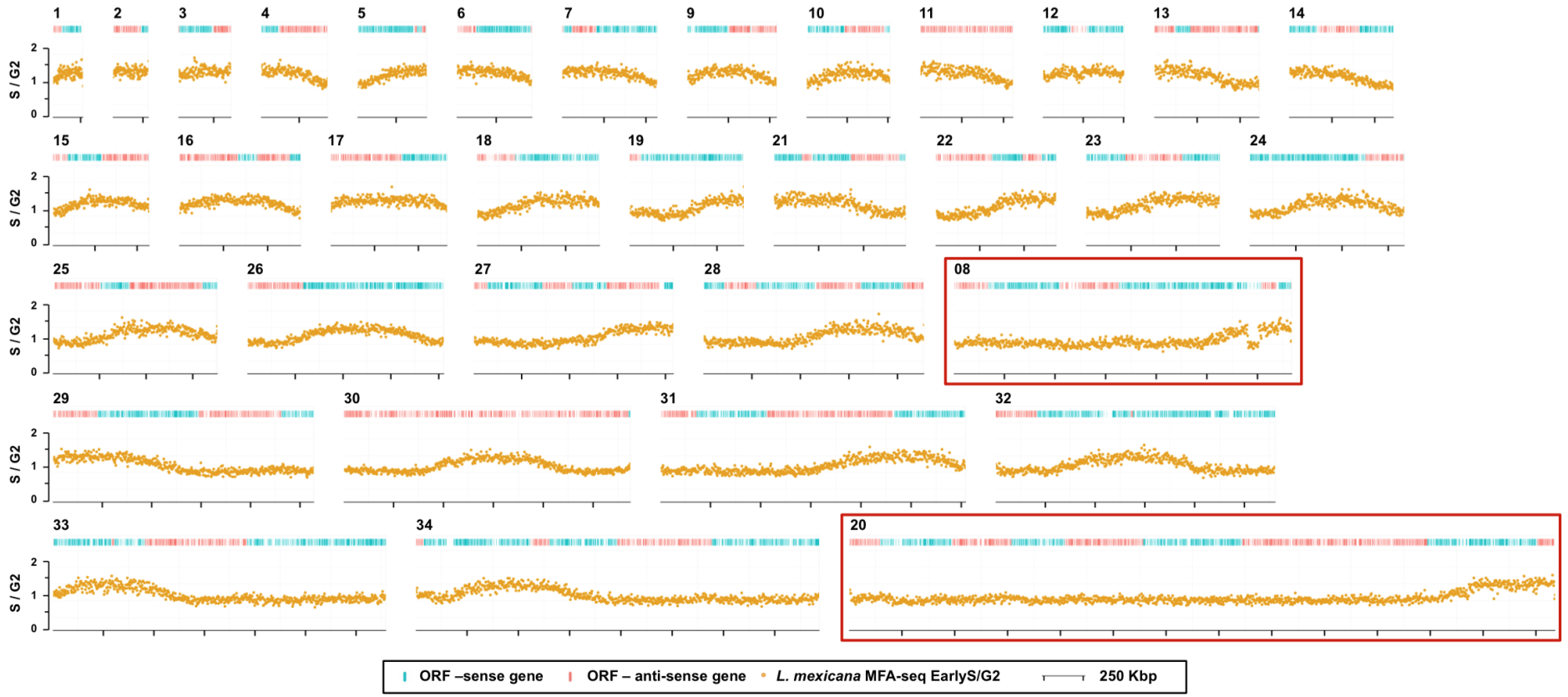


Figure 5.11. Origins of replication in late S phase *L. mexicana* promastigote cells.
Description as in Figure 5.10, with the exception that the MFA-seq data shows the ratio between late S and G2 phases.

Like in *L. major*, all peaks in *L. mexicana* appeared to have the same height (around 1.5 in early S phase) (Figure 5.10), suggesting again the absence of an origin activation timing programme, as seen in *T. brucei* (Tiengwe *et al.*, 2012a). Consistent with this, origin width was similar between *L. mexicana* origins, independent of the chromosome size, and identical to the values observed in *L. major*: peaks in smaller chromosomes such as 15 (0.57 Mbp) and 18 (0.69 Mbp) had a width of ~0.36 Mbp in early S phase, and of 0.53-0.6 Mbp in late S phase, like in larger chromosomes, such as 29 to 34 (1.32-1.97 Mbp), where peaks had a width of 0.36-0.43 Mbp in early S phase and about 0.78-0.89 Mbp in late S phase (Figure 5.10 and Figure 5.11). The broad similarity in peak width may suggest that both *Leishmania* species replicate at similar rates, albeit within the uncertainty of how comparable the S-phase sampling by FACS was. Nonetheless these values again suggest that the smaller chromosomes could be replicated from a single origin; for instance, data from late S phase suggest that chromosome 15 was already 93% replicated while chromosome 18 was 87%. In contrast, a single origin appears insufficient to replicate the larger chromosomes, such as chromosomes 29 and 34, which in the late S sample were only 59% and 43% replicated, respectively (Figure 5.10 and Figure 5.11). Although S phase in *L. mexicana* (Wheeler *et al.*, 2011) appears to be shorter than the one from *L. major* (Ambit *et al.*, 2011), lasting 2.9 h of its 7.1 h cell cycle, at a fork rate similar to the one reported for other eukaryotes, of 2 Kbp.min^{-1} (reviewed in Masai *et al.*, 2010; Mechali, 2010), bi-directional replication forks would be able to replicate only about 0.69 Mb of DNA, which would be enough to completely replicate chromosomes 1 to 19, but would not be enough to replicate the larger ones. Even if *L. mexicana* replicates at a faster rate, such as has been described for *T. brucei* PCF cells (3.7 Kbp.min^{-1} rate) (Calderano *et al.*, 2015), it would still be insufficient, replicating a maximum of 1.28 Mbp per origin. Again, this would not be enough to completely replicate the larger chromosomes, such as 30 to 34 (1.44 to 1.96 Mbp), and chromosome 20 (3.3 Mbp).

Together, this suggests that origin usage in *L. mexicana* is highly similar to *L. major*, and both greatly different from *T. brucei*, supporting the idea that *Leishmania* parasites may employ a different strategy from *T. brucei* to ensure the complete replication of their genomes.

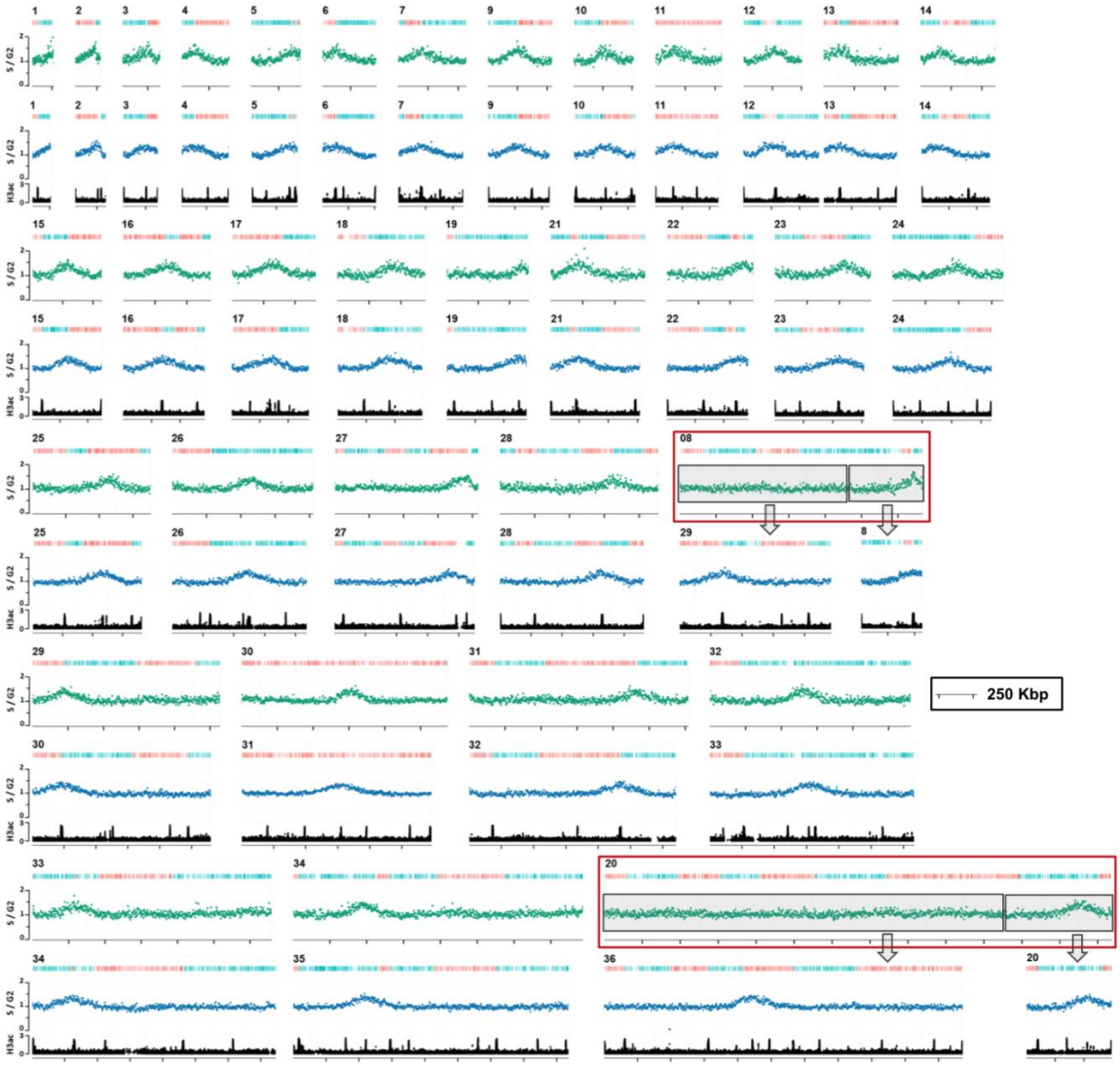


Figure 5.12. Comparison of the origins of replication in *L. major* and *L. mexicana*. Individual graphs of the MFA-seq data (S/G2 ratio) from early S samples of *L. mexicana* (green, shown in Figure 5.10) and *L. major* (blue, as shown in Figure 5.4) are shown. Graphs are shown according to synteny between the two genomes, with the chromosomes of *L. mexicana* on top and the respective syntenic *L. major* chromosome below. *L. mexicana* chromosome 8 (red box) is syntenic with both *L. major* chromosomes 29 and 8, while chromosome 20 (red box) is syntenic with *L. major* chromosomes 26 and 20; the syntenic regions in the *L. mexicana* chromosomes are highlighted by a grey box, and the arrows indicate the respective syntenic chromosome in *L. major*.

5.3.2 Confirmation of a single origin in *L. mexicana* chromosomes 8 and 20 by MFA-qPCR

In order to test the validity of the MFA-seq data further we tested the prediction of a single origin in *L. mexicana* chromosomes 8 and 20 by MFA-qPCR (Figure 5.). Employing the qPCR strategy described in (Tiengwe *et al.*, 2012a), and as detailed in section 5.2.2, gDNA from *L. major* and *L. mexicana* early S and G2 samples was analysed by qPCR. In this case, primers were designed in regions that covered the origins detected in the *L. major* chromosomes 8 (primers N3,

P2, P3 and Q1) and 29 (primers N1, M2, M4 and O2), 20 (U1, V1 and W1) and 36 (R2, S1, S3 and T1), and the corresponding syntenic regions in the *L. mexicana* chromosomes 8 (O2, M4, M2, N1, N3, P2, P3 and Q1) and 20 (R2, S1, S3, T1, U1, V1 and W1), respectively (Figure 5., A). Primer sequences are shown in the materials and methods Chapter 2, Table 2-10. All primers were tested for efficiency (data not shown) as described in section 5.2.2, with all pairs being close to 100%, and with a maximum difference of 12%. Primers pair R3 was used as the "reference gene" for normalisation, and the corresponding G2 phase sample as the calibrator for the $\Delta\Delta C_t$ analysis (Livak and Schmittgen, 2001). As shown in Figure 5., B), only one origin was detected in each of the *L. mexicana* chromosomes, with the MFA-seq and MFA-qPCR S/G2 ratios matching well. In the syntenic regions in which MFA-seq predicted origin activity in *L. major*, but no activity in *L. mexicana*, the MFA-qPCR S/G2 ratios confirmed this prediction. Together, these results corroborate the MFA-seq analysis (Figure 5.10), and support the existence of a single origin per *Leishmania* chromosome.

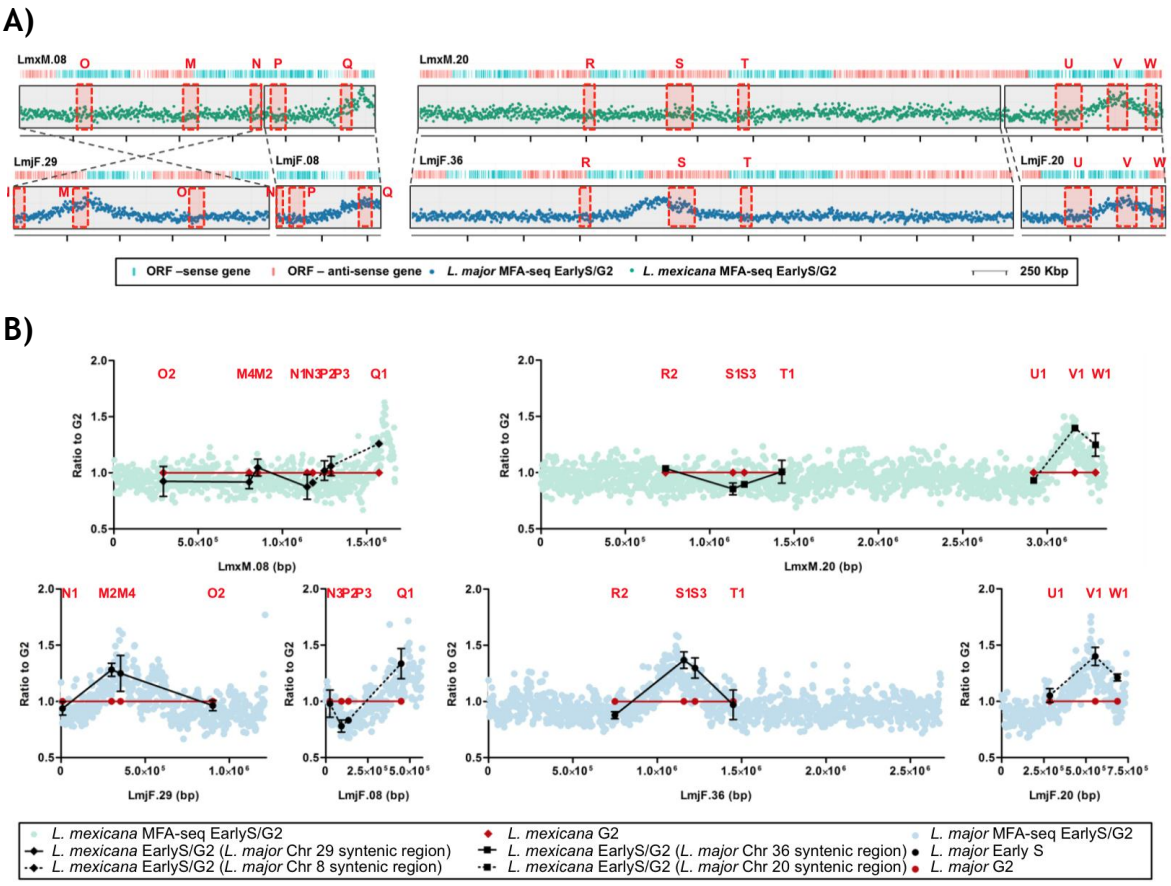


Figure 5.13. *L. mexicana* and *L. major* fusion/fission chromosomes.
A) MFA-seq graphs of *L. mexicana* chromosomes 8 (LmxM.08) and 20 (LmxM.20), over the corresponding *L. major* syntenic chromosomes 29 (LmjF.29) and 8 (LmjF.08), 36 (LmjF.36) and 20 (LmjF.20), respectively, as shown in Figure 5.12. The y-axes are not shown for simplicity. The grey shaded boxes highlight the synteny blocks, while the dotted lines show their orientation. Dashed lines demark the regions (loci O to W) chosen for the design of primers for MFA analysis by qPCR.

Primers pair R3 (within R locus) was used as the "endogenous gene" in the qPCR analysis by the $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001). B) Validation of the replication origins in the *L. mexicana* chromosomes 8 and 20, as well as in *L. major* chromosomes 29, 8, 36 and 20, by MFA-qPCR at a number of specific loci predicted to display origin activity in *L. major*. At each loci (O2 to W1 refer to the primers pairs used per loci), the values from the G2 sample (red) are set as 1 (completely replicated DNA), while the relative quantity of the early S phase over G2 is shown (black). Each data point (black) represents the mean of three qPCR runs (n = 3) of the same sample, each run performed with three experimental replicates. The error bars depict the standard error of the mean of the three independent qPCR runs. For ease of comparison, the MFA-seq data was plotted in the background. All data is represented to their position in the chromosome, as the x-axis is represented to scale, depicting the chromosome's length in bp.

5.4 Do any sequence features define an origin in kinetoplastids?

Similar to *T. brucei* (Tiengwe *et al.*, 2012a), analysis of the mapped *L. major* and *L. mexicana* origins of replication (performed by Dr Nicholas J. Dickens, data not shown) failed to find any consensus DNA sequence, motif or pattern, defining the sites that become origins of replication, perhaps suggesting that, like other eukaryotes, origins of replication in kinetoplastids are defined by chromatin structure, context and status (reviewed in Masai *et al.*, 2010; Mechali, 2010; Leonard and Mechali, 2013). In line with this, a common feature between *T. brucei*, *L. major* and *L. mexicana* origins is their localisation at SSRs, irrespective of whether they are divergent (dSSRs) or head-to-tail (h-t SSRs), both demarcated by acetylated histones and known RNA polymerase II transcription starting sites (Thomas *et al.*, 2009; Siegel *et al.*, 2009), or convergent (cSSRs), where transcription terminates. However, origins are not found at all SSRs, and thus we asked whether there are differences between origin-active SSRs and non-origin SSRs. To answer this, the length of each origin-active SSR was assessed by measuring of the distance (in Kbp) between the most proximal gene ORFs at the SSR: between the predicted start codons of the genes at a dSSR; the two stop codons of the genes at a cSSR; and the stop and start codons of the two genes at a h-t SSR. The length of multiple non-origin SSRs was also measured, and plotted alongside the origin-active SSR data, in order to infer statistically significant differences (Figure 5.14, A). Individual measurements are shown in the appendices, section 7.9. Because the distributions of the *L. major* and *L. mexicana* origin-active and non-origin SSR data sets failed to pass normality tests (D'Agostino and Pearson normality test, Shapiro-Wilk normality test, and the Kolmogorov-Smirnov normality test), a non-parametric test was preferred. For

the three sets of data, *T. brucei*, *L. major* and *L. mexicana*, the two unpaired groups (origin-active SSR and non-origin SSR) were compared using the Mann-Whitney test, one-tailed, with statistical significance defined at a p value < 0.05. In *T. brucei*, no difference in length was seen between origin-active and non-origin SSRs (Figure 5.14, A), orange graph), suggesting that they could not be distinguished by size. In striking contrast, in both *Leishmania* species there was a clear difference ($p < 0.0001$) between these two types of SSRs: the distance between the most proximal genes in SSRs containing an origin was significantly larger than the non-origin SSRs (Figure 5.14, A), blue and green graphs). Separating the SSRs into the three types (dSSR, cSSR or h-t SSR) and repeating the analysis (due to the small number of data points in some groups, normality was assumed for the data distribution, and the analysis was performed using one-way ANOVA, un-matched, with statistical significance defined at a p value < 0.05) showed that the size differences in *Leishmania* between origin-active and inactive SSRs are independent from the configuration of transcription direction around them (Figure 5.14, B).

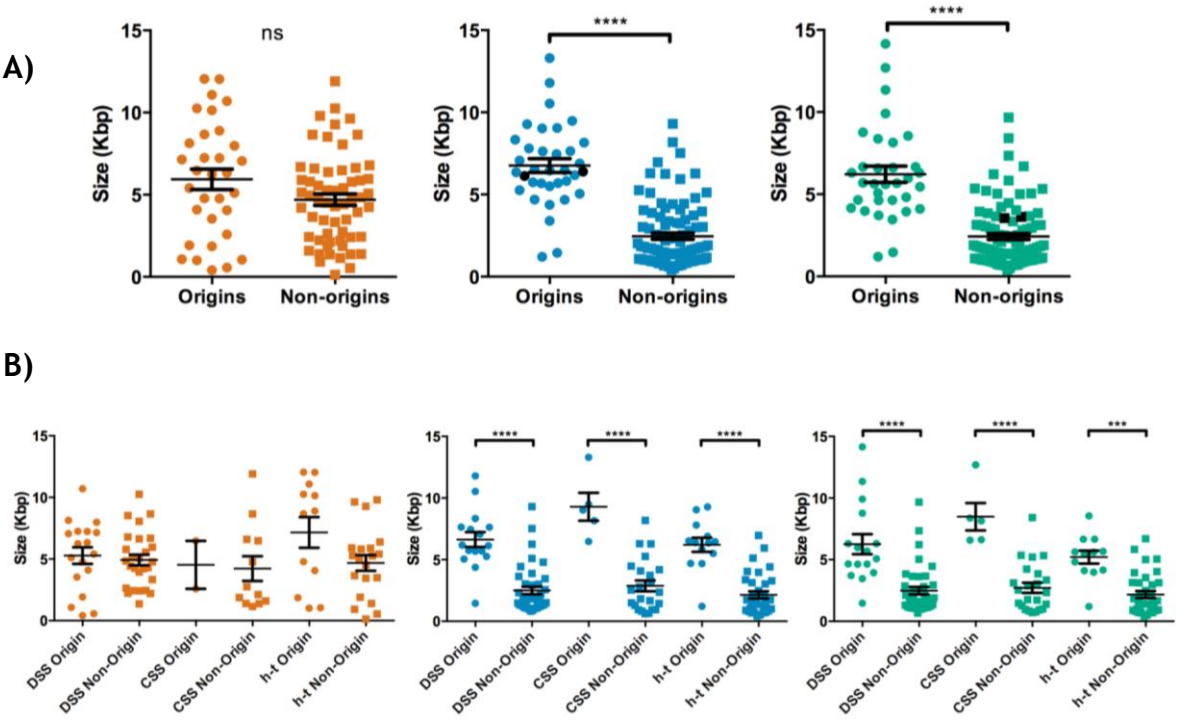


Figure 5.14. Size of origin and non-origin containing strand switch regions.
A) Scatter plot showing the lengths of each origin SSR (circles) and of multiple non-origin SSRs (squares). The graph on the left represents the data (in orange) from *T. brucei*; the graph in the middle shows the data (in blue) from *L. major*; while the graph on the right shown the data (in green) from *L. mexicana*. In the *L. major* and *L. mexicana* graphs, black points represent the origin-SSR in *L. major* chromosomes 29 and 36 that are non-origin SSRs in *L. mexicana* chromosomes 8 and 20, respectively. The mean is shown as a black horizontal line, and the error bars represent the standard error of the mean. Statistical significance was inferred with the non-parametric Mann-

Whitney test, one tailed, with a p value threshold of < 0.05. (ns) refers to non-significant, while (****) refers to a p value < 0.0001. B) Same data as in A) but represented per SSR type: dSSR, cSSR and h-t SSR. As in A), the graph on the left shows the data from *T. brucei* (orange), the one in the middle refers to *L. major* (blue), while the one on the right represents the data from *L. mexicana* (green). The mean is shown as a black horizontal line, and the error bars represent the standard error of the mean. Statistical significance was inferred with the parametric one-way ANOVA test, un-matched, multiple group comparison, with a p value threshold of < 0.05. (****) refers to a p value < 0.0001.

In the previous sections it was shown that in *L. mexicana* chromosome 8, only the origin from the region syntenic with *L. major* chromosome 8 was kept, while the one referring to the region of *L. major* chromosome 29 was lost; and in *L. mexicana* chromosome 20, the origin from *L. major* chromosome 20 was kept, while the one in *L. major* chromosome 36 was lost (Figure 5.12). A closer analysis of these regions showed that the general SSRs trend was reflected in these specific sites: in *L. major*, the origin-active SSRs have a length of 6.13 Kbp (chromosome 29) and 6.4 Kbp (chromosome 36) (Figure 5.14, A, middle graph, black points), while the corresponding syntenic, non-origin SSRs in *L. mexicana*, are 3.63 Kbp (chromosome 8) and 3.55 Kbp (chromosome 20) in size (Figure 5.14, A, far right graph, black points). These differences are shown in more detail in Figure 5.15.

In both *T. brucei* and *Leishmania*, origins localise to a subset of SSRs, and appear to be maintained during differentiation and prolonged culture (in *T. brucei*), and between strains or species. However, the data above suggests that the origins of replication in *Leishmania*, but not in *T. brucei*, localise to a specific subset of SSRs that can be distinguished by size from non-active SSRs. This lends weight to the argument that replication in *Leishmania* initiates from only a single origin in each chromosome, as these are the SSRs in each case whose size differs from all others. The observation that the SSRs in *L. mexicana* chromosomes 8 and 20 that do not show origin activity in *L. mexicana* are smaller than the syntenic SSRs in *L. major*, where origin activity is observed, gives support to the argument that SSR features are altered as these loci move from an origin-active to inactive (or vice versa) status.

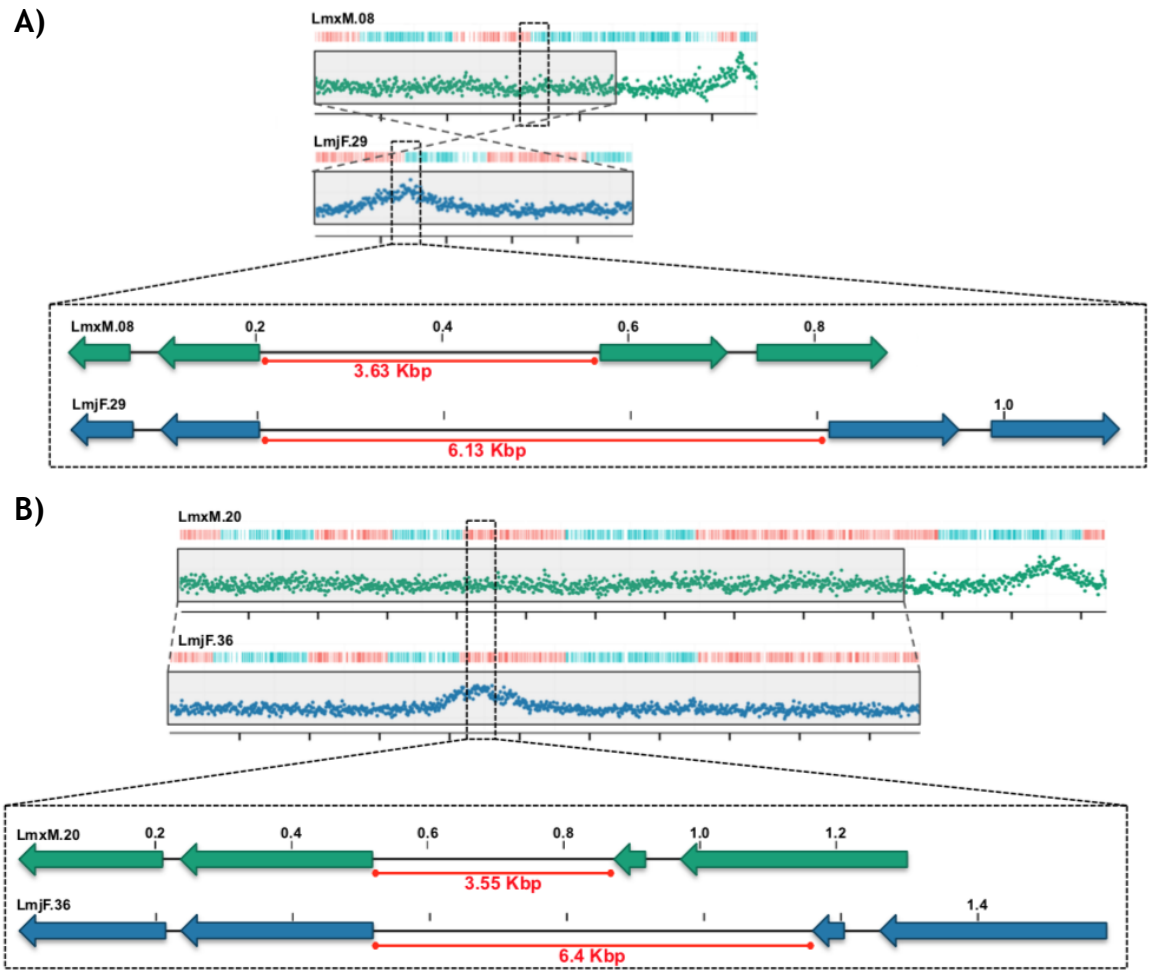


Figure 5.15. Syntenic regions and *L. mexicana* and *L. major* that display different origin activity.
Diagrams representing the two regions of *L. mexicana* chromosomes 8 (A) and 20 (B) that do not show origin activity but correspond to syntenic regions in *L. major* chromosomes 29 (A) and 36 (B), respectively, that show origin activity. As in Figure 5., the blocks of synteny are highlighted within the grey boxes. Dashed boxes identify the syntenic SSR, which are shown in more detail below (each gap represents 0.2 Mbp). The distance between the two most proximal ORFs within the SSR is highlighted in red. Distances are shown in Kbp.

5.5 Discussion

Presently, it is universally accepted that eukaryotic chromosomes are invariably replicated from multiple origins of replication (reviewed in Costa *et al.*, 2013): origin multiplicity in eukaryotic genomes is needed since complete, timely duplication of their large linear chromosomes is only possible with the establishment of various replication forks at different loci distributed throughout the chromosomes (reviewed in Leonard and Mechali, 2013; Mechali, 2010). In addition, although all pre-RC-bound sites at the end of G1 phase are deemed potential origins, only a small subset of these are activated per cell cycle, with

the remaining acting as a dormant pool of origins than can be activated during S phase if needed (reviewed in Blow *et al.*, 2011; McIntosh and Blow, 2012). Nevertheless, the origins that are activated in every cell cycle, the number of which appears to be dependent on chromosome size, do not do so at the same time, as their activation is governed by a specific timing programme of origin activation during the cell cycle, with some origins being activated earlier (early firing) than others (late firing) (reviewed in Mechali, 2010; Rhind and Gilbert, 2013). With the exception of budding yeast, where origins are defined by consensus sequences (Bell and Stillman, 1992), eukaryotic origins of replication appear to be defined by contextual cues, such as chromatin structure, status and environment, including proximity to promoter elements, DNA accessibility and nucleosome occupancy (reviewed in Costa *et al.*, 2013). Finally, DNA replication and transcription are intimately connected processes (reviewed in Helmrich *et al.*, 2013; Barlow and Nussenzweig, 2014).

Mapping of the origins of replication in *T. brucei* PCF cells (Tiengwe *et al.*, 2012a) revealed that this parasite appears to conform to the above 'universal' features of eukaryotic DNA replication: multiple origins of replication are detected per chromosome; origin number relates to chromosome size; the number of TbORC1/CDC6-binding sites outnumbers the activated origins, thus showing that only a subset of potential origins are activated per cell cycle; the detected active origins appear to have different strengths, suggesting that *T. brucei* origins of replication are activated at different times during the cell cycle, and thus an activation timing programme is present; and no specific sequences were found to define either TbORC1/CDC6-binding sites or active origin sites. In addition, and perhaps to an even greater extent than in other eukaryotes, origins of replication overlap with regions of transcription initiation and termination, suggesting that DNA replication and transcription are interconnected in *T. brucei*. Furthermore, a recent report measured replication rate in *T. brucei*, and provided very preliminary evidence for non-active TbORC1/CDC6 binding sites acting as dormant origins, since one such site appears to be activated after hydroxyurea (HU) treatment (Calderano *et al.*, 2015).

Though believed to have diverged around 250 million years ago (Lukes *et al.*, 2014), the genomes of *T. brucei* and *L. major* are highly conserved and syntenic, and share the same unusual gene expression organisation in large polycistronic transcription units (El-Sayed *et al.*, 2005b). As origins of replication in eukaryotes appear to be defined by chromatin cues and genomic context, we sought to analyse the origins of replication in *L. major*, since it might be predicted that kinetoplastid gene expression constrains the potential location of origins (Tiengwe *et al.*, 2012a). The result of origins mapping that was obtained, however, was unprecedented: only a single origin could be mapped in each *L. major* chromosome. Further analysis of another *Leishmania* species, *L. mexicana*, revealed the same pattern: a unique origin per chromosome. If this organisation of replication is correct, it overturns the accepted orthodoxy of origin multiplicity in eukaryotes, and the suggestion that origin singularity is an exclusive feature of prokaryotic genomes, which are small enough to be replicated from a single origin (reviewed in Masai *et al.*, 2010; Costa *et al.*, 2013). Moreover, analyses of *L. major* and *L. mexicana* replication patterns by MFA-seq around the mapped origins, suggests that all origins are activated at the same time and replicate at the same rate in every chromosome. As a result, these findings indicate that an origin activation timing program is absent in *Leishmania*, in contrast to *T. brucei* (Tiengwe *et al.*, 2012a) and all characterised eukaryotes to date (reviewed in Rhind and Gilbert, 2013). Like in *T. brucei* (Tiengwe *et al.*, 2012a), all *Leishmania* origins co-localised with SSRs. Indeed, a remarkably high level of origin location conservation was seen: 40% of origins were conserved between the two parasites, in contrast with only 12-21% between *S. cerevisiae* and *Lachancea waltii*, two yeasts believed to have diverged from each other at a similar time to *T. brucei* and *L. major* (Di Rienzi *et al.*, 2012). However, analysis comparing the size of origin-active and inactive SSRs revealed a striking difference between *Leishmania* and *T. brucei*. In the latter, there is no difference between origin-active and inactive SSRs, and thus SSR size in itself is not defining which origins are activated. In contrast, in both *L. major* and *L. mexicana* it is clear that the origin-containing SSRs are considerably larger than the ones with no origin activity. Indeed, SSRs with origin activity in *L. major* are considerably larger than syntenic SSRs in *L. mexicana* that do not display any origin activity. These data suggest that in *Leishmania*, origins of replication are activated only at specific SSRs that, most

likely amongst other features, are larger. How this is reflected in factor recruitment to the SSRs is unknown; for instance, does the *Leishmania* pre-RC machinery bind to all SSRs or only to the distinctive origin-active SSRs? Nonetheless, the fact that SSRs size is equivalent at origin-active and inactive SSRs in *T. brucei* is consistent with TbORC1/CDC6 binding to all these sites, and the potential that some act as dormant origins. In *Leishmania*, this flexibility may be absent, with no possibility to initiate replication via an origin of replication at any other discrete sites than the single mapped SSRs in each chromosome. In *T. brucei*, gene density is higher than in *L. major*, where the average length between polycistronic transcription units has been shown to be almost twice the size than in *T. brucei* (El-Sayed *et al.*, 2005b). Perhaps this difference in gene density may account for the possibility of *Leishmania* parasites to use the length of SSRs as an origin-defining feature, but not *T. brucei*, which might have developed origin-activating mechanisms common to other eukaryotes.

The observation of a single origin per chromosome in *Leishmania* raises an important question: how is the parasite capable of replicating its genome within S phase? Data from late S samples suggest that it is, most likely, not possible to complete the replication of all chromosomes from one site. In *L. major* and *L. mexicana*, chromosomes size rages from 0.27-2.7 Mbp and 0.27-3.3 Mbp, respectively. Data from late S phase samples suggest that a single *L. major* origin replicated between 0.5 and 0.83 Mbp of DNA, while in *L. mexicana* 0.5 to 0.89 Mbp had been replicated. Although the replication rate has not been measured for any *Leishmania* species, the assumption that these might have a similar rate to other eukaryotes and *T. brucei* (2-3.7 Kbp.min^{-1}), and a 3-4 h S phase (Ambit *et al.*, 2011; Wheeler *et al.*, 2011), suggests that *L. major* and *L. mexicana* origins would only be able to replicate a maximum of 0.7-0.96 Mbp (rate of 2 Kbp.min^{-1}), similar to the observations made for the late S phase data, or 1.28-1.77 Mbp (rate of 3.7 Kbp.min^{-1}). In making these estimates, it should be noted that this assumes that replication fork movement is unperturbed, and that origin location is central in the chromosome. The former assumption is unlikely, as all organisms use recombination and by-pass mechanisms to overcome lesions. The latter is demonstrably incorrect, with the most striking example of a non-central origin seen in *L. mexicana* chromosome 20; here, the single mapped

origin is close one telomere and therefore replication needs to traverse most of the ~3.3 Mb molecule. It seems very likely that although a single origin is most likely able to efficiently direct the complete replication of the smaller chromosomes of *L. major* and *L. mexicana*, a unique origin is not sufficient to allow the replication of the larger chromosomes. In fact, in *T. brucei* PCF cells the chromosomes can only be completely replicated because various origins are activated per chromosome and the replication rate of 3.7 Kbp.min^{-1} is slightly faster than other eukaryotes (Calderano *et al.*, 2015), as each origin is only able to replicate around 0.66 Mbp per S phase, which lasts 1.5 h (Woodward and Gull, 1990). Therefore, the main question arising from this study is: how do *Leishmania* parasites completely replicate their larger chromosomes?

One hypothesis is that there are other origins of replication per chromosome, but that the MFA-seq approach is not sensitive enough to detect them, as it maps predominant origins in a population. For instance, if a dominant origin is activated in every cell per cell cycle (constitutive origins), but others are activated randomly throughout the chromosome in different cells (flexible origins), these origins will not be identified by MFA analysis, as these are not activated in a large proportion of the cell population. It will be crucial to map the binding sites of *Leishmania* ORC1/CDC6 throughout the genome in order to address this. This will answer whether there are more potential origins in the *Leishmania* chromosomes, and if so, whether these can be activated. Moreover, mapping *L. mexicana* ORC1/CDC6 to the fusion chromosomes, 8 and 20, could be informative, as these molecules possess SSRs that may have recently lost origin activity. Nonetheless, indirect data from various studies on chromosome stability support the existence of a single origin of replication per *Leishmania* chromosome (Figure 5.16). Fragmentation of chromosome 5 of *L. donovani* and of chromosome 23 of *L. tarantolae* by telomere-associated chromosome fragmentation into two fragments each, showed that in both cases one fragment was stably maintained while the other was lost after several generations (Tamar and Papadopoulou, 2001). This suggested that there were elements, most likely centromere-like, in these two maintained fragments that conferred mitotic stability, thus allowing their retention and segregation during cell divisions, while these were absent from the fragments that were lost during growth (Tamar and Papadopoulou, 2001). However, unlike in *T. cruzi* (in which similar

fragmentation assays (Obado *et al.*, 2005; Obado *et al.*, 2007) led to the functional mapping of the centromeres in chromosomes 1 and 3) and *T. brucei*, where the mapping of topoisomerase II activity after etoposide treatment allowed the biochemical mapping of centromere location , the implementation of such approach failed to identify *L. major* centromeres (Obado *et al.*, 2007). It therefore remains unknown where centromeres are localised in the *Leishmania* genome. Data from the mapping of the origins of replication in *T. brucei*, however, suggest that the strongest origins co-localise with the mapped centromeres in chromosomes 1 to 8 (Tiengwe *et al.*, 2012a), similarly to other eukaryotes, in which the centromeres have been shown to be the earliest replicating regions in the genome (reviewed in Rhind and Gilbert, 2013; Masai *et al.*, 2010). It is thus possible that the elements conferring mitotic stability in the chromosome fragments highlighted above (Tamar and Papadopoulou, 2001) include not just a potential centromere-like region, but also the origin of replication of that chromosome. Assuming that the origins of replication here mapped are conserved in other *Leishmania* species (*L. major* and *L. mexicana* genomes are highly syntenic with the *L. donovani* and *L. tarantolae* ones; data available on TriTrypDB), comparison of the chromosome fragmentation data from *L. donovani* and *L. tarantolae* with the origins mapped in both *L. major* and *L. mexicana* show that the stable chromosome fragments enclosed the region of the chromosomes containing the origin, whereas the unstable fragments did not, thus supporting the results of a single origin at least in chromosomes 5 and 23. Fragmentation of a duplicated region of chromosome 19 in *L. donovani* retrieved similar results (Dubessay *et al.*, 2001), as the fragment stably maintained by the parasites contained the syntenic region of the *L. major* and *L. mexicana* mapped origins, while the non-origin containing fragment was lost with growth. It was also shown that the “right end” of chromosome 1 of *L. major* is stably maintained, a region that corresponds with the origin here mapped (Dubessay *et al.*, 2002). It will be of interest to explore this further in *L. major* and *L. mexicana*, now that origin locations are known, either by chromosome fragmentation, as described in the abovementioned studies, or by cloning origin-active and non-origin SSRs into plasmids and asking whether these are maintained by the parasites in the absence of drug selection (already being performed in *L. major* by Dr Daniel Paape using a vector system from (Roy *et al.*, 2000). It will be perhaps of value to explore the *L. mexicana* origin-active and

non-active SSRs from the “fusion” chromosomes. Indeed, if it was possible, it would be interesting to artificially fuse two single origin *L. major* chromosomes, such as 8 and 29, recapitulating the proposed evolutionary history, and ask what effect the imposition of two origin-active SSRs has on their usage and on chromosome stability.

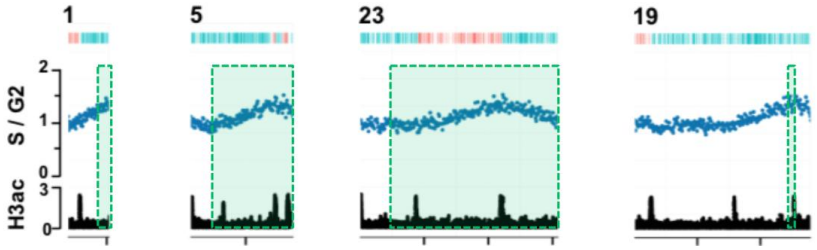


Figure 5.16. Summary of fragmentation assays.
Diagrams summarising data published from chromosome fragmentation stability in *Leishmania.* Green dashed box represents the section of the *L. major* chromosome syntenic to the fragments showed to be maintained in various *Leishmania* species. Chromosome 1 represents the data from (Dubessay *et al.*, 2002), study performed in *L. major*. Chromosome 5 summarises the finding in *L. donovani* and chromosome 23 from *L. tarantolae*, both from (Tamar and Papadopoulou, 2001). Chromosome 19 depicts the data from (Dubessay *et al.*, 2001), in *L. donovani*. The *L. major* plots are the same as in Figure 5.5, showing the origins detected by MFA-seq in early S. The reader is referred there for the legend in Figure 5.5 for more detail.

The evidence outlined above supports the existence of a single origin per *Leishmania* chromosome, although it is very unlikely that the parasites are able to completely replicate their larger chromosomes within S phase, at least by origin-dependent DNA replication. A plausible alternative hypothesis relies on the parasite finishing DNA replication, of perhaps mainly its large chromosomes, by origin-independent initiation of DNA replication. Such replication is well known to occur by recombination-dependent initiation of DNA replication (also known as break-induced DNA replication, BIR), which is, for example, an intrinsic part of the bacteriophage T4 life cycle, wherein early studies allowed the study of such process in bacterial and eukaryotic cells (reviewed in Anand *et al.*, 2013). Origin-independent initiation can take place when recombination (D-loops) or transcription (R-loops) intermediates are used to prime replication, being converted into replication forks, which are composed of the molecular machineries used in normal DNA replication, including the replicative helicase MCM2-7, Cdc45 and GINS complex, but excluding ORC (reviewed in Anand *et al.*, 2013). A recent example of recombination-dependent initiation of DNA replication comes from an archaeal organism, *Heloferax volcanii* (Hawkins *et al.*, 2013). *H. volcanii* possesses a main circular chromosome of 2.8 Mbp (similar in size to *L. major* and *L. mexicana* larger chromosomes), which is normally

replicated from 4 different origins, recognised by ORC1/CDC6. However, it was possible to generate viable cells in which all four origins were deleted (Hawkins *et al.*, 2013). Moreover, these cells grew faster, and presented a similar DNA profile to the wild type cells when analysed for DNA content by flow cytometry, suggesting that they had successfully replicated their DNA. MFA-seq analysis revealed that no defined origins were used, though the largely flat replication profile suggested widespread origin-independent initiation of DNA replication (Hawkins *et al.*, 2013). Attempts to generate origin-less cells deficient in one of the central players of homologous recombination, RadA (homologous to the bacterial RecA and the eukaryotic Rad51), showed that in these origin-lacking cells, the entire genome was being replicated by recombination-dependent initiation alone, apparently with no associated fitness cost. *H. volcanii* is highly polyploid, tolerating variations in genome copy number, and it has been suggested that this enables the origin-less cells to grow (Hawkins *et al.*, 2013). This contrasts with *oriC* deletion in bacteria, which though possible, is severely detrimental to growth (reviewed in Kogoma, 1997).

Leishmania species have been shown to be prone to chromosome aneuploidy, which appears to arise frequently and to be well tolerated (Rogers *et al.*, 2011; Sterkers *et al.*, 2011; Downing *et al.*, 2011; Lachaud *et al.*, 2014). For instance, in *L. major* strain Friedlin (used here) chromosome 31 (1.48 Mbp) is tetraploid, while in *L. mexicana* strain U1103 (used here) several chromosomes are triploid, or intermediate between monoploid and tetraploid (Rogers *et al.*, 2011). It is possible that *Leishmania* species' tendency to chromosome aneuploidy might be necessary to ensure complete replication of their genomes. In addition, *Leishmania* parasites are known to resort to gene amplification of specific loci as a response to the environment, through the generation of extra-chromosomal circular (episomes) or linear DNA molecules by homologous recombination between direct or inverted repeated sequences, respectively, which were recently shown to be widespread throughout the *L. major*, *L. infantum*, and *L. braziliensis* genomes (Ubeda *et al.*, 2014). These repeats are mostly non-coding sequences and present in low-copy number, being used as a platform for the amplification of segments of the parasite genome, and as a consequence, allowing the *Leishmania* genome to be continuously subjected to gene rearrangements, thus conferring high plasticity to the genome (Ubeda *et al.*,

2014; Rogers *et al.*, 2011). *Leishmania* parasites are able to replicate and maintain these extra-chromosomal elements as long as selective pressure is present, which also happens with plasmids not containing any *Leishmania* specific sequences (Papadopoulou *et al.*, 1994), suggesting that *Leishmania* parasites are highly permissive to extra-chromosomal elements without sequences needed for stability (for e.g. origins of replication) (Ubeda *et al.*, 2014). Could the genome-wide distribution of these repeated sequences and origin-independent replication of extra-chromosomal elements be linked to the need for recombination-dependent initiation of DNA replication to facilitate the complete replication of the *Leishmania* genome (or at least the larger chromosomes)? Like *Leishmania,* it was shown that *T. brucei* also possesses repeated sequences where gene rearrangements could take place (Ubeda *et al.*, 2014). However, no gene amplification has been described in *T. brucei*, thus suggesting that although the parasite has the potential (at the sequence level) to generate extra-chromosomal elements, this is not a common event, as the parasite is most likely unable to efficiently replicate and maintain them (Ubeda *et al.*, 2014). Though apparently uncommon, the presence and maintenance of a naturally occurring extra-chromosomal circular DNA element (NR-element) has nevertheless been reported in *T. brucei* Lister 427 and field strains (but not in the TREU 927 strain) (Alsford *et al.*, 2003). Could this difference in generating and maintaining extra-chromosomal elements between *Leishmania* and *T. brucei* result from them employing different strategies of DNA replication initiation, in which a replication origin is required for *T. brucei* to replicate and maintain any extra-chromosomal element, while in *Leishmania* these elements will be replicated completely via recombination-dependent initiation of replication? Null mutants of RAD51, a key player in homologous recombination, in *L. infantum* promastigotes were not reported to have a growth defect, besides being more sensitive to double strand break-inducing agents (Ubeda *et al.*, 2014). However, growth is a weak measure of genome stability, and it will be interesting to analyse DNA replication dynamics in these mutants as well as in other mutants, deficient in other factors involved in DNA recombination.

Another question arising from the results here presented is: how do *Leishmania* parasites cope with replicative stress if they only have one origin? As already suggested, it will be of importance to map the ORC1/CDC6-binding sites in

Leishmania in order to reveal whether there are other potential origins in the chromosomes that can be activated under replicative stress, such as dormant origins. Indeed, it will be interesting to generate ORC1/CDC6 mutants in *Leishmania*. In *T. brucei*, TbORC1/CDC6 is essential for parasite survival, and it was not possible to generate null mutants of the gene (Tiengwe *et al.*, 2012b), in principle because the parasite depends on TbORC1/CDC6 for origin of replication recognition and activation. If DNA replication can be initiated in *Leishmania* parasites independently of origins of replication, ORC1/CDC6 may not be as essential in these parasites. Nonetheless, one prediction of the single origin hypothesis is that in conditions of replicative stress no further discrete origins are present to restart replication. If so, it might be predicted that replicative stress must be overcome by recombination-dependent replication, perhaps leading to increased chromosome copy variation. Whole genome sequencing could be used to test this.

In *T. brucei* the strongest origins co-localised with the mapped centromeres (Tiengwe *et al.*, 2012a). It is thus possible that *L. major* and *L. mexicana* origins localise to the centromere of each chromosome. Although centromeres have not been successfully mapped in *Leishmania*, the kinetochore proteins of *T. brucei* have been recently identified and shown to localise to the centromeres, and orthologues for all of these were also shown be present in *L. mexicana* (Akiyoshi and Gull, 2014), and potentially in *L. major*. Logically, the next step would be to localise these proteins in *L. major* or *L. mexicana* to map centromere location and test whether these overlap with origins of replication.

In summary, mapping of the origins of replication in *Leishmania* has revealed an unprecedented case of origin usage in a eukaryote, in which a single origin is used per chromosome. Whether this feature is unique to the *Leishmania* genus or whether it is a common feature among early-diverged unicellular organisms it is not known, and only the analysis of a broader spectrum of eukaryotes will answer this question. Nevertheless, it appears that initiation of DNA replication in *Leishmania* is considerably different from *T. brucei*, though there is a high conservation of origin location. Moreover, it appears that origins of replication in *Leishmania*, but not *T. brucei*, are associated with specific genomic loci that differ from others that do not possess origin activity at least in one specific

aspect: the distance between the first two genes at strand switch regions. Further understanding of *Leishmania* DNA replication will be essential to provide insight into the evolution of the eukaryotic initiation of DNA replication machinery (apparently highly diverged in both *Leishmania* and *T. brucei*, see Chapter 3), as well as the evolution of the mechanisms behind co-ordination of initiation of DNA replication in eukaryotes.

6 Concluding Remarks & Future Perspectives

Since the *T. brucei* genome was sequenced in 2005 (Berriman *et al.*, 2005), it has become clear that the process of DNA replication in *T. brucei*, as well as in the related kinetoplastids *T. cruzi* and *L. major* (El-Sayed *et al.*, 2005b), may differ from that described in most of the eukaryotes studied so far (reviewed in Tiengwe *et al.*, 2013). In eukaryotes, the process of DNA replication has been mostly studied in a limited group of organisms, all of which belong to the Opisthokonta supergroup of the eukaryotic tree. Consequently, this focus might have led to the idea that this essential process is highly conserved among all eukaryotes, not just functionally, but also in terms of the molecular machineries involved. With the recent availability of sequenced genomes from a wider range of eukaryotes, belonging to the various eukaryotic supergroups, it has become clear that, most likely, there is a great variability between eukaryotes, especially regarding the factors involved in the first steps of DNA replication, such as the members of the pre-replication complex (pre-RC) (Aves *et al.*, 2012). For example, in the Excavata supergroup, to which *T. brucei* belongs, no sequence orthologues have been found for the Orc3 and Orc6 subunits of ORC, or for Cdc6 or Cdt1, suggesting that in these organisms these factors are either absent, present but so diverged that they cannot be identified bioinformatically, or non-homologous proteins perform their roles (Aves *et al.*, 2012). In addition, the Orc2, Orc4 and Orc5 subunits could not be found in all species analysed from this supergroup (Aves *et al.*, 2012). It is therefore not surprising that it has been a cumbersome task to identify the pre-RC components in *T. brucei*, *T. cruzi* or *L. major* (El-Sayed *et al.*, 2005a). The results presented in this thesis appear to suggest that a mix of the above scenarios may be present in *T. brucei*: there is evidence for the existence of a multi-subunit complex, but this might be composed of a combination of well-conserved Orc-like factors, some highly diverged Orc-like factors, and some non-homologous factors. To date, no evidence supports the existence of the key mediators Cdc6 and Cdt1, and it remains possible that these may be absent, with their functions/regulatory roles being performed by other factors.

At the start of this work, it was suggested that TbORC1/CDC6 interacts with TbORC1B (Dang and Li, 2011), TbORC4, Tb7980, and Tb3120 (Tiengwe *et al.*, 2012b), and also potentially with Tb1120 (Tiengwe, 2010). It was not known, however, if these interacted with TbORC1/CDC6 in an ORC-like complex,

whether these factors were involved in DNA replication, or if these were functionally replacing any of the ORC subunits still not identified at the time (reviewed in Tiengwe *et al.*, 2013). Presently, it is clear that TbORC1/CDC6 and TbORC4 are Orc-like subunits (Tiengwe *et al.*, 2012b; Godoy *et al.*, 2009), and both have been confirmed to be involved in DNA replication in PCF and BSF cells. Though initially identified as being most likely kinetoplastid-specific factors (Tiengwe *et al.*, 2012b), a re-analysis of Tb7980 and Tb3120 protein sequences tentatively suggests that these are, respectively, highly diverged Orc5-like and Orc2-like subunits. However, it was not possible to confirm the role of Tb7980 in DNA replication in either PCF or BSF cells, and therefore it remains premature to firmly re-label it as TbORC5. It will thus be essential, in the future, to associate it with DNA replication in *T. brucei*. In the case of Tb3120, evidence from PCF cells confirm its role in DNA replication, and protein sequence re-analysis was able to identify diverged Walker A and Walker B motifs defined by signature sequences present in Orc2 subunits, strongly suggesting that Tb3120 is, most likely, TbORC2. However, analysis of Tb3120 and corresponding orthologues in *T. cruzi* and *L. major* revealed that these proteins are considerably larger than other Orc2 proteins; when comparing these with other eukaryotes Orc2 subunits, it is clear that the kinetoplastid Orc2-like factors have an enlarged N-terminus, especially in *L. major*. It will be of interest, in the future, to try and explore what function this enlargement might provide: for instance, it may be that this domain provides an key activity or conformation to the putative kinetoplastid ORC, or even provide non-replication functions. It has not to date proved possible to infer homology with any other factors, replication-related or not, when analysing Tb1120. Though it was here confirmed that Tb1120 interacts with TbORC1/CDC6, it was not possible to show that it is involved in DNA replication, and thus it remains inconclusive whether it is part of, or not, an ORC-like complex. Very recently, the crystal structure of *D. melanogaster* ORC has been published (Bleichert *et al.*, 2015). It will be of value to repeat the structure prediction analysis (using RaptorX and I-Tasser), in order to infer whether any of these subunits would be used as templates for modelling some of the *T. brucei* Orc-like proteins. Moreover, it would be also interesting to predict the *T. brucei* factors' structures and try to model these into the *D. melanogaster* ORC structure (Bleichert *et al.*, 2015; Bleichert *et al.*, 2013), in order to infer putative interactions between the various factors.

Overall, if Tb7980 and Tb3120 are confirmed to be *T. brucei* TbORC5 and TbORC2, respectively, one can argue that *T. brucei* has orthologues for Orc1, Orc4, Orc5 and Orc2 subunits of ORC (Figure 6.1), which were the ones predicted to be present in the organisms from the Excavata supergroup (Aves *et al.*, 2012), while Orc3 and Orc6 are still missing, together with Cdc6 and Cdt1. TbORC1/CDC6 has been shown to complement functionally Cdc6 in yeast (Godoy *et al.*, 2009), so it is possible that in *T. brucei*, TbORC1/CDC6 might have a dual role, functioning both as Orc1 and Cdc6, and thus not requiring a independent Cdc6 factor. In yeast, for instance, Orc6 has two Cdt1-binding sites, and allows the recruitment and loading of the MCM2-7 helicase via Cdt1 (Takara and Bell, 2011). If Orc6 is truly absent in kinetoplastids, it may then make sense that Cdt1 is also absent, and it is thus possible that in the absence of both, the TbMCM2-7 helicase is recruited and loaded to the origin-bound ORC-like complex via other factors. It was shown that both TbORC1/CDC6 and TbORC1B interact with, at least, TbMCM3 (Dang and Li, 2011), but whether these interactions are enough to recruit and load the helicase onto the origins of replication (Figure 6.1), is not known and requires further investigation.

The gel filtration assay presented here suggests the existence of TbORC1/CDC6 in a complex, or various complexes of different sizes. Though this must be repeated, and the presence of the various TbORC1/CDC6-interacting factors in these complexes still require testing, these preliminary results support the existence of an ORC-like complex in *T. brucei*, which had not been shown before (reviewed in Tiengwe *et al.*, 2013). As already explored in the dedicated discussion section, TbORC1/CDC6 was detected in fractions suggesting complexes large enough to include TbORC1/CDC6, TbORC4, Tb7980, Tb3120 and Tb1120 as members of an ORC-like complex, but also other factors, which might have not been yet identified (potential model in Figure 6.1). Nevertheless, the most predominant putative complexes are too big to just harbour a putative ORC-like complex, and are large enough to include, in addition to the abovementioned factors, the six subunits of the TbMCM2-7 helicase, which would make up the pre-RC complex (if Cdc6 and Cdt1 are indeed lacking). Even larger complexes were also observed, that could in addition include the four subunits of TbGINS and TbCDC45, making up the pre-IC complex, and even TbMCM10. This all, of course, requires testing, which could be done by performing the same

assay using the PCF cell lines containing both TbORC1/CDC6 tagged endogenously with 12myc and each of the other factors tagged with 6HA, or performing an immunoprecipitation assay of the isolated gel filtration fractions, using TbORC1/CDC6^{12myc} as bait, followed by mass spectrometry analysis. Immunoprecipitation of TbORC4^{12myc} followed by mass spectrometry analysis, suggests that TbORC4 interacts not only with TbORC1/CDC6 but also with Tb3120 and Tb1120, and it will be essential to confirm these interactions by directed co-Immunoprecipitation. Nevertheless, these results offer further support for the existence of an ORC-like complex in *T. brucei*, in which, at least, TbORC1/CDC6, TbORC4, Tb3120 (possibly as TbORC2) and Tb1120, interact. Furthermore, the Immunoprecipitation of TbORC1/CDC6, TbORC4, and TbORC1B tagged with 12myc followed by mass spectrometry analysis retrieved a common hit, Tb927.10.2240 (Tb2240), which had not previously been examined. It will be important to investigate this new factor and test whether it is also another member of the ORC-like complex and if it is involved in DNA replication.

Immunolocalisation of the TbORC1/CDC6, TbORC4, Tb7980, Tb3120 and Tb1120 12myc-tagged versions revealed that all these factors localise to the nucleus of both PCF and BSF cells as punctate throughout the cell cycle, during which all showed identical localisation dynamics, perhaps supporting the idea that these are interacting in a complex. In the future, it will be important to co-localise these factors, perhaps by super resolution imaging. Though it requires further validation, it appears that TbORC1/CDC6 co-localises with newly replicated DNA during S phase; it has been described in other eukaryotes that ORC binds to newly synthesised origins of replication as S phase progresses, so it is possible that as the DNA is replicated, TbORC1/CDC6 recognises and binds to newly replicated origins of replication, a hypothesis that is somehow supported by the previous report that TbORC1/CDC6 is bound to the DNA throughout the cell cycle (Godoy *et al.*, 2009). Whether this is also true for the other factors requires testing.

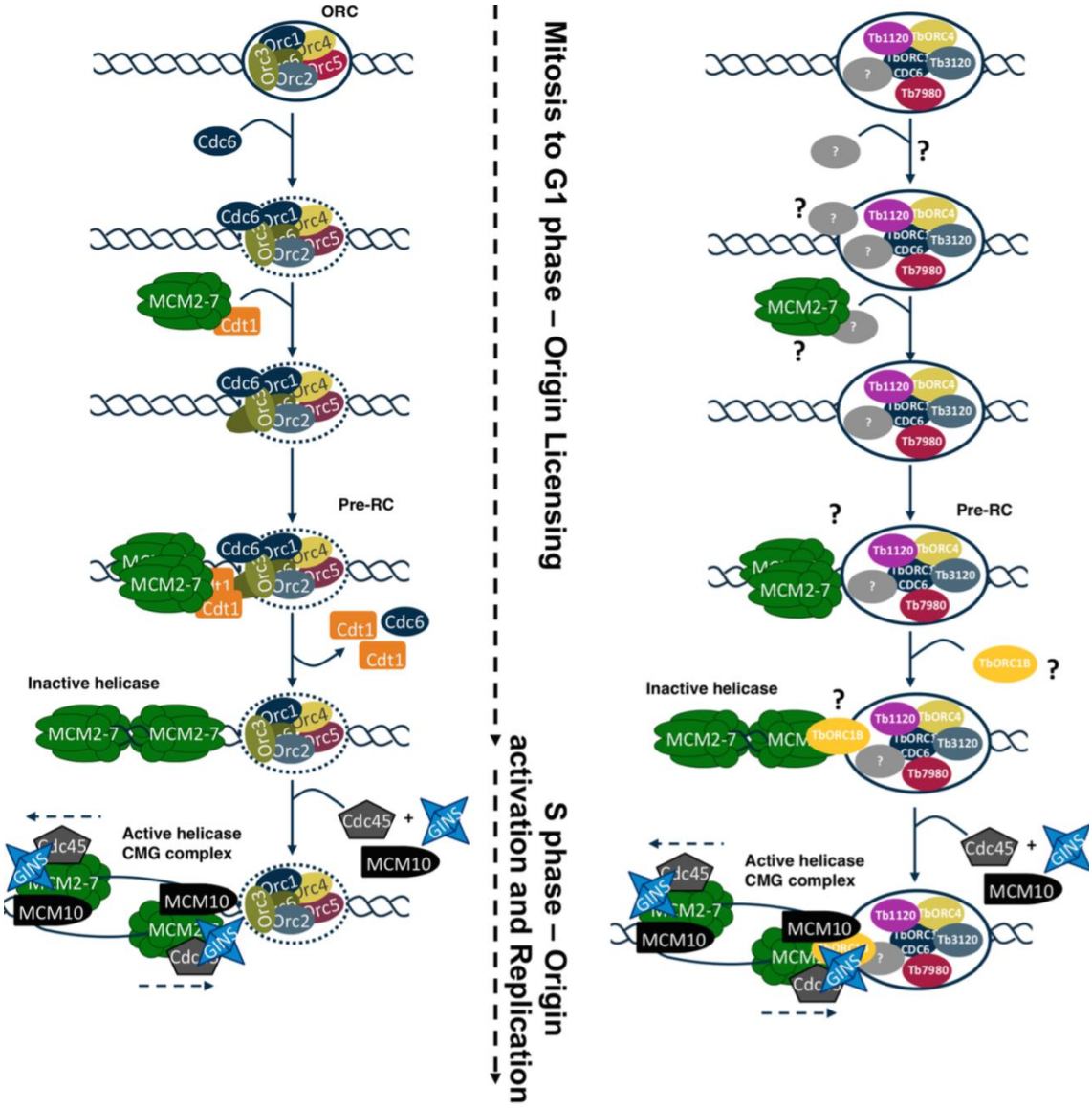


Figure 6.1. Models of pre-RC and pre-IC formation.
Simplified models of pre-RC and pre-IC formation in model eukaryotes (left), and *T. brucei* (right); both not to scale. In model eukaryotes, ORC binds to the origin of replication; next, Cdc6 binds, leading to conformational changes in ORC, exposing the Cdt1-binding domains in Orc6; Cdt1, bound to the MCM2-7 helicase, is recruited to the origin, where it binds to ORC via Orc6; Cdc6 and Orc1 undergo ATP hydrolysis, and this releases both Cdc6 and Cdt1 from the origins; at the onset of S phase, the GINS complex and Cdc45 are recruited, and form the CMG complex with MCM2-7, activating it, and unwinding the origin DNA, allowing the recruitment of the components of the replication fork. In *T. brucei*, TbORC1/CDC6 is known to bind to the origins of replication, and has been shown to interact with TbORC4, Tb7980 (potential TbORC5), Tb3120 (potential TbORC2), and Tb1120, possibly as a divergent ORC-like complex. There's indication that TbORC4 interacts with both Tb3120 and Tb1120, though this requires further validation. It is not known whether other factors are members of this ORC-like complex (represented in grey). It is not known if a factor such as Cdc6 binds to the ORC-like complex, or if the helicase is recruited to it directly or via a loader, in place of Cdt1. At the end of G1 and onset of S phase, TbORC1B is expressed and potentially binds to the ORC-like complex, at least via TbORC1/CDC6, and also to the TbMCM2-7 helicase at least via TbMCM3. Whether TbORC1B is recruited together with the TbGINS complex and TbCDC45, or if it interacts with them at all at the origin, it is not known. Supposedly, the helicase is then activated, the origin DNA is unwind, and DNA replication starts. Question marks represent the speculative steps in process of initiation of DNA replication in *T. brucei*, which must be further investigated in the near future. Diagrams adapted from (Tiengwe *et al.*, 2013), with permission.

Unlike all the other analysed factors, TbORC1B, both in PCF and BSF cells, localised to the nucleus of cells from late G1 phase/start of S phase to late S phase/early G2 phase. Further analysis revealed that TbORC1B is mainly expressed in S phase cells and is completely absent from G1 phase cells, strongly suggesting that TbORC1B expression is cell cycle dependent, though how its expression is controlled is not known, and requires further investigation. Together with the strikingly rapid phenotype following its silencing by inducible RNAi, which results in the virtual abolishment of DNA replication, the results herein presented suggest that, more than being a potential member of a ORC-like complex (indeed, the data suggest it is not a static member of such a complex), TbORC1B appears to be, instead, a regulatory factor. However, how it provides a regulatory function is presently unknown, as it does not seem to follow the pattern of either Cdc6 or Cdt1 localisation dynamics described in other organisms (Delmolino *et al.*, 2001; Petersen *et al.*, 1999; Piatti *et al.*, 1995; Luo *et al.*, 2003; Borlado and Mendez, 2008; Diffley, 2010; Li and Jin, 2010; Tanaka and Diffley, 2002; Nishitani *et al.*, 2000; Nishitani *et al.*, 2001; Wohlschlegel *et al.*, 2000). The evidence to date suggests that it might be a positive regulatory factor whose expression or nuclear localisation coincides with the onset of S phase and its presence continues to the late stages of S phase (and perhaps early G2 phase), but is removed when the cells have committed into mitosis. It remains possible then, that TbORC1B acts in a truly novel strategy in the control of nuclear DNA replication, and deserves further investigation.

Unlike TbORC1/CDC6, the preliminary localisation data suggest that TbORC1B does not appear to bind to newly synthesised DNA. If TbORC1B is acting from the beginning of S phase up to G2 phase as a regulator of DNA replication initiation, it is logical that it would not associate with newly replicated DNA, but instead with origins that are activated during the S phase. This would accord with the observations that TbORC1B signal is more intense in the nucleus of cells that are in the beginning and during S phase (1N1K and 1N1eK cells), decreasing as the cells progress into G2 phase (1N2K cells): at the start of S phase, TbORC1B would be present at the origins of DNA replication; as S phase progresses, less origins would remain to be activated, and thus TbORC1B signal would decrease as the cells progress into the end of S phase and beginning of the G2 phase. Once all

origins have been used, TbORC1B would no longer be needed, and therefore would no longer be detected in the nucleus. The complete absence of TbORC1B in G1 phase cells may suggest that it is important for the start of DNA replication and, therefore, acts as the factor that directs the activation of origins from the larger pool of TbORC1/CDC6 (and potentially ORC) binding sites. These suggestions require further investigation, and it will be of great value to perform a chromatin-immunoprecipitation using TbORC1B as bait, and ask whether this binds to all potential origins, like TbORC1/CDC6, or whether it only binds to the origins that are activated during S phase, which were detected by the MFA-seq analysis (Tiengwe *et al.*, 2012b).

Initially proposed as another Orc1-like factor devoid of ATPase activity, TbORC1B was shown to interact with both TbORC1/CDC6 and TbMCM3 (Dang and Li, 2011). Though various attempts were made in this study to confirm that TbORC1B interacts with TbORC1/CDC6, this proved unsuccessful. This would be consistent with TbORC1B acting as a regulatory factor: if it interacts with the ORC-like complex at the origins of replication, such interaction may be rather transient, taking place between late G1/early S and late S/early G2 phase, and thus might be difficult to detect. The interaction between TbORC1B and TbORC1/CDC6, or any other ORC and pre-RC factors should be further analysed (perhaps via immunoprecipitation assays coupled with cross-linking, if the interaction is transient) to probe in detail how it might influence replication initiation.

In order to better understand DNA replication in *T. brucei*, the origins of replication were also a focus of attention in the present work. Previously, the origins of replication have been mapped in *T. brucei* PCF cells, revealing a small number of active origins, all localising to the SSR boundaries of the polycistronic transcription units, and the existence of an origin activation timing programme (Tiengwe *et al.*, 2012a). Though all the analysed factors' subcellular localisation dynamics was identical in PCF and BSF cells, we sought to map, using the same strategy as before (MFA-seq) (Tiengwe *et al.*, 2012a), the activated origins of replication in BSF cells, driven by reports that in other eukaryotes origin usage differs between cell types (reviewed in Leonard and Mechali, 2013; Rhind and Gilbert, 2013). Perhaps not surprisingly, though the two life cycle stages of the parasite differ in many biological aspects (detailed in the introduction and in

Chapter 4), the same activated origins were detected, and appear to follow the same, or very similar, activation programme in BSF cells as in PCF cells. This reveals that the programme of origin usage is rather inflexible in *T. brucei*, which might be related to the fact that, unlike other eukaryotes, gene expression (of housekeeping genes, by RNA pol II) in *T. brucei*, and related kinetoplastids, is not mainly controlled at the transcription level but at the post-transcriptional level (reviewed in Clayton, 2002; Clayton, 2013; Kramer, 2012; Siegel *et al.*, 2011). Thus, pronounced differences in chromatin status, structure and composition that may underlie replication variation in different developmental stages in other eukaryotes may be absent between PCF and BSF cells. Nonetheless, if there is no variation in transcription initiation or termination at any SSR in *T. brucei*, it remains unclear what features determine the relatively rigid pattern of origin firing (as seen in varying MFA-seq peak heights), which is also maintained between BSF and PCF cells. The main difference in transcription patterns between PCF and BSF cells lies on the expression of their surface antigens: procyclins in PCF cells and VSGs in BSF cells, both expressed in a life cycle-dependent fashion by RNA Pol I (Gunzl *et al.*, 2003). Indeed, the only apparent difference between origin mapping in PCF and BSF cells lies in the region from where the single VSG is being expressed in the analysed BSF cells (VSG 221), the active bloodstream expression site (BES 1). This genome feature appeared to be replicated early during S phase in BSF cells (where it is transcribed), but be late replicating in PCF cells (where it is silenced). All other analysed BES, whose expression is silenced in both life cycle stages, were replicated late in both BSF and PCF cells. In order to validate these observations, it will be necessary to analyse BSF cell lines in which another BES is active and BES 1 is inactive, and ask whether the distinct, active BES is replicated early in BSF cells while BES 1 is replicated late. Nevertheless, these data support previous findings (Tiengwe *et al.*, 2012a; Benmerzouga *et al.*, 2013) in which DNA replication and transcription appear to be intimately connected in *T. brucei*, and that DNA replication might be involved in antigenic variation.

With the initial intention to better understand origins of replication in *T. brucei*, and in an attempt to identify origin-defining features, origins were also mapped in *Leishmania major* and *L. mexicana*, taking advantage of their genomes being

highly syntenic between each other as well as with the genome of *T. brucei* (El-Sayed *et al.*, 2005b). Surprisingly, only one origin of replication could be identified per chromosome in both *Leishmania* species. More strikingly, in two *L. mexicana* chromosomes, each of which is syntenic with two *L. major* chromosomes (each with a single detectable origin), only one origin was detected, suggesting that a single origin per chromosome is preferred in *Leishmania*. This is unprecedented in eukaryotes, which are assumed to have multiple origins per chromosome (reviewed in Masai *et al.*, 2010; Mechali, 2010; Leonard and Mechali, 2013), and contrasts with what has been consistently observed in *T. brucei* (Tiengwe *et al.*, 2012a; and shown in Chapter 4). In addition, all detected origins appeared to be of the same strength, suggesting that, again in contrast to *T. brucei*, *Leishmania* does not possess an origin activation timing programme, and instead all origins are activated at the same time. The analysis of some chromosomes by MFA coupled with qPCR (MFA-qPCR) supports the existence of a single origin per *Leishmania* chromosome, but it will be crucial to confirm this, as it goes against the orthodoxy established in the field: eukaryotic chromosomes are replicated from multiple origins of replication. One option would be to perform single molecule analysis of replicated DNA (SMARD). However, if DNA replication can be initiated via other mechanisms that are independent from origins of replication, such as recombination-dependent DNA replication (below), these events will be detected by SMARD, and will be difficult to distinguish from the existence of 'extra' origins. Another option would be to perform MFA-seq on a single cell basis, although how feasible single cell sequencing (Nawy, 2014; Eberwine *et al.*, 2014) would be in *Leishmania* is not clear at present. Therefore, confirmation of a single origin per *Leishmania* chromosome by other experimental procedures beyond MFA-seq will have to be planned carefully, in order to exclude replication-initiation events that are not dependent on origins of replication. Although the rate of replication fork movement has been recently calculated in both *T. brucei* PCF and BSF cells (Calderano *et al.*, 2015), and appears to be slightly faster than in other eukaryotes (reviewed in Mechali, 2010), the same has not been done in *Leishmania*. Even if it were predicted that *Leishmania* replication fork rate is similar to the one of *T. brucei*, this would not be enough to allow full replication of the *Leishmania* larger chromosomes within S phase; it would be enough to only completely replicate the smaller chromosomes. The

MFA-seq approach is able to detect origins that are activated at the population level (constitutive origins), and it is thus possible that origins that are activated only at the individual cell level (flexible or dormant origins) are below detection. It is therefore possible that other origins are present per *Leishmania* chromosome, but these are activated differently from cell to cell and, consequently, are not detected by MFA-seq/MFA-qPCR. Nevertheless, even this scenario is still completely different from what was observed in *T. brucei*, suggesting that DNA replication dynamics in these two closely related organisms differ. Whether this is because the machineries involved in replication differ is unknown, but may be hinted at by our observation of differences in what features define what SSR acts as an origin relative to those SSRs that do not.

Like in *T. brucei*, no specific sequence was detected that defines what an origin is in *Leishmania*. Furthermore, *Leishmania* origins localised to SSRs. However, analysis of the size of origin and non-origin SSRs revealed that there is a significant difference between the two SSR classes in *Leishmania*, but not in *T. brucei*. Besides, in the two unique syntenic regions which are origin-active SSRs in *L. major* but non-origin SSRs in *L. mexicana*, a difference in size was also observed, suggesting that SSR size might be an origin-defining feature in *Leishmania*. What feature of origin designation SSR size in *Leishmania* dictates is unknown, but may relate to a requirement for the binding of specific factors or the generation of specific chromatin structures. Given the likely absence of such SSR differences in *T. brucei*, comparison of the machinery that localises to SSRs in the two parasites may be revealing.

If indeed there is only one origin per chromosome, *Leishmania* must have developed a strategy to achieve complete replication of each chromosome within S phase. One strategy could be to initiate DNA replication independently from origins of replication, via recombination-dependent DNA replication, as suggested and discussed in Chapter 5, which could also explain the variation in and tolerance of chromosome ploidy changes in *Leishmania*, but not *T. brucei*. It will be essential to test this hypothesis by, for instance, mapping the origins of replication in cell lines deficient for recombination factors such as Rad51. If *Leishmania* is indeed able to initiate DNA replication in a non-origin dependent

fashion, it will also be of value to investigate the DNA replication and cell cycle effects of *Leishmania* ORC1/CDC6 depletion.

Origin singularity in *Leishmania*, besides raising the question on how the parasite fully replicates its larger chromosomes within S phase, also raises the question on how the parasite copes with replicative stress. Although recombination-dependent DNA replication could be the answer for both of these, it will be of importance to map the *Leishmania* ORC1/CDC6-binding sites at the genome-wide level by chromatin-immunoprecipitation followed by sequencing, in order to infer whether the initiator only binds to the origins that are activated or if it also bind to the other SSRs, potentially allowing them to act as origins under stress. It will also be interesting to map the origins of replication in cells in which DNA replication stress has been induced (e.g. hydroxyurea). In other eukaryotes (Leonard and Mechali, 2013; Rhind and Gilbert, 2013), including in *T. brucei* (Tiengwe *et al.*, 2012a), the centromeres are one of the earliest regions of the genome to be replicated. Unfortunately, to date, the centromeres have not been mapped in *Leishmania*. Recently, however, the kinetochore subunits have been identified in *T. brucei* and appear to be conserved in *Leishmania* (Akiyoshi and Gull, 2014). It will, therefore, be of interest to map, as for ORC1/CDC6, the kinetochore-binding sites, and ask whether these localise to the single origins detected per *Leishmania* chromosome.

In conclusion, much still has to be done in the field of DNA replication in *T. brucei* and related kinetoplastids, and it will be a very exciting area of research in the next years. Whether the divergences, both in the initiation machinery and in the dynamics of replication, relative to mammals are sufficient to warrant investigation as routes towards drug development is unclear and requires further fundamental studies.

List of References

Abbas, T., Keaton, M. A. & Dutta, A. (2013). Genomic instability in cancer. *Cold Spring Harb Perspect Biol,* **5**, a012914.

Acosta-Serrano, A., Vassella, E., Liniger, M., Kunz Renggli, C., Brun, R., Roditi, I. & Englund, P. T. (2001). The surface coat of procyclic Trypanosoma brucei: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. *Proc Natl Acad Sci U S A,* **98**, 1513-8.

Adl, S. M., Simpson, A. G., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., Bowser, S. S., Brugerolle, G., Fensome, R. A., Fredericq, S., James, T. Y., Karpov, S., Kugrens, P., Krug, J., Lane, C. E., Lewis, L. A., Lodge, J., Lynn, D. H., Mann, D. G., McCourt, R. M., Mendoza, L., Moestrup, O., Mozley-Standridge, S. E., Nerad, T. A., Shearer, C. A., Smirnov, A. V., Spiegel, F. W. & Taylor, M. F. (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol,* **52**, 399-451.

Adl, S. M., Simpson, A. G., Lane, C. E., Lukes, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Le Gall, L., Lynn, D. H., McManus, H., Mitchell, E. A., Mozley-Stanridge, S. E., Parfrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, R. S., Schoch, C. L., Smirnov, A. & Spiegel, F. W. (2012). The revised classification of eukaryotes. *J Eukaryot Microbiol,* **59**, 429-93.

Akita, M., Adachi, A., Takemura, K., Yamagami, T., Matsunaga, F. & Ishino, Y. (2010). Cdc6/Orc1 from Pyrococcus furiosus may act as the origin recognition protein and Mcm helicase recruiter. *Genes Cells,* **15**, 537-52.

Akiyoshi, B. & Gull, K. (2014). Discovery of unconventional kinetochores in kinetoplastids. *Cell,* **156**, 1247-58.

Aladjem, M. I. (2007). Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nat Rev Genet,* **8**, 588-600.

Alsford, N. S., Navarro, M., Jamnadass, H. R., Dunbar, H., Ackroyd, M., Murphy, N. B., Gull, K. & Ersfeld, K. (2003). The identification of circular extrachromosomal DNA in the nuclear genome of Trypanosoma brucei. *Mol Microbiol,* **47**, 277-89.

Alsford, S. & Horn, D. (2008). Single-locus targeting constructs for reliable regulated RNAi and transgene expression in Trypanosoma brucei. *Mol Biochem Parasitol,* **161**, 76-9.

Alsford, S., Kawahara, T., Glover, L. & Horn, D. (2005). Tagging a T. brucei RRNA locus improves stable transfection efficiency and circumvents inducible expression position effects. *Mol Biochem Parasitol,* **144**, 142-8.

Alsford, S., Turner, D. J., Obado, S. O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-Fowler, C. & Horn, D. (2011). High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res,* **21**, 915-24.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol,* **215**, 403-10.

Alvar, J., Velez, I. D., Bern, C., Herrero, M., Desjeux, P., Cano, J., Jannin, J. & den Boer, M. (2012). Leishmaniasis worldwide and global estimates of its incidence. *PLoS One,* **7**, e35671.

Alvarez-Fernandez, R. (2013). Explanatory chapter: PCR primer design. *Methods Enzymol,* **529**, 1-21.

Ambit, A., Woods, K. L., Cull, B., Coombs, G. H. & Mottram, J. C. (2011). Morphological events during the cell cycle of Leishmania major. *Eukaryot Cell*, **10**, 1429-38.

Anand, R. P., Lovett, S. T. & Haber, J. E. (2013). Break-induced DNA replication. *Cold Spring Harb Perspect Biol*, **5**, a010397.

Aparicio, T., Guillou, E., Coloma, J., Montoya, G. & Mendez, J. (2009). The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. *Nucleic Acids Res*, **37**, 2087-95.

Araki, M., Wharton, R. P., Tang, Z., Yu, H. & Asano, M. (2003). Degradation of origin recognition complex large subunit by the anaphase-promoting complex in Drosophila. *Embo j*, **22**, 6115-26.

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

Arias, E. E. & Walter, J. C. (2007). Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes Dev*, **21**, 497-518.

Aronesty, E. (2011). Command-line tools for processing biological sequencing data.

Asano, M. & Wharton, R. P. (1999). E2F mediates developmental and cell cycle regulation of ORC1 in Drosophila. *Embo j*, **18**, 2435-48.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., Gao, X., Gardner, M. J., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Logan, F. J., Miller, J. A., Mitra, S., Myler, P. J., Nayak, V., Pennington, C., Phan, I., Pinney, D. F., Ramasamy, G., Rogers, M. B., Roos, D. S., Ross, C., Sivam, D., Smith, D. F., Srinivasamoorthy, G., Stoeckert, C. J., Jr., Subramanian, S., Thibodeau, R., Tivey, A., Treatman, C., Velarde, G. & Wang, H. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res*, **38**, D457-62.

Auth, T., Kunkel, E. & Grummt, F. (2006). Interaction between HP1alpha and replication proteins in mammalian cells. *Exp Cell Res*, **312**, 3349-59.

Aves, S. J., Liu, Y. & Richards, T. A. (2012). Evolutionary diversification of eukaryotic DNA replication machinery. *Subcell Biochem*, **62**, 19-35.

Azuara, V. (2006). Profiling of DNA replication timing in unsynchronized cell populations. *Nat Protoc*, **1**, 2171-7.

Bailey, S., Eliason, W. K. & Steitz, T. A. (2007). Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science*, **318**, 459-63.

Balasov, M., Huijbregts, R. P. & Chesnokov, I. (2007). Role of the Orc6 protein in origin recognition complex-dependent DNA binding and replication in Drosophila melanogaster. *Mol Cell Biol*, **27**, 3143-53.

Barlow, J. H. & Nussenzweig, A. (2014). Replication initiation and genome instability: a crossroads for DNA and RNA synthesis. *Cell Mol Life Sci*.

Barry, E. R. & Bell, S. D. (2006). DNA replication in the archaea. *Microbiol Mol Biol Rev*, **70**, 876-87.

Barry, J. D., Graham, S. V., Fotheringham, M., Graham, V. S., Kobryn, K. & Wymer, B. (1998). VSG gene control and infectivity strategy of metacyclic stage Trypanosoma brucei. *Mol Biochem Parasitol*, **91**, 93-105.

Becker, M., Aitcheson, N., Byles, E., Wickstead, B., Louis, E. & Rudenko, G. (2004). Isolation of the repertoire of VSG expression site containing

telomeres of Trypanosoma brucei 427 using transformation-associated recombination in yeast. *Genome Res,* **14**, 2319-29.
Bell, S. D. & Botchan, M. R. (2013). The minichromosome maintenance replicative helicase. *Cold Spring Harb Perspect Biol,* **5**, a012807.
Bell, S. D., Méchali, M. & DePamphilis, M. L. (2013). *DNA Replication*. 1 ed. United States of America: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
Bell, S. P. & Kaguni, J. M. (2013). Helicase loading at chromosomal origins of replication. *Cold Spring Harb Perspect Biol,* **5**.
Bell, S. P. & Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature,* **357**, 128-34.
Benmerzouga, I., Concepcion-Acevedo, J., Kim, H. S., Vandoros, A. V., Cross, G. A., Klingbeil, M. M. & Li, B. (2013). *Trypanosoma brucei* Orc1 is essential for nuclear DNA replication and affects both VSG silencing and VSG switching. *Mol Microbiol,* **87**, 196-210.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res,* **28**, 235-42.
Berquist, B. R. & DasSarma, S. (2003). An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, Halobacterium sp. strain NRC-1. *J Bacteriol,* **185**, 5959-66.
Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L., Turner, C. M., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., et al. (2005). The genome of the African trypanosome Trypanosoma brucei. *Science,* **309**, 416-22.
Blaesing, F., Weigel, C., Welzeck, M. & Messer, W. (2000). Analysis of the DNA-binding domain of Escherichia coli DnaA protein. *Mol Microbiol,* **36**, 557-69.
Bleichert, F., Balasov, M., Chesnokov, I., Nogales, E., Botchan, M. R. & Berger, J. M. (2013). A Meier-Gorlin syndrome mutation in a conserved C-terminal helix of Orc6 impedes origin recognition complex formation. *Elife,* **2**, e00882.
Bleichert, F., Botchan, M. R. & Berger, J. M. (2015). Crystal structure of the eukaryotic origin recognition complex. *Nature*.
Blow, J. J., Ge, X. Q. & Jackson, D. A. (2011). How dormant origins promote complete genome replication. *Trends Biochem Sci,* **36**, 405-14.

Blow, J. J. & Gillespie, P. J. (2008). Replication licensing and cancer--a fatal entanglement? *Nat Rev Cancer,* **8**, 799-806.

Bochman, M. L., Bell, S. P. & Schwacha, A. (2008). Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. *Mol Cell Biol,* **28**, 5865-73.

Bochman, M. L. & Schwacha, A. (2009). The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol Mol Biol Rev,* **73**, 652-83.

Boothroyd, C. E., Dreesen, O., Leonova, T., Ly, K. I., Figueiredo, L. M., Cross, G. A. & Papavasiliou, F. N. (2009). A yeast-endonuclease-generated DNA break induces antigenic switching in Trypanosoma brucei. *Nature,* **459**, 278-81.

Borlado, L. R. & Mendez, J. (2008). CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. *Carcinogenesis,* **29**, 237-43.

Borst, P. & Sabatini, R. (2008). Base J: discovery, biosynthesis, and possible functions. *Annu Rev Microbiol,* **62**, 235-51.

Bowers, J. L., Randell, J. C., Chen, S. & Bell, S. P. (2004). ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. *Mol Cell,* **16**, 967-78.

Brand, N., Faul, T. & Grummt, F. (2007). Interactions and subcellular distribution of DNA replication initiation proteins in eukaryotic cells. *Mol Genet Genomics,* **278**, 623-32.

Braun, K. A. & Breeden, L. L. (2007). Nascent transcription of MCM2-7 is important for nuclear localization of the minichromosome maintenance complex in G1. *Mol Biol Cell,* **18**, 1447-56.

Braun, R. E., O'Day, K. & Wright, A. (1985). Autoregulation of the DNA replication gene dnaA in E. coli K-12. *Cell,* **40**, 159-69.

Brems, S., Guilbride, D. L., Gundlesdodjir-Planck, D., Busold, C., Luu, V. D., Schanne, M., Hoheisel, J. & Clayton, C. (2005). The transcriptomes of Trypanosoma brucei Lister 427 and TREU927 bloodstream and procyclic trypomastigotes. *Mol Biochem Parasitol,* **139**, 163-72.

Britto, C., Ravel, C., Bastien, P., Blaineau, C., Pages, M., Dedet, J. P. & Wincker, P. (1998). Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World Leishmania genomes. *Gene,* **222**, 107-17.

Brizzard, B. (2008). Epitope tagging. *Biotechniques,* **44**, 693-5.

Bruck, I. & Kaplan, D. (2009). Dbf4-Cdc7 phosphorylation of Mcm2 is required for cell growth. *J Biol Chem,* **284**, 28823-31.

Brun, R., Blum, J., Chappuis, F. & Burri, C. (2010). Human African trypanosomiasis. *Lancet,* **375**, 148-59.

Brun, R. & Schonenberger (1979). Cultivation and in vitro cloning or procyclic culture forms of Trypanosoma brucei in a semi-defined medium. Short communication. *Acta Trop,* **36**, 289-92.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J. & Wittwer, C. T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem,* **55**, 611-22.

Cadoret, J. C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H. & Prioleau, M. N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc Natl Acad Sci U S A,* **105**, 15837-42.

Caillat, C. & Perrakis, A. (2012). Cdt1 and geminin in DNA replication initiation. *Subcell Biochem,* **62**, 71-87.

Calderano, S., Godoy, P., Soares, D., Sant'Anna, O. A., Schenkman, S. & Elias, M. C. (2014). ORC1/CDC6 and MCM7 distinct associate with chromatin through Trypanosoma cruzi life cycle. *Molecular and Biochemical Parasitology,* **193**, 110-113.

Calderano, S. G., de Melo Godoy, P. D., da Cunha, J. P. & Elias, M. C. (2011a). Trypanosome prereplication machinery: a potential new target for an old problem. *Enzyme Res,* **2011**, 518258.

Calderano, S. G., de Melo Godoy, P. D., Motta, M. C., Mortara, R. A., Schenkman, S. & Elias, M. C. (2011b). Trypanosoma cruzi DNA replication includes the sequential recruitment of pre-replication and replication machineries close to nuclear periphery. *Nucleus,* **2**, 136-45.

Calderano, S. G., Drosopoulos, W. C., Quaresma, M. M., Marques, C. A., Kosiyatrakul, S., McCulloch, R., Schildkraut, C. L. & Elias, M. C. (2015). Single molecule analysis of Trypanosoma brucei DNA replication dynamics. *Nucleic Acids Res*.

Cann, A. J. (2003). *Maths from Scratch for Biologists*. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, UK: John Wiley & Sons, Ltd.

Cassler, M. R., Grimwade, J. E. & Leonard, A. C. (1995). Cell cycle-specific changes in nucleoprotein complexes at a chromosomal replication origin. *Embo j,* **14**, 5833-41.

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R. & Mechali, M. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res,* **21**, 1438-49.

Chaves, I., Zomerdijk, J., Dirks-Mulder, A., Dirks, R. W., Raap, A. K. & Borst, P. (1998). Subnuclear localization of the active variant surface glycoprotein gene expression site in Trypanosoma brucei. *Proc Natl Acad Sci U S A,* **95**, 12328-33.

Chen, S. & Bell, S. P. (2011). CDK prevents Mcm2-7 helicase loading by inhibiting Cdt1 interaction with Orc6. *Genes Dev,* **25**, 363-72.

Chen, S., de Vries, M. A. & Bell, S. P. (2007). Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading. *Genes Dev,* **21**, 2897-907.

Cheng, L., Collyer, T. & Hardy, C. F. (1999). Cell cycle regulation of DNA replication initiator factor Dbf4p. *Mol Cell Biol,* **19**, 4270-8.

Chesnokov, I., Remus, D. & Botchan, M. (2001). Functional analysis of mutant and wild-type Drosophila origin recognition complex. *Proc Natl Acad Sci U S A,* **98**, 11997-2002.

Chesnokov, I. N., Chesnokova, O. N. & Botchan, M. (2003). A cytokinetic function of Drosophila ORC6 protein resides in a domain distinct from its replication activity. *Proc Natl Acad Sci U S A,* **100**, 9150-5.

Chuang, R. Y. & Kelly, T. J. (1999). The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc Natl Acad Sci U S A,* **96**, 2656-61.

Clarey, M. G., Botchan, M. & Nogales, E. (2008). Single particle EM studies of the Drosophila melanogaster origin recognition complex and evidence for DNA wrapping. *J Struct Biol,* **164**, 241-9.

Clarey, M. G., Erzberger, J. P., Grob, P., Leschziner, A. E., Berger, J. M., Nogales, E. & Botchan, M. (2006). Nucleotide-dependent conformational changes in the DnaA-like core of the origin recognition complex. *Nat Struct Mol Biol,* **13**, 684-90.

Clayton, C. (2013). The Regulation of Trypanosome Gene Expression by RNA-Binding Proteins. *PLoS Pathog,* **9**, e1003680.

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

Costa, A., Hood, I. V. & Berger, J. M. (2013). Mechanisms for initiating cellular DNA replication. *Annu Rev Biochem,* **82**, 25-54.

Costa, A., Ilves, I., Tamberg, N., Petojevic, T., Nogales, E., Botchan, M. R. & Berger, J. M. (2011). The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol,* **18**, 471-7.

Costa, A. & Onesti, S. (2009). Structural biology of MCM helicases. *Crit Rev Biochem Mol Biol,* **44**, 326-42.

Costa, A., Pape, T., van Heel, M., Brick, P., Patwardhan, A. & Onesti, S. (2006). Structural studies of the archaeal MCM complex in different functional states. *J Struct Biol,* **156**, 210-9.

Costa, A., Renault, L., Swuec, P., Petojevic, T., Pesavento, J. J., Ilves, I., MacLellan-Gibson, K., Fleck, R. A., Botchan, M. R. & Berger, J. M. (2014). DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome. *Elife,* **3**, e03273.

Crampton, A., Chang, F., Pappas, D. L., Jr., Frisch, R. L. & Weinreich, M. (2008). An ARS element inhibits DNA replication through a SIR2-dependent mechanism. *Mol Cell,* **30**, 156-66.

Creek, D. J., Anderson, J., McConville, M. J. & Barrett, M. P. (2012). Metabolomic analysis of trypanosomatid protozoa. *Mol Biochem Parasitol,* **181**, 73-84.

Crevel, G., Ivetic, A., Ohno, K., Yamaguchi, M. & Cotterill, S. (2001). Nearest neighbour analysis of MCM protein complexes in Drosophila melanogaster. *Nucleic Acids Res,* **29**, 4834-42.

Cross, G. A., Kim, H. S. & Wickstead, B. (2014). Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427. *Mol Biochem Parasitol,* **195**, 59-73.

Dacks, J. B., Walker, G. & Field, M. C. (2008). Implications of the new eukaryotic systematics for parasitologists. *Parasitol Int,* **57**, 97-104.

Dang, H. Q. & Li, Z. (2011). The Cdc45.Mcm2-7.GINS protein complex in trypanosomes regulates DNA replication and interacts with two Orc1-like proteins in the origin recognition complex. *J Biol Chem,* **286**, 32424-35.

Daniels, J. P., Gull, K. & Wickstead, B. (2010). Cell biology of the trypanosome genome. *Microbiol Mol Biol Rev,* **74**, 552-69.

Davey, M. J., Fang, L., McInerney, P., Georgescu, R. E. & O'Donnell, M. (2002a). The DnaC helicase loader is a dual ATP/ADP switch protein. *Embo j,* **21**, 3148-59.

Davey, M. J., Indiani, C. & O'Donnell, M. (2003). Reconstitution of the Mcm2-7p heterohexamer, subunit arrangement, and ATP site architecture. *J Biol Chem,* **278**, 4491-9.

Davey, M. J., Jeruzalmi, D., Kuriyan, J. & O'Donnell, M. (2002b). Motors and switches: AAA+ machines within the replisome. *Nat Rev Mol Cell Biol,* **3**, 826-35.

Dayn, A., Malkhosyan, S. & Mirkin, S. M. (1992). Transcriptionally driven cruciform formation in vivo. *Nucleic Acids Res,* **20**, 5991-7.

De Biasio, A. & Blanco, F. J. (2013). Proliferating cell nuclear antigen structure and interactions: too many partners for one dancer? *Adv Protein Chem Struct Biol,* **91**, 1-36.

de la Paz Sanchez, M. & Gutierrez, C. (2009). Arabidopsis ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. *Proc Natl Acad Sci U S A,* **106**, 2065-70.

de Moura, A. P., Retkute, R., Hawkins, M. & Nieduszynski, C. A. (2010). Mathematical modelling of whole chromosome replication. *Nucleic Acids Res,* **38**, 5623-33.

Dean, S., Marchetti, R., Kirk, K. & Matthews, K. R. (2009). A surface transporter family conveys the trypanosome differentiation signal. *Nature,* **459**, 213-7.

Delgado, S., Gomez, M., Bird, A. & Antequera, F. (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. *Embo j,* **17**, 2426-35.

Delmolino, L. M., Saha, P. & Dutta, A. (2001). Multiple mechanisms regulate subcellular localization of human CDC6. *J Biol Chem,* **276**, 26947-54.

Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. (2009). TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol Cell,* **35**, 403-13.

DePamphilis, M. L. (2005). Cell cycle dependent regulation of the origin recognition complex. *Cell Cycle,* **4**, 70-9.

Di Rienzi, S. C., Lindstrom, K. C., Mann, T., Noble, W. S., Raghuraman, M. K. & Brewer, B. J. (2012). Maintaining replication origins in the face of genomic change. *Genome Res,* **22**, 1940-52.

Diffley, J. F. (2010). The many faces of redundancy in DNA replication control. *Cold Spring Harb Symp Quant Biol,* **75**, 135-42.

Diffley, J. F. (2011). Quality control in the initiation of eukaryotic DNA replication. *Philos Trans R Soc Lond B Biol Sci,* **366**, 3545-53.

Downing, T., Imamura, H., Decuypere, S., Clark, T. G., Coombs, G. H., Cotton, J. A., Hilley, J. D., de Doncker, S., Maes, I., Mottram, J. C., Quail, M. A., Rijal, S., Sanders, M., Schonian, G., Stark, O., Sundar, S., Vanaerschot, M., Hertz-Fowler, C., Dujardin, J. C. & Berriman, M. (2011). Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res,* **21**, 2143-56.

Dreesen, O., Li, B. & Cross, G. A. (2007). Telomere structure and function in trypanosomes: a proposal. *Nat Rev Microbiol,* **5**, 70-5.

Drosopoulos, W. C., Kosiyatrakul, S. T., Yan, Z., Calderano, S. G. & Schildkraut, C. L. (2012). Human telomeres replicate using chromosome-specific, rather than universal, replication programs. *J Cell Biol,* **197**, 253-66.

Dubessay, P., Ravel, C., Bastien, P., Lignon, M. F., Ullman, B., Pages, M. & Blaineau, C. (2001). Effect of large targeted deletions on the mitotic stability of an extra chromosome mediating drug resistance in Leishmania. *Nucleic Acids Res,* **29**, 3231-40.

Dubessay, P., Ravel, C., Bastien, P., Stuart, K., Dedet, J. P., Blaineau, C. & Pages, M. (2002). Mitotic stability of a coding DNA sequence-free version of Leishmania major chromosome 1 generated by targeted chromosome fragmentation. *Gene,* **289**, 151-9.

Duderstadt, K. E. & Berger, J. M. (2008). AAA+ ATPases in the initiation of DNA replication. *Crit Rev Biochem Mol Biol,* **43**, 163-87.

Duderstadt, K. E., Mott, M. L., Crisona, N. J., Chuang, K., Yang, H. & Berger, J. M. (2010). Origin remodeling and opening in bacteria rely on distinct assembly states of the DnaA initiator. *J Biol Chem,* **285**, 28229-39.

Dueber, E. C., Costa, A., Corn, J. E., Bell, S. D. & Berger, J. M. (2011). Molecular determinants of origin discrimination by Orc1 initiators in archaea. *Nucleic Acids Res,* **39**, 3621-31.

Dueber, E. L., Corn, J. E., Bell, S. D. & Berger, J. M. (2007). Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science,* **317**, 1210-3.

Duncker, B. P., Chesnokov, I. N. & McConkey, B. J. (2009). The origin recognition complex protein family. *Genome Biol,* **10**, 214.

Duque, A. & Rakic, P. (2011). Different effects of bromodeoxyuridine and [3H]thymidine incorporation into DNA on cell proliferation, position, and fate. *J Neurosci,* **31**, 15205-17.

Durand-Dubief, M. & Bastin, P. (2003). TbAGO1, an argonaute protein required for RNA interference, is involved in mitosis and chromosome segregation in Trypanosoma brucei. *BMC Biol,* **1**, 2.

Durand-Dubief, M., Kohl, L. & Bastin, P. (2003). Efficiency and specificity of RNA interference generated by intra- and intermolecular double stranded RNA in Trypanosoma brucei. *Mol Biochem Parasitol,* **129**, 11-21.

Dyer, N. A., Rose, C., Ejeh, N. O. & Acosta-Serrano, A. (2013). Flying tryps: survival and maturation of trypanosomes in tsetse flies. *Trends Parasitol,* **29**, 188-96.

Dymond, J. S. (2013). Explanatory chapter: quantitative PCR. *Methods Enzymol,* **529**, 279-89.

Eberwine, J., Sul, J.-Y., Bartfai, T. & Kim, J. (2014). The promise of single-cell sequencing. *Nat Meth,* **11**, 25-27.

El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D. & Andersson, B. (2005a). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science,* **309**, 409-15.

El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J. M., Cerqueira, G. C., Creasy, T., Delcher, A. L., Djikeng, A., Embley, T. M., Hauser, C., Ivens, A. C., Kummerfeld, S. K., Pereira-Leal, J. B., Nilsson, D., Peterson, J., Salzberg, S. L., Shallom, J., Silva, J. C., Sundaram, J., Westenberger, S., White, O., Melville, S. E.,

Donelson, J. E., Andersson, B., Stuart, K. D. & Hall, N. (2005b). Comparative genomics of trypanosomatid parasitic protozoa. *Science,* **309**, 404-9.

Elias, M. C., Faria, M., Mortara, R. A., Motta, M. C., de Souza, W., Thiry, M. & Schenkman, S. (2002). Chromosome localization changes in the Trypanosoma cruzi nucleus. *Eukaryot Cell,* **1**, 944-53.

Errington, J., Daniel, R. A. & Scheffers, D. J. (2003). Cytokinesis in bacteria. *Microbiol Mol Biol Rev,* **67**, 52-65, table of contents.

Ersfeld, K. (2011). Nuclear architecture, genome and chromatin organisation in Trypanosoma brucei. *Res Microbiol,* **162**, 626-36.

Erzberger, J. P. & Berger, J. M. (2006). Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct,* **35**, 93-114.

Erzberger, J. P., Pirruccello, M. M. & Berger, J. M. (2002). The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *Embo j,* **21**, 4763-73.

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B. & Speck, C. (2009). A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proc Natl Acad Sci U S A,* **106**, 20240-5.

Evrin, C., Fernandez-Cid, A., Riera, A., Zech, J., Clarke, P., Herrera, M. C., Tognetti, S., Lurz, R. & Speck, C. (2013a). The ORC/Cdc6/MCM2-7 complex facilitates MCM2-7 dimerization during prereplicative complex formation. *Nucleic Acids Res.*

Evrin, C., Fernandez-Cid, A., Zech, J., Herrera, M. C., Riera, A., Clarke, P., Brill, S., Lurz, R. & Speck, C. (2013b). In the absence of ATPase activity, pre-RC formation is blocked prior to MCM2-7 hexamer dimerization. *Nucleic Acids Res,* **41**, 3162-72.

Fenn, K. & Matthews, K. R. (2007). The cell biology of Trypanosoma brucei differentiation. *Curr Opin Microbiol,* **10**, 539-46.

Fernandez-Cid, A., Riera, A., Tognetti, S., Herrera, M. C., Samel, S., Evrin, C., Winkler, C., Gardenal, E., Uhle, S. & Speck, C. (2013). An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamer assembly. *Mol Cell,* **50**, 577-88.

Ferreira, M. F., Santocanale, C., Drury, L. S. & Diffley, J. F. (2000). Dbf4p, an essential S phase-promoting factor, is targeted for degradation by the anaphase-promoting complex. *Mol Cell Biol,* **20**, 242-8.

Fevre, E. M., Picozzi, K., Jannin, J., Welburn, S. C. & Maudlin, I. (2006). Human African trypanosomiasis: Epidemiology and control. *Adv Parasitol,* **61**, 167-221.

Field, M. C. & Carrington, M. (2009). The trypanosome flagellar pocket. *Nat Rev Microbiol,* **7**, 775-86.

Figueiredo, L. M., Cross, G. A. & Janzen, C. J. (2009). Epigenetic regulation in African trypanosomes: a new kid on the block. *Nat Rev Microbiol,* **7**, 504-13.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J. & Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Res,* **42**, D222-30.

Fletcher, R. J., Bishop, B. E., Leon, R. P., Sclafani, R. A., Ogata, C. M. & Chen, X. S. (2003). The structure and function of MCM from archaeal M. Thermoautotrophicum. *Nat Struct Biol,* **10**, 160-7.

Flynn, R. L., Centore, R. C., O'Sullivan, R. J., Rai, R., Tse, A., Songyang, Z., Chang, S., Karlseder, J. & Zou, L. (2011). TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature,* **471**, 532-6.

Franco, J. R., Simarro, P. P., Diarra, A. & Jannin, J. G. (2014). Epidemiology of human African trypanosomiasis. *Clin Epidemiol,* **6**, 257-75.

Frigola, J., Remus, D., Mehanna, A. & Diffley, J. F. (2013). ATPase-dependent quality control of DNA replication origin licensing. *Nature,* **495**, 339-43.

Froelich, C. A., Kang, S., Epling, L. B., Bell, S. P. & Enemark, E. J. (2014). A conserved MCM single-stranded DNA binding element is essential for replication initiation. *Elife,* **3**, e01993.

Fu, Y. V., Yardimci, H., Long, D. T., Ho, T. V., Guainazzi, A., Bermudez, V. P., Hurwitz, J., van Oijen, A., Scharer, O. D. & Walter, J. C. (2011). Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell,* **146**, 931-41.

Fujii, M., Ito, H., Hasegawa, T., Suzuki, T., Adachi, N. & Ayusawa, D. (2002). 5-Bromo-2'-deoxyuridine efficiently suppresses division potential of the yeast Saccharomyces cerevisiae. *Biosci Biotechnol Biochem,* **66**, 906-9.

Gaczynska, M., Osmulski, P. A., Jiang, Y., Lee, J. K., Bermudez, V. & Hurwitz, J. (2004). Atomic force microscopic analysis of the binding of the Schizosaccharomyces pombe origin recognition complex and the spOrc4 protein with origin DNA. *Proc Natl Acad Sci U S A,* **101**, 17952-7.

Gambus, A., Khoudoli, G. A., Jones, R. C. & Blow, J. J. (2011). MCM2-7 form double hexamers at licensed origins in Xenopus egg extract. *J Biol Chem,* **286**, 11855-64.

Gassen, A., Brechtefeld, D., Schandry, N., Arteaga-Salas, J. M., Israel, L., Imhof, A. & Janzen, C. J. (2012). DOT1A-dependent H3K76 methylation is required for replication regulation in Trypanosoma brucei. *Nucleic Acids Res,* **40**, 10302-11.

Gaudier, M., Schuwirth, B. S., Westcott, S. L. & Wigley, D. B. (2007). Structural basis of DNA replication origin recognition by an ORC protein. *Science,* **317**, 1213-6.

Ge, X. Q., Jackson, D. A. & Blow, J. J. (2007). Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev,* **21**, 3331-41.

Ghosh, S., Vassilev, A. P., Zhang, J., Zhao, Y. & DePamphilis, M. L. (2011). Assembly of the human origin recognition complex occurs through independent nuclear localization of its components. *J Biol Chem,* **286**, 23831-41.

Giardini, M. A., Segatto, M., da Silva, M. S., Nunes, V. S. & Cano, M. I. (2014). Telomere and telomerase biology. *Prog Mol Biol Transl Sci,* **125**, 1-40.

Ginger, M. L., Blundell, P. A., Lewis, A. M., Browitt, A., Gunzl, A. & Barry, J. D. (2002). Ex vivo and in vitro identification of a consensus promoter for VSG genes expressed by metacyclic-stage trypanosomes in the tsetse fly. *Eukaryot Cell,* **1**, 1000-9.

Giraldo, R. (2003). Common domains in the initiators of DNA replication in *Bacteria*, *Archaea* and *Eukarya*: combined structural, functional and phylogenetic perspectives. *FEMS Microbiol Rev,* **26**, 533-54.

Glover, L., Alsford, S. & Horn, D. (2013a). DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. *PLoS Pathog,* **9**, e1003260.

Glover, L. & Horn, D. (2014). Locus-specific control of DNA resection and suppression of subtelomeric VSG recombination by HAT3 in the African trypanosome. *Nucleic Acids Res*.

Glover, L., Hutchinson, S., Alsford, S., McCulloch, R., Field, M. C. & Horn, D. (2013b). Antigenic variation in African trypanosomes: the importance of chromosomal and nuclear context in VSG expression control. *Cell Microbiol,* **15**, 1984-93.

Godoy, P. D., Nogueira-Junior, L. A., Paes, L. S., Cornejo, A., Martins, R. M., Silber, A. M., Schenkman, S. & Elias, M. C. (2009). Trypanosome prereplication machinery contains a single functional orc1/cdc6 protein, which is typical of archaea. *Eukaryot Cell,* **8**, 1592-603.

Goranov, A. I., Katz, L., Breier, A. M., Burge, C. B. & Grossman, A. D. (2005). A transcriptional response to replication status mediated by the conserved bacterial replication protein DnaA. *Proc Natl Acad Sci U S A,* **102**, 12932-7.

Graham, B. W., Schauer, G. D., Leuba, S. H. & Trakselis, M. A. (2011). Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding. *Nucleic Acids Res,* **39**, 6585-95.

Grainge, I., Gaudier, M., Schuwirth, B. S., Westcott, S. L., Sandall, J., Atanassova, N. & Wigley, D. B. (2006). Biochemical analysis of a DNA replication origin in the archaeon Aeropyrum pernix. *J Mol Biol,* **363**, 355-69.

Gunzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L. C., Chung, H. M., Lee, P. T. & Lee, M. G. (2003). RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Eukaryot Cell,* **2**, 542-51.

Gunzl, A., Kirkham, J. K., Nguyen, T. N., Badjatia, N. & Park, S. H. (2014). Mono-allelic VSG expression by RNA polymerase I in Trypanosoma brucei: expression site control from both ends? *Gene*.

Gupta, A., Mehra, P., Deshmukh, A., Dar, A., Mitra, P., Roy, N. & Dhar, S. K. (2009). Functional dissection of the catalytic carboxyl-terminal domain of origin recognition complex subunit 1 (PfORC1) of the human malaria parasite Plasmodium falciparum. *Eukaryot Cell,* **8**, 1341-51.

Gupta, A., Mehra, P. & Dhar, S. K. (2008). *Plasmodium falciparum* origin recognition complex subunit 5: functional characterization and role in DNA replication foci formation. *Mol Microbiol,* **69**, 646-65.

Gupta, M. K., Atkinson, J. & McGlynn, P. (2010). DNA structure specificity conferred on a replicative helicase by its loader. *J Biol Chem,* **285**, 979-87.

Haenni, S., Renggli, C. K., Fragoso, C. M., Oberle, M. & Roditi, I. (2006). The procyclin-associated genes of Trypanosoma brucei are not essential for cyclical transmission by tsetse. *Mol Biochem Parasitol,* **150**, 144-56.

Haenni, S., Studer, E., Burkard, G. S. & Roditi, I. (2009). Bidirectional silencing of RNA polymerase I transcription by a strand switch region in Trypanosoma brucei. *Nucleic Acids Res,* **37**, 5007-18.

Hall, B. S., Gabernet-Castello, C., Voak, A., Goulding, D., Natesan, S. K. & Field, M. C. (2006). TbVps34, the trypanosome orthologue of Vps34, is required for Golgi complex segregation. *J Biol Chem,* **281**, 27600-12.

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

Hammarton, T. C., Clark, J., Douglas, F., Boshart, M. & Mottram, J. C. (2003). Stage-specific differences in cell cycle control in Trypanosoma brucei revealed by RNA interference of a mitotic cyclin. *J Biol Chem,* **278**, 22877-86.

Hanson, P. I. & Whiteheart, S. W. (2005). AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol,* **6**, 519-29.

Hara, K., Zheng, G., Qu, Q., Liu, H., Ouyang, Z., Chen, Z., Tomchick, D. R. & Yu, H. (2014). Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. *Nat Struct Mol Biol*.

Hartley, J. L., Temple, G. F. & Brasch, M. A. (2000). DNA cloning using in vitro site-specific recombination. *Genome Res,* **10**, 1788-95.

Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A. & Allers, T. (2013). Accelerated growth in the absence of DNA replication origins. *Nature,* **503**, 544-7.

Heller, R. C., Kang, S., Lam, W. M., Chen, S., Chan, C. S. & Bell, S. P. (2011). Eukaryotic origin-dependent DNA replication in vitro reveals sequential action of DDK and S-CDK kinases. *Cell,* **146**, 80-91.

Helmrich, A., Ballarino, M., Nudler, E. & Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. *Nat Struct Mol Biol,* **20**, 412-8.

Hemerly, A. S., Prasanth, S. G., Siddiqui, K. & Stillman, B. (2009). Orc1 controls centriole and centrosome copy number in human cells. *Science,* **323**, 789-93.

Hertz-Fowler, C., Figueiredo, L. M., Quail, M. A., Becker, M., Jackson, A., Bason, N., Brooks, K., Churcher, C., Fahkro, S., Goodhead, I., Heath, P., Kartvelishvili, M., Mungall, K., Harris, D., Hauser, H., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Taylor, J. E., Walker, D., White, B., Young, R., Cross, G. A., Rudenko, G., Barry, J. D., Louis, E. J. & Berriman, M. (2008). Telomeric expression sites are highly conserved in Trypanosoma brucei. *PLoS One,* **3**, e3527.

Heun, P., Laroche, T., Raghuraman, M. K. & Gasser, S. M. (2001). The positioning and dynamics of origins of replication in the budding yeast nucleus. *J Cell Biol,* **152**, 385-400.

Hirumi, H. & Hirumi, K. (1989). Continuous cultivation of Trypanosoma brucei blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J Parasitol,* **75**, 985-9.

Ho, H. H., He, C. Y., de Graffenried, C. L., Murrells, L. J. & Warren, G. (2006). Ordered assembly of the duplicating Golgi in Trypanosoma brucei. *Proc Natl Acad Sci U S A,* **103**, 7676-81.

Hoek, M., Engstler, M. & Cross, G. A. (2000). Expression-site-associated gene 8 (ESAG8) of Trypanosoma brucei is apparently essential and accumulates in the nucleolus. *J Cell Sci,* **113 (Pt 22)**, 3959-68.

Hopwood, B. & Dalton, S. (1996). Cdc45p assembles into a complex with Cdc46p/Mcm5p, is required for minichromosome maintenance, and is essential for chromosomal DNA replication. *Proc Natl Acad Sci U S A,* **93**, 12309-14.

Horn, D. & McCulloch, R. (2010). Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. *Curr Opin Microbiol,* **13**, 700-5.

Hou, Z., Bernstein, D. A., Fox, C. A. & Keck, J. L. (2005). Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. *Proc Natl Acad Sci U S A,* **102**, 8489-94.

Hu, H., Yu, Z., Liu, Y., Wang, T., Wei, Y. & Li, Z. (2014). The Aurora B kinase in Trypanosoma brucei undergoes post-translational modifications and is targeted to various subcellular locations through binding to TbCPC1. *Mol Microbiol,* **91**, 256-74.

Hughes, L., Towers, K., Starborg, T., Gull, K. & Vaughan, S. (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**, 5748-57.

Huijbregts, R. P., Svitin, A., Stinnett, M. W., Renfrow, M. B. & Chesnokov, I. (2009). Drosophila Orc6 facilitates GTPase activity and filament formation of the septin complex. *Mol Biol Cell,* **20**, 270-81.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., de Castro, E., Coggill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Fraser, M., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., McMenamin, C., Mi, H., Mutowo-Muellenet, P., Mulder, N., Natale, D., Orengo, C., Pesseat, S., Punta, M., Quinn, A. F., Rivoire, C., Sangrador-Vegas, A., Selengut, J. D., Sigrist, C. J., Scheremetjew, M., Tate, J., Thimmajanarthanan, M., Thomas, P. D., Wu, C. H., Yeats, C. & Yong, S. Y. (2012). InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res,* **40**, D306-12.

Iizuka, M., Matsui, T., Takisawa, H. & Smith, M. M. (2006). Regulation of replication licensing by acetyltransferase Hbo1. *Mol Cell Biol,* **26**, 1098-108.

Ilves, I., Petojevic, T., Pesavento, J. J. & Botchan, M. R. (2010). Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell,* **37**, 247-58.

Ishino, S., Fujino, S., Tomita, H., Ogino, H., Takao, K., Daiyasu, H., Kanai, T., Atomi, H. & Ishino, Y. (2011). Biochemical and genetical analyses of the three mcm genes from the hyperthermophilic archaeon, Thermococcus kodakarensis. *Genes Cells,* **16**, 1176-89.

Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchettin, G., Borzym, K., Bothe, G., Bruschi, C. V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R. M., Cronin, A., Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J. C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B., et al. (2005). The genome of the kinetoplastid parasite, Leishmania major. *Science,* **309**, 436-42.

Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol,* **146**, 11-31.

Jackson, A. P. (2014). Genome evolution in trypanosomatid parasites. *Parasitology*, 1-17.

Jackson, D., Wang, X. & Rudner, D. Z. (2012). Spatio-temporal organization of replication in bacteria and eukaryotes (nucleoids and nuclei). *Cold Spring Harb Perspect Biol,* **4**, a010389.

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

Jehi, S. E., Wu, F. & Li, B. (2014). Trypanosoma brucei TIF2 suppresses VSG switching by maintaining subtelomere integrity. *Cell Res,* **24**, 870-85.

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

Jensen, B. C., Sivam, D., Kifer, C. T., Myler, P. J. & Parsons, M. (2009). Widespread variation in transcript abundance within and across developmental stages of Trypanosoma brucei. *BMC Genomics,* **10**, 482.

Jensen, B. C., Wang, Q., Kifer, C. T. & Parsons, M. (2003). The NOG1 GTP-binding protein is required for biogenesis of the 60 S ribosomal subunit. *J Biol Chem,* **278**, 32204-11.

Jensen, R. E. & Englund, P. T. (2012). Network news: the replication of kinetoplast DNA. *Annu Rev Microbiol,* **66**, 473-91.

Jezewska, M. J., Rajendran, S., Bujalowska, D. & Bujalowski, W. (1998). Does single-stranded DNA pass through the inner channel of the protein hexamer in the complex with the Escherichia coli DnaB Helicase? Fluorescence energy transfer studies. *J Biol Chem,* **273**, 10515-29.

Johnson, A. & O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. *Annu Rev Biochem,* **74**, 283-315.

Jones, N. G., Thomas, E. B., Brown, E., Dickens, N. J., Hammarton, T. C. & Mottram, J. C. (2014). Regulators of Trypanosoma brucei cell cycle progression and differentiation identified using a kinome-wide RNAi screen. *PLoS Pathog,* **10**, e1003886.

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

Kaguni, J. M. (2011). Replication initiation at the Escherichia coli chromosomal origin. *Curr Opin Chem Biol,* **15**, 606-13.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H. & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nat. Protocols,* **7**, 1511-1522.

Kang, Y. H., Galal, W. C., Farina, A., Tappin, I. & Hurwitz, J. (2012). Properties of the human Cdc45/Mcm2-7/GINS helicase complex and its action with DNA polymerase epsilon in rolling circle DNA synthesis. *Proc Natl Acad Sci U S A,* **109**, 6042-7.

Kanke, M., Kodama, Y., Takahashi, T. S., Nakagawa, T. & Masukata, H. (2012). Mcm10 plays an essential role in origin DNA unwinding after loading of the CMG components. *Embo j,* **31**, 2182-94.

Kaplan, D. L. & O'Donnell, M. (2002). DnaB drives DNA branch migration and dislodges proteins while encircling two DNA strands. *Mol Cell,* **10**, 647-57.

Kaufmann, D., Gassen, A., Maiser, A., Leonhardt, H. & Janzen, C. J. (2012). Regulation and spatial organization of PCNA in *Trypanosoma brucei*. *Biochem Biophys Res Commun,* **419**, 698-702.

Kawakami, H. & Katayama, T. (2010). DnaA, ORC, and Cdc6: similarity beyond the domains of life and diversity. *Biochem Cell Biol,* **88**, 49-62.

Kawakami, H., Keyamura, K. & Katayama, T. (2005). Formation of an ATP-DnaA-specific initiation complex requires DnaA Arginine 285, a conserved motif in the AAA+ protein family. *J Biol Chem,* **280**, 27420-30.

Kaye, P. & Scott, P. (2011). Leishmaniasis: complexity at the host-pathogen interface. *Nat Rev Microbiol,* **9**, 604-15.

Kearsey, S. E. & Cotterill, S. (2003). Enigmatic variations: divergent modes of regulating eukaryotic DNA replication. *Mol Cell,* **12**, 1067-75.

Kelly, S., Reed, J., Kramer, S., Ellis, L., Webb, H., Sunter, J., Salje, J., Marinsek, N., Gull, K., Wickstead, B. & Carrington, M. (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**, 103-9.

Kelman, L. M. & Kelman, Z. (2014). Archaeal DNA replication. *Annu Rev Genet,* **48**, 71-97.

Keyamura, K., Fujikawa, N., Ishida, T., Ozaki, S., Su'etsugu, M., Fujimitsu, K., Kagawa, W., Yokoyama, S., Kurumizaka, H. & Katayama, T. (2007). The interaction of DiaA and DnaA regulates the replication cycle in E. coli by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev,* **21**, 2083-99.

Kim, H. S. & Cross, G. A. (2010). TOPO3alpha influences antigenic variation by monitoring expression-site-associated VSG switching in Trypanosoma brucei. *PLoS Pathog,* **6**, e1000992.

Kim, H. S. & Cross, G. A. (2011). Identification of Trypanosoma brucei RMI1/BLAP75 homologue and its roles in antigenic variation. *PLoS One,* **6**, e25313.

Kim, H. S., Park, S. H., Gunzl, A. & Cross, G. A. (2013). MCM-BP is required for repression of life-cycle specific genes transcribed by RNA polymerase I in the mammalian infectious form of *Trypanosoma brucei*. *PLoS One,* **8**, e57001.

Kim, S. M., Dubey, D. D. & Huberman, J. A. (2003). Early-replicating heterochromatin. *Genes Dev,* **17**, 330-5.

Klemm, R. D., Austin, R. J. & Bell, S. P. (1997). Coordinate binding of ATP and origin DNA regulates the ATPase activity of the origin recognition complex. *Cell,* **88**, 493-502.

Kogoma, T. (1997). Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev,* **61**, 212-38.

Kolev, N. G., Tschudi, C. & Ullu, E. (2011). RNA interference in protozoan parasites: achievements and challenges. *Eukaryot Cell,* **10**, 1156-63.

Kong, D. & DePamphilis, M. L. (2001). Site-specific DNA binding of the Schizosaccharomyces pombe origin recognition complex is determined by the Orc4 subunit. *Mol Cell Biol,* **21**, 8095-103.

Kramer, S. (2012). Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplastids. *Mol Biochem Parasitol,* **181**, 61-72.

Krastanova, I., Sannino, V., Amenitsch, H., Gileadi, O., Pisani, F. M. & Onesti, S. (2012). Structural and functional insights into the DNA replication factor

Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases. *J Biol Chem,* **287**, 4121-8.

Kumar, D., Kumar, D. & Saha, S. (2012). A highly basic sequence at the N-terminal region is essential for targeting the DNA replication protein ORC1 to the nucleus in *Leishmania donovani*. *Microbiology,* **158**, 1775-82.

Kumar, D., Minocha, N., Rajanala, K. & Saha, S. (2009). The distribution pattern of proliferating cell nuclear antigen in the nuclei of *Leishmania donovani*. *Microbiology,* **155**, 3748-57.

Kumar, D., Mukherji, A. & Saha, S. (2008). Expression and subcellular localization of ORC1 in *Leishmania major*. *Biochem Biophys Res Commun,* **375**, 74-9.

Kurokawa, K., Nishida, S., Emoto, A., Sekimizu, K. & Katayama, T. (1999). Replication cycle-coordinated change of the adenine nucleotide-bound forms of DnaA protein in Escherichia coli. *Embo j,* **18**, 6642-52.

Kurth, I. & Gautier, J. (2010). Origin-dependent initiation of DNA replication within telomeric sequences. *Nucleic Acids Res,* **38**, 467-76.

Lachaud, L., Bourgeois, N., Kuk, N., Morelle, C., Crobu, L., Merlin, G., Bastien, P., Pages, M. & Sterkers, Y. (2014). Constitutive mosaic aneuploidy is a unique genetic feature widespread in the Leishmania genus. *Microbes Infect,* **16**, 61-6.

LaCount, D. J., Bruse, S., Hill, K. L. & Donelson, J. E. (2000). Double-stranded RNA interference in Trypanosoma brucei using head-to-head promoters. *Mol Biochem Parasitol,* **111**, 67-76.

Landeira, D. & Navarro, M. (2007). Nuclear repositioning of the VSG promoter during developmental silencing in Trypanosoma brucei. *J Cell Biol,* **176**, 133-9.

Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Meth,* **9**, 357-359.

Langousis, G. & Hill, K. L. (2014). Motility and more: the flagellum of Trypanosoma brucei. *Nat Rev Microbiol,* **12**, 505-18.

Larsen, F., Gundersen, G., Lopez, R. & Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics,* **13**, 1095-107.

Lee, C., Hong, B., Choi, J. M., Kim, Y., Watanabe, S., Ishimi, Y., Enomoto, T., Tada, S., Kim, Y. & Cho, Y. (2004). Structural basis for inhibition of the replication licensing factor Cdt1 by geminin. *Nature,* **430**, 913-7.

Lee, J. K., Moon, K. Y., Jiang, Y. & Hurwitz, J. (2001). The Schizosaccharomyces pombe origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. *Proc Natl Acad Sci U S A,* **98**, 13589-94.

Lee, P. H., Meng, X. & Kapler, G. M. (2015). Developmental regulation of the Tetrahymena thermophila origin recognition complex. *PLoS Genet,* **11**, e1004875.

Leipe, D. D., Aravind, L., Grishin, N. V. & Koonin, E. V. (2000). The bacterial replicative helicase DnaB evolved from a RecA duplication. *Genome Res,* **10**, 5-16.

Leonard, A. C. & Mechali, M. (2013). DNA replication origins. *Cold Spring Harb Perspect Biol,* **5**, a010116.

Leprohon, P., Fernandez-Prada, C., Gazanion, É., Monte-Neto, R. & Ouellette, M. (2014). Drug resistance analysis by next generation sequencing in Leishmania. *International Journal for Parasitology: Drugs and Drug Resistance*.

Li, C. & Jin, J. (2010). DNA replication licensing control and rereplication prevention. *Protein Cell,* **1**, 227-36.

Li, C. J. & DePamphilis, M. L. (2002). Mammalian Orc1 protein is selectively released from chromatin and ubiquitinated during the S-to-M transition in the cell division cycle. *Mol Cell Biol,* **22**, 105-16.

Li, H. & Stillman, B. (2012). The origin recognition complex: a biochemical and structural view. *Subcell Biochem,* **62**, 37-58.

Li, J. J., Schnick, J., Hayles, J. & MacNeill, S. A. (2011a). Purification and functional inactivation of the fission yeast MCM(MCM-BP) complex. *FEBS Lett,* **585**, 3850-5.

Li, W., Kondratowicz, B., McWilliam, H., Nauche, S. & Lopez, R. (2013). The annotation-enriched non-redundant patent sequence databases. *Database (Oxford),* **2013**, bat005.

Li, Y. & Araki, H. (2013). Loading and activation of DNA replicative helicases: the key step of initiation of DNA replication. *Genes Cells,* **18**, 266-77.

Li, Z. (2012). Regulation of the cell division cycle in *Trypanosoma brucei*. *Eukaryot Cell,* **11**, 1180-90.

Li, Z., Pan, M., Santangelo, T. J., Chemnitz, W., Yuan, W., Edwards, J. L., Hurwitz, J., Reeve, J. N. & Kelman, Z. (2011b). A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res,* **39**, 6114-23.

Li, Z., Santangelo, T. J., Cubonova, L., Reeve, J. N. & Kelman, Z. (2010). Affinity purification of an archaeal DNA replication protein network. *MBio,* **1**.

Li, Z. & Wang, C. C. (2006). Changing roles of aurora-B kinase in two life cycle stages of Trypanosoma brucei. *Eukaryot Cell,* **5**, 1026-35.

Liang, C. & Stillman, B. (1997). Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants. *Genes Dev,* **11**, 3375-86.

Lindas, A. C. & Bernander, R. (2013). The cell cycle of archaea. *Nat Rev Microbiol,* **11**, 627-38.

Liu, B., Liu, Y., Motyka, S. A., Agbo, E. E. & Englund, P. T. (2005). Fellowship of the rings: the replication of kinetoplast DNA. *Trends Parasitol,* **21**, 363-9.

Liu, J., Phillips, B. T., Amaya, M. F., Kimble, J. & Xu, W. (2008). The C. elegans SYS-1 protein is a bona fide beta-catenin. *Dev Cell,* **14**, 751-61.

Liu, J., Smith, C. L., DeRyckere, D., DeAngelis, K., Martin, G. S. & Berger, J. M. (2000). Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. *Mol Cell,* **6**, 637-48.

Liu, S., Balasov, M., Wang, H., Wu, L., Chesnokov, I. N. & Liu, Y. (2011). Structural analysis of human Orc6 protein reveals a homology with transcription factor TFIIB. *Proc Natl Acad Sci U S A,* **108**, 7373-8.

Liu, Y., Richards, T. A. & Aves, S. J. (2009). Ancient diversification of eukaryotic MCM DNA replication proteins. *BMC Evol Biol,* **9**, 60.

Livak, K. J. & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* **25**, 402-8.

Lo, Y. H., Tsai, K. L., Sun, Y. J., Chen, W. T., Huang, C. Y. & Hsiao, C. D. (2009). The crystal structure of a replicative hexameric helicase DnaC and its complex with single-stranded DNA. *Nucleic Acids Res,* **37**, 804-14.

Lukes, J., Skalicky, T., Tyc, J., Votypka, J. & Yurchenko, V. (2014). Evolution of parasitism in kinetoplastid flagellates. *Mol Biochem Parasitol,* **195**, 115-22.

Lundgren, M., Andersson, A., Chen, L., Nilsson, P. & Bernander, R. (2004). Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A,* **101**, 7046-51.

Luo, K. Q., Elsasser, S., Chang, D. C. & Campbell, J. L. (2003). Regulation of the localization and stability of Cdc6 in living yeast cells. *Biochemical and Biophysical Research Communications,* **306**, 851-859.

Lygerou, Z. & Nurse, P. (1999). The fission yeast origin recognition complex is constitutively associated with chromatin and is differentially modified through the cell cycle. *J Cell Sci,* **112 (Pt 21)**, 3703-12.

MacAlpine, D. M., Rodriguez, H. K. & Bell, S. P. (2004). Coordination of replication and transcription along a Drosophila chromosome. *Genes Dev,* **18**, 3094-105.

MacAlpine, H. K., Gordan, R., Powell, S. K., Hartemink, A. J. & MacAlpine, D. M. (2010). Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res,* **20**, 201-11.

MacGregor, P., Savill, N. J., Hall, D. & Matthews, K. R. (2011). Transmission stages dominate trypanosome within-host dynamics during chronic infections. *Cell Host Microbe,* **9**, 310-8.

MacGregor, P., Szoor, B., Savill, N. J. & Matthews, K. R. (2012). Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act. *Nat Rev Microbiol,* **10**, 431-8.

Mackiewicz, P., Zakrzewska-Czerwinska, J., Zawilak, A., Dudek, M. R. & Cebrat, S. (2004). Where does bacterial replication start? Rules for predicting the oriC region. *Nucleic Acids Res,* **32**, 3781-91.

MacNeill, S. A. (2010). Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem J,* **425**, 489-500.

MacNeill, S. A. (2011). Protein-protein interactions in the archaeal core replisome. *Biochem Soc Trans,* **39**, 163-8.

Makarova, K. S., Koonin, E. V. & Kelman, Z. (2012). The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol Direct,* **7**, 7.

Makowska-Grzyska, M. & Kaguni, J. M. (2010). Primase directs the release of DnaC from DnaB. *Mol Cell,* **37**, 90-101.

Mannaert, A., Downing, T., Imamura, H. & Dujardin, J. C. (2012). Adaptive mechanisms in pathogens: universal aneuploidy in Leishmania. *Trends Parasitol,* **28**, 370-6.

Marcello, L. & Barry, J. D. (2007). Analysis of the VSG gene silent archive in Trypanosoma brucei reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome Res,* **17**, 1344-52.

Maree, J. P. & Patterton, H. G. (2014). The epigenome of Trypanosoma brucei: A regulatory interface to an unconventional transcriptional machine. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* **1839**, 743-750.

Marinsek, N., Barry, E. R., Makarova, K. S., Dionne, I., Koonin, E. V. & Bell, S. D. (2006). GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep,* **7**, 539-45.

Marsh, V. L., McGeoch, A. T. & Bell, S. D. (2006). Influence of chromatin and single strand binding proteins on the activity of an archaeal MCM. *J Mol Biol,* **357**, 1345-50.

Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N. & Oda, M. (2010). Eukaryotic chromosome DNA replication: where, when, and how? *Annu Rev Biochem,* **79**, 89-130.

Matsunaga, F., Forterre, P., Ishino, Y. & Myllykallio, H. (2001). In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc Natl Acad Sci U S A,* **98**, 11152-7.

Matsunaga, F., Glatigny, A., Mucchielli-Giorgi, M. H., Agier, N., Delacroix, H., Marisa, L., Durosay, P., Ishino, Y., Aggerbeck, L. & Forterre, P. (2007). Genomewide and biochemical analyses of DNA-binding activity of Cdc6/Orc1 and Mcm proteins in Pyrococcus sp. *Nucleic Acids Res,* **35**, 3214-22.

Matsunaga, F., Takemura, K., Akita, M., Adachi, A., Yamagami, T. & Ishino, Y. (2010). Localized melting of duplex DNA by Cdc6/Orc1 at the DNA replication origin in the hyperthermophilic archaeon Pyrococcus furiosus. *Extremophiles,* **14**, 21-31.

Matte, A. & Delbaere, L. T. J. (2010). ATP-binding Motifs. *eLS.* John Wiley & Sons, Ltd.

Matthews, K. R. (2005). The developmental cell biology of Trypanosoma brucei. *J Cell Sci,* **118**, 283-90.

Matthews, K. R. & Gull, K. (1997). Commitment to differentiation and cell cycle re-entry are coincident but separable events in the transformation of African trypanosomes from their bloodstream to their insect form. *J Cell Sci,* **110 (Pt 20)**, 2609-18.

McCulloch, R., Morrison, L. J. & Hall, J. P. (2014). DNA recombination strategies during antigenic variation in the african trypanosome. *In:* Gellert, M. & Craig, N. (eds.) *MDNA3-0016-2014.* American Society for Microbiology Press: Microbiology Spectrum.

McDonagh, P. D., Myler, P. J. & Stuart, K. (2000). The unusual gene organization of Leishmania major chromosome 1 may reflect novel transcription processes. *Nucleic Acids Res,* **28**, 2800-3.

McGarry, K. C., Ryan, V. T., Grimwade, J. E. & Leonard, A. C. (2004). Two discriminatory binding sites in the Escherichia coli replication origin are required for DNA strand opening by initiator DnaA-ATP. *Proc Natl Acad Sci U S A,* **101**, 2811-6.

McGeoch, A. T., Trakselis, M. A., Laskey, R. A. & Bell, S. D. (2005). Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. *Nat Struct Mol Biol,* **12**, 756-62.

McIntosh, D. & Blow, J. J. (2012). Dormant origins, the licensing checkpoint, and the response to replicative stresses. *Cold Spring Harb Perspect Biol,* **4**.

McKean, P. G. (2003). Coordination of cell cycle and cytokinesis in Trypanosoma brucei. *Curr Opin Microbiol,* **6**, 600-7.

McNairn, A. J., Okuno, Y., Misteli, T. & Gilbert, D. M. (2005). Chinese hamster ORC subunits dynamically associate with chromatin throughout the cell-cycle. *Exp Cell Res,* **308**, 345-56.

Mechali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol,* **11**, 728-38.

Melville, S. E., Leech, V., Gerrard, C. S., Tait, A. & Blackwell, J. M. (1998). The molecular karyotype of the megabase chromosomes of Trypanosoma brucei and the assignment of chromosome markers. *Mol Biochem Parasitol,* **94**, 155-73.

Messer, W. (2002). The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev,* **26**, 355-74.

Messer, W., Blaesing, F., Majka, J., Nardmann, J., Schaper, S., Schmidt, A., Seitz, H., Speck, C., Tungler, D., Wegrzyn, G., Weigel, C., Welzeck, M. & Zakrzewska-Czerwinska, J. (1999). Functional domains of DnaA proteins. *Biochimie,* **81**, 819-25.

Michaeli, S. (2011). Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol,* **6**, 459-74.

Miller, J. M., Arachea, B. T., Epling, L. B. & Enemark, E. J. (2014). Analysis of the crystal structure of an active MCM hexamer. *Elife,* **3**, e03433.

Minocha, N., Kumar, D., Rajanala, K. & Saha, S. (2011). Characterization of *Leishmania donovani* MCM4: expression patterns and interaction with PCNA. *PLoS One,* **6**, e23107.

Miotto, B. & Struhl, K. (2010). HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. *Mol Cell,* **37**, 57-66.

Mohammad, M., York, R. D., Hommel, J. & Kapler, G. M. (2003). Characterization of a novel origin recognition complex-like complex: implications for DNA recognition, cell cycle control, and locus-specific gene amplification. *Mol Cell Biol,* **23**, 5005-17.

Mohammad, M. M., Donti, T. R., Sebastian Yakisich, J., Smith, A. G. & Kapler, G. M. (2007). Tetrahymena ORC contains a ribosomal RNA fragment that participates in rDNA origin recognition. *Embo j,* **26**, 5048-60.

Moldovan, G. L., Pfander, B. & Jentsch, S. (2007). PCNA, the maestro of the replication fork. *Cell,* **129**, 665-79.

Mony, B. M., MacGregor, P., Ivens, A., Rojas, F., Cowton, A., Young, J., Horn, D. & Matthews, K. (2014). Genome-wide dissection of the quorum sensing signalling pathway in Trypanosoma brucei. *Nature,* **505**, 681-5.

Moore, A., Donahue, C. J., Bauer, K. D. & Mather, J. P. (1998). Simultaneous measurement of cell cycle and apoptotic cell death. *Methods Cell Biol,* **57**, 265-78.

Moreau, M. J., McGeoch, A. T., Lowe, A. R., Itzhaki, L. S. & Bell, S. D. (2007). ATPase site architecture and helicase mechanism of an archaeal MCM. *Mol Cell,* **28**, 304-14.

Moreira, D., Lopez-Garcia, P. & Vickerman, K. (2004). An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *Int J Syst Evol Microbiol,* **54**, 1861-75.

Morty, R. E., Lonsdale-Eccles, J. D., Morehead, J., Caler, E. V., Mentele, R., Auerswald, E. A., Coetzer, T. H., Andrews, N. W. & Burleigh, B. A. (1999). Oligopeptidase B from Trypanosoma brucei, a new member of an emerging subgroup of serine oligopeptidases. *J Biol Chem,* **274**, 26149-56.

Mott, M. L. & Berger, J. M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. *Nat Rev Microbiol,* **5**, 343-54.

Mott, M. L., Erzberger, J. P., Coons, M. M. & Berger, J. M. (2008). Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell,* **135**, 623-34.

Moyer, S. E., Lewis, P. W. & Botchan, M. R. (2006). Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci U S A,* **103**, 10236-41.

Muller, C. A., Hawkins, M., Retkute, R., Malla, S., Wilson, R., Blythe, M. J., Nakato, R., Komata, M., Shirahige, K., de Moura, A. P. & Nieduszynski, C.

A. (2014). The dynamics of genome replication using deep sequencing. *Nucleic Acids Res,* **42**, e3.

Muller, C. A. & Nieduszynski, C. A. (2012). Conservation of replication timing reveals global and local regulation of replication origin activity. *Genome Res,* **22**, 1953-62.

Munday, J. C., McLuskey, K., Brown, E., Coombs, G. H. & Mottram, J. C. (2011). Oligopeptidase B deficient mutants of Leishmania major. *Mol Biochem Parasitol*, **175**, 49-57.

Muramatsu, S., Hirai, K., Tak, Y. S., Kamimura, Y. & Araki, H. (2010). CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon}, and GINS in budding yeast. *Genes Dev,* **24**, 602-12.

Myler, P. J., Audleman, L., deVos, T., Hixson, G., Kiser, P., Lemley, C., Magness, C., Rickel, E., Sisk, E., Sunkin, S., Swartzell, S., Westlake, T., Bastien, P., Fu, G., Ivens, A. & Stuart, K. (1999). Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci U S A,* **96**, 2902-6.

Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H. & Forterre, P. (2000). Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science,* **288**, 2212-5.

Narayanan, M. S. & Rudenko, G. (2013). TDP1 is an HMG chromatin protein facilitating RNA polymerase I transcription in African trypanosomes. *Nucleic Acids Res,* **41**, 2981-92.

Natesan, S. K., Peacock, L., Matthews, K., Gibson, W. & Field, M. C. (2007). Activation of endocytosis as an adaptation to the mammalian host by trypanosomes. *Eukaryot Cell,* **6**, 2029-37.

Navarro, M. & Gull, K. (2001). A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. *Nature,* **414**, 759-63.

Nawy, T. (2014). Single-cell sequencing. *Nat Meth,* **11**, 18-18.

Názer, E. & Sánchez, D. O. (2011). Nucleolar Accumulation of RNA Binding Proteins Induced by ActinomycinD Is Functional in <italic>Trypanosoma cruzi</italic> and <italic>Leishmania mexicana</italic> but Not in <italic>T. brucei</italic>. *PLoS ONE,* **6**, e24184.

Nguyen, T., Jagannathan, M., Shire, K. & Frappier, L. (2012). Interactions of the human MCM-BP protein with MCM complex components and Dbf4. *PLoS One,* **7**, e35931.

Nguyen, V. Q., Co, C., Irie, K. & Li, J. J. (2000). Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. *Curr Biol,* **10**, 195-205.

Nguyen, V. Q., Co, C. & Li, J. J. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature,* **411**, 1068-73.

Nishitani, H., Lygerou, Z., Nishimoto, T. & Nurse, P. (2000). The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature,* **404**, 625-8.

Nishitani, H., Taraviras, S., Lygerou, Z. & Nishimoto, T. (2001). The human licensing factor for DNA replication Cdt1 accumulates in G1 and is destabilized after initiation of S-phase. *J Biol Chem,* **276**, 44905-11.

Nishiyama, A., Frappier, L. & Mechali, M. (2011). MCM-BP regulates unloading of the MCM2-7 helicase in late S phase. *Genes Dev,* **25**, 165-75.

Norais, C., Hawkins, M., Hartman, A. L., Eisen, J. A., Myllykallio, H. & Allers, T. (2007). Genetic and physical mapping of DNA replication origins in Haloferax volcanii. *PLoS Genet,* **3**, e77.

Nordlund, P. & Reichard, P. (2006). Ribonucleotide reductases. *Annu Rev Biochem,* **75**, 681-706.

Norio, P., Kosiyatrakul, S., Yang, Q., Guan, Z., Brown, N. M., Thomas, S., Riblet, R. & Schildkraut, C. L. (2005). Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. *Mol Cell,* **20**, 575-87.

O'Donnell, M., Langston, L. & Stillman, B. (2013). Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb Perspect Biol,* **5**.

Obado, S. O., Bot, C., Nilsson, D., Andersson, B. & Kelly, J. M. (2007). Repetitive DNA is associated with centromeric domains in Trypanosoma brucei but not Trypanosoma cruzi. *Genome Biol,* **8**, R37.

Obado, S. O., Taylor, M. C., Wilkinson, S. R., Bromley, E. V. & Kelly, J. M. (2005). Functional mapping of a trypanosome centromere by chromosome fragmentation identifies a 16-kb GC-rich transcriptional "strand-switch" domain as a major feature. *Genome Res,* **15**, 36-43.

Oberholzer, M., Morand, S., Kunz, S. & Seebeck, T. (2006). A vector series for rapid PCR-mediated C-terminal in situ tagging of Trypanosoma brucei genes. *Mol Biochem Parasitol,* **145**, 117-20.

Ogbadoyi, E., Ersfeld, K., Robinson, D., Sherwin, T. & Gull, K. (2000). Architecture of the Trypanosoma brucei nucleus during interphase and mitosis. *Chromosoma,* **108**, 501-13.

Ogino, H., Ishino, S., Mayanagi, K., Haugland, G. T., Birkeland, N. K., Yamagishi, A. & Ishino, Y. (2011). The GINS complex from the thermophilic archaeon, Thermoplasma acidophilum may function as a homotetramer in DNA replication. *Extremophiles,* **15**, 529-39.

Ogura, Y., Imai, Y., Ogasawara, N. & Moriya, S. (2001). Autoregulation of the dnaA-dnaN operon and effects of DnaA protein levels on replication initiation in Bacillus subtilis. *J Bacteriol,* **183**, 3833-41.

Olliver, A., Saggioro, C., Herrick, J. & Sclavi, B. (2010). DnaA-ATP acts as a molecular switch to control levels of ribonucleotide reductase expression in Escherichia coli. *Mol Microbiol,* **76**, 1555-71.

Ooi, C. P. & Bastin, P. (2013). More than meets the eye: understanding Trypanosoma brucei morphology in the tsetse. *Front Cell Infect Microbiol,* **3**, 71.

Oshiro, G., Owens, J. C., Shellman, Y., Sclafani, R. A. & Li, J. J. (1999). Cell cycle control of Cdc7p kinase activity through regulation of Dbf4p stability. *Mol Cell Biol,* **19**, 4888-96.

Overath, P. & Engstler, M. (2004). Endocytosis, membrane recycling and sorting of GPI-anchored proteins: Trypanosoma brucei as a model system. *Mol Microbiol,* **53**, 735-44.

Oyola, S. O., Bringaud, F. & Melville, S. E. (2009). A kinetoplastid BRCA2 interacts with DNA replication protein CDC45. *Int J Parasitol,* **39**, 59-69.

Ozaydin, B. & Rine, J. (2010). Expanded roles of the origin recognition complex in the architecture and function of silenced chromatin in Saccharomyces cerevisiae. *Mol Cell Biol,* **30**, 626-39.

Pace, N. R., Sapp, J. & Goldenfeld, N. (2012). Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life. *Proc Natl Acad Sci U S A,* **109**, 1011-8.

Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C. V., Hau, J., Martin, O., Kuznetsov, D. & Falquet, L. (2007). MyHits: improvements to

an interactive resource for analyzing protein sequences. *Nucleic Acids Res,* **35**, W433-7.

Pak, D. T., Pflumm, M., Chesnokov, I., Huang, D. W., Kellum, R., Marr, J., Romanowski, P. & Botchan, M. R. (1997). Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. *Cell,* **91**, 311-23.

Pan, M., Santangelo, T. J., Li, Z., Reeve, J. N. & Kelman, Z. (2011). Thermococcus kodakarensis encodes three MCM homologs but only one is essential. *Nucleic Acids Res,* **39**, 9671-80.

Pandya, U. M., Sandhu, R. & Li, B. (2013). Silencing subtelomeric VSGs by Trypanosoma brucei RAP1 at the insect stage involves chromatin structure changes. *Nucleic Acids Res,* **41**, 7673-82.

Papadopoulou, B., Roy, G. & Ouellette, M. (1994). Autonomous replication of bacterial DNA plasmid oligomers in Leishmania. *Mol Biochem Parasitol,* **65**, 39-49.

Park, J. H., Jensen, B. C., Kifer, C. T. & Parsons, M. (2001). A novel nucleolar G-protein conserved in eukaryotes. *J Cell Sci,* **114**, 173-185.

Park, S. Y. & Asano, M. (2011). Function of the origin recognition complex 1 (ORC1) outside DNA replication in Drosophila. *Cell Cycle,* **10**, 3957-63.

Park, S. Y. & Asano, M. (2012). An orc1 allele with a mutated APC motif is female sterile with amplification defects. *Cell Cycle,* **11**, 2828-32.

Petersen, B. O., Lukas, J., Sorensen, C. S., Bartek, J. & Helin, K. (1999). Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. *Embo j,* **18**, 396-410.

Pfeiffer, V. & Lingner, J. (2013). Replication of telomeres and the regulation of telomerase. *Cold Spring Harb Perspect Biol,* **5**, a010405.

Piatti, S., Lengauer, C. & Nasmyth, K. (1995). Cdc6 is an unstable protein whose de novo synthesis in G1 is important for the onset of S phase and for preventing a 'reductional' anaphase in the budding yeast Saccharomyces cerevisiae. *Embo j,* **14**, 3788-99.

Ploubidou, A., Robinson, D. R., Docherty, R. C., Ogbadoyi, E. O. & Gull, K. (1999). Evidence for novel cell cycle checkpoints in trypanosomes: kinetoplast segregation and cytokinesis in the absence of mitosis. *J Cell Sci,* **112 (Pt 24)**, 4641-50.

Prasanth, S. G., Prasanth, K. V. & Stillman, B. (2002). Orc6 involved in DNA replication, chromosome segregation, and cytokinesis. *Science,* **297**, 1026-31.

R Development Core Team (2010). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Raghuraman, M. K., Winzeler, E. A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D. J., Davis, R. W., Brewer, B. J. & Fangman, W. L. (2001). Replication dynamics of the yeast genome. *Science,* **294**, 115-21.

Ramer, M. D., Suman, E. S., Richter, H., Stanger, K., Spranger, M., Bieberstein, N. & Duncker, B. P. (2013). Dbf4 and Cdc7 proteins promote DNA replication through interactions with distinct Mcm2-7 protein subunits. *J Biol Chem,* **288**, 14926-35.

Randell, J. C., Bowers, J. L., Rodriguez, H. K. & Bell, S. P. (2006). Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell,* **21**, 29-39.

Randell, J. C., Fan, A., Chan, C., Francis, L. I., Heller, R. C., Galani, K. & Bell, S. P. (2010). Mec1 is one of multiple kinases that prime the Mcm2-7 helicase for phosphorylation by Cdc7. *Mol Cell,* **40**, 353-63.

Ravel, C., Dubessay, P., Bastien, P., Blackwell, J. M. & Ivens, A. C. (1998). The complete chromosomal organization of the reference strain of the Leishmania Genome Project, L. major ;Friedlin'. *Parasitol Today,* **14**, 301-3.

Raymann, K., Forterre, P., Brochier-Armanet, C. & Gribaldo, S. (2014). Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. *Genome Biol Evol,* **6**, 192-212.

Redmond, S., Vadivelu, J. & Field, M. C. (2003). RNAit: an automated web-based tool for the selection of RNAi targets in Trypanosoma brucei. *Mol Biochem Parasitol,* **128**, 115-8.

Remus, D., Beall, E. L. & Botchan, M. R. (2004). DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *Embo j,* **23**, 897-907.

Remus, D., Beuron, F., Tolun, G., Griffith, J. D., Morris, E. P. & Diffley, J. F. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell,* **139**, 719-30.

Reuner, B., Vassella, E., Yutzy, B. & Boshart, M. (1997). Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. *Mol Biochem Parasitol,* **90**, 269-80.

Reynolds, D., Cliffe, L., Forstner, K. U., Hon, C. C., Siegel, T. N. & Sabatini, R. (2014). Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in Leishmania major and Trypanosoma brucei. *Nucleic Acids Res.*

Rhind, N. & Gilbert, D. M. (2013). DNA replication timing. *Cold Spring Harb Perspect Biol,* **5**, a010132.

Richardson, J. P., Beecroft, R. P., Tolson, D. L., Liu, M. K. & Pearson, T. W. (1988). Procyclin: an unusual immunodominant glycoprotein surface antigen from the procyclic stage of African trypanosomes. *Mol Biochem Parasitol,* **31**, 203-16.

Robinson, D. R., Sherwin, T., Ploubidou, A., Byard, E. H. & Gull, K. (1995). Microtubule polarity and dynamics in the control of organelle positioning, segregation, and cytokinesis in the trypanosome cell cycle. *J Cell Biol,* **128**, 1163-72.

Robinson, N. P. & Bell, S. D. (2005). Origins of DNA replication in the three domains of life. *Febs j,* **272**, 3757-66.

Robinson, N. P. & Bell, S. D. (2007). Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci U S A,* **104**, 5806-11.

Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R. & Bell, S. D. (2004). Identification of two origins of replication in the single chromosome of the archaeon Sulfolobus solfataricus. *Cell,* **116**, 25-38.

Rogers, M. B., Hilley, J. D., Dickens, N. J., Wilkes, J., Bates, P. A., Depledge, D. P., Harris, D., Her, Y., Herzyk, P., Imamura, H., Otto, T. D., Sanders, M., Seeger, K., Dujardin, J. C., Berriman, M., Smith, D. F., Hertz-Fowler, C. & Mottram, J. C. (2011). Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. *Genome Res,* **21**, 2129-42.

Romanowski, P., Madine, M. A., Rowles, A., Blow, J. J. & Laskey, R. A. (1996). The Xenopus origin recognition complex is essential for DNA replication and MCM binding to chromatin. *Curr Biol,* **6**, 1416-25.

Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. (2002). A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective

"ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl,* **41**, 2596-9.

Rothenberg, E., Trakselis, M. A., Bell, S. D. & Ha, T. (2007). MCM forked substrate specificity involves dynamic interaction with the 5'-tail. *J Biol Chem,* **282**, 34229-34.

Rotureau, B., Subota, I., Buisson, J. & Bastin, P. (2012). A new asymmetric division contributes to the continuous production of infective trypanosomes in the tsetse fly. *Development,* **139**, 1842-50.

Rout, M. P. & Field, M. C. (2001). Isolation and characterization of subnuclear compartments from Trypanosoma brucei. Identification of a major repetitive nuclear lamina component. *J Biol Chem,* **276**, 38261-71.

Roy, A., Kucukural, A. & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc,* **5**, 725-38.

Roy, G., Dumas, C., Sereno, D., Wu, Y., Singh, A. K., Tremblay, M. J., Ouellette, M., Olivier, M. & Papadopoulou, B. (2000). Episomal and stable expression of the luciferase reporter gene for quantifying Leishmania spp. infections in macrophages and in animal models. *Mol Biochem Parasitol,* **110**, 195-206.

Rudd, S. G., Glover, L., Jozwiakowski, S. K., Horn, D. & Doherty, A. J. (2013). PPL2 translesion polymerase is essential for the completion of chromosomal DNA replication in the African trypanosome. *Mol Cell,* **52**, 554-65.

Rudenko, G. (2010). Epigenetics and transcriptional control in African trypanosomes. *Essays Biochem,* **48**, 201-19.

Ryan, V. T., Grimwade, J. E., Camara, J. E., Crooke, E. & Leonard, A. C. (2004). Escherichia coli prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. *Mol Microbiol,* **51**, 1347-59.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T. C., Robins, A. J., Dalton, S. & Gilbert, D. M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res,* **20**, 761-70.

Saha, S., Shan, Y., Mesner, L. D. & Hamlin, J. L. (2004). The promoter of the Chinese hamster ovary dihydrofolate reductase gene regulates the activity of the local origin and helps define its boundaries. *Genes Dev,* **18**, 397-410.

Sakwe, A. M., Nguyen, T., Athanasopoulos, V., Shire, K. & Frappier, L. (2007). Identification and characterization of a novel component of the human minichromosome maintenance complex. *Mol Cell Biol,* **27**, 3044-55.

Salic, A. & Mitchison, T. J. (2008). A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proc Natl Acad Sci U S A,* **105**, 2415-20.

Sambrook, J. F. & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*. 3rd ed.: Cold Spring Harbor Laboratory Press.

Sasaki, T. & Gilbert, D. M. (2007). The many faces of the origin recognition complex. *Curr Opin Cell Biol,* **19**, 337-43.

Schaper, S. & Messer, W. (1995). Interaction of the initiator protein DnaA of Escherichia coli with its DNA target. *J Biol Chem,* **270**, 17622-6.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. & Cardona, A.

(2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods,* **9**, 676-82.

Schlimme, W., Burri, M., Bender, K., Betschart, B. & Hecker, H. (1993). Trypanosoma brucei brucei: differences in the nuclear chromatin of bloodstream forms and procyclic culture forms. *Parasitology,* **107 (Pt 3)**, 237-47.

Scholefield, G., Veening, J. W. & Murray, H. (2011). DnaA and ORC: more than DNA replication initiators. *Trends Cell Biol,* **21**, 188-94.

Segurado, M., de Luis, A. & Antequera, F. (2003). Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe. *EMBO Rep,* **4**, 1048-53.

Seitz, H., Weigel, C. & Messer, W. (2000). The interaction domains of the DnaA and DnaB replication proteins of Escherichia coli. *Mol Microbiol,* **37**, 1270-9.

Sekimizu, K., Bramhill, D. & Kornberg, A. (1987). ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the E. coli chromosome. *Cell,* **50**, 259-65.

Semple, J. W., Da-Silva, L. F., Jervis, E. J., Ah-Kee, J., Al-Attar, H., Kummer, L., Heikkila, J. J., Pasero, P. & Duncker, B. P. (2006). An essential role for Orc6 in DNA replication through maintenance of pre-replicative complexes. *Embo j,* **25**, 5150-8.

Sequeira-Mendes, J., Diaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N. & Gomez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet,* **5**, e1000446.

Sequeira-Mendes, J. & Gomez, M. (2012). On the opportunistic nature of transcription and replication initiation in the metazoan genome. *Bioessays,* **34**, 119-25.

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W. & Van der Oost, J. (2001). The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proc Natl Acad Sci U S A,* **98**, 7835-40.

Shen, S., Arhin, G. K., Ullu, E. & Tschudi, C. (2001). In vivo epitope tagging of Trypanosoma brucei genes using a one step PCR-based strategy. *Mol Biochem Parasitol,* **113**, 171-3.

Shi, H., Djikeng, A., Mark, T., Wirtz, E., Tschudi, C. & Ullu, E. (2000). Genetic interference in Trypanosoma brucei by heritable and inducible double-stranded RNA. *Rna,* **6**, 1069-76.

Shreeram, S., Sparks, A., Lane, D. P. & Blow, J. J. (2002). Cell type-specific responses of human cells to inhibition of replication licensing. *Oncogene,* **21**, 6624-32.

Siddiqui, K., On, K. F. & Diffley, J. F. (2013). Regulating DNA replication in eukarya. *Cold Spring Harb Perspect Biol,* **5**.

Siddiqui, K. & Stillman, B. (2007). ATP-dependent assembly of the human origin recognition complex. *J Biol Chem,* **282**, 32370-83.

Siegel, T. N., Gunasekera, K., Cross, G. A. & Ochsenreiter, T. (2011). Gene expression in Trypanosoma brucei: lessons from high-throughput RNA sequencing. *Trends Parasitol,* **27**, 434-41.

Siegel, T. N., Hekstra, D. R. & Cross, G. A. (2008). Analysis of the Trypanosoma brucei cell cycle by quantitative DAPI imaging. *Mol Biochem Parasitol,* **160**, 171-4.

Siegel, T. N., Hekstra, D. R., Kemp, L. E., Figueiredo, L. M., Lowell, J. E., Fenyo, D., Wang, X., Dewell, S. & Cross, G. A. (2009). Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. *Genes Dev,* **23**, 1063-76.

Siegel, T. N., Hekstra, D. R., Wang, X., Dewell, S. & Cross, G. A. (2010). Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. *Nucleic Acids Res,* **38**, 4946-57.

Sigrist, C. J., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L. & Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Res,* **41**, D344-7.

Simarro, P. P., Cecchi, G., Franco, J. R., Paone, M., Diarra, A., Ruiz-Postigo, J. A., Fevre, E. M., Mattioli, R. C. & Jannin, J. G. (2012). Estimating and mapping the population at risk of sleeping sickness. *PLoS Negl Trop Dis,* **6**, e1859.

Simarro, P. P., Cecchi, G., Paone, M., Franco, J. R., Diarra, A., Ruiz, J. A., Fevre, E. M., Courtin, F., Mattioli, R. C. & Jannin, J. G. (2010). The Atlas of human African trypanosomiasis: a contribution to global mapping of neglected tropical diseases. *Int J Health Geogr,* **9**, 57.

Simmons, L. A., Felczak, M. & Kaguni, J. M. (2003). DnaA Protein of Escherichia coli: oligomerization at the E. coli chromosomal origin is required for initiation and involves specific N-terminal amino acids. *Mol Microbiol,* **49**, 849-58.

Simpson, A. G., Stevens, J. R. & Lukes, J. (2006). The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol,* **22**, 168-74.

Simpson, R. T. (1990). Nucleosome positioning can affect the function of a cis-acting DNA element in vivo. *Nature,* **343**, 387-9.

Skarstad, K. & Katayama, T. (2013). Regulating DNA replication in bacteria. *Cold Spring Harb Perspect Biol,* **5**, a012922.

Slaymaker, I. M. & Chen, X. S. (2012). MCM structure and mechanics: what we have learned from archaeal MCM. *Subcell Biochem,* **62**, 89-111.

Sonneville, R., Querenet, M., Craig, A., Gartner, A. & Blow, J. J. (2012). The dynamics of replication licensing in live Caenorhabditis elegans embryos. *J Cell Biol,* **196**, 233-46.

Speck, C., Chen, Z., Li, H. & Stillman, B. (2005). ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nat Struct Mol Biol,* **12**, 965-71.

Speck, C. & Messer, W. (2001). Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *Embo j,* **20**, 1469-76.

Speck, C. & Stillman, B. (2007). Cdc6 ATPase activity regulates ORC.Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication. *J Biol Chem,* **282**, 11705-14.

Speck, C., Weigel, C. & Messer, W. (1999). ATP- and ADP-dnaA protein, a molecular switch in gene regulation. *Embo j,* **18**, 6169-76.

Sterkers, Y., Lachaud, L., Bourgeois, N., Crobu, L., Bastien, P. & Pages, M. (2012). Novel insights into genome plasticity in Eukaryotes: mosaic aneuploidy in Leishmania. *Mol Microbiol,* **86**, 15-23.

Sterkers, Y., Lachaud, L., Crobu, L., Bastien, P. & Pages, M. (2011). FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in Leishmania major. *Cell Microbiol,* **13**, 274-83.

Steverding, D. (2008). The history of African trypanosomiasis. *Parasit Vectors,* **1**, 3.

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

Stuart, K. D. & Myler, P. J. (2006). Comparative genomics of the trypanosomatids. *In:* Katz, L. A. & Bhattacharya, D. (eds.) *Genomics and Evolution of Microbial Eukaryotes.* United States: Oxford University Press Inc., New York.

Sun, J., Evrin, C., Samel, S. A., Fernandez-Cid, A., Riera, A., Kawakami, H., Stillman, B., Speck, C. & Li, H. (2013). Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA. *Nat Struct Mol Biol,* **20**, 944-51.

Sun, J., Kawakami, H., Zech, J., Speck, C., Stillman, B. & Li, H. (2012). Cdc6-induced conformational changes in ORC bound to origin DNA revealed by cryo-electron microscopy. *Structure,* **20**, 534-44.

Szoor, B., Ruberto, I., Burchmore, R. & Matthews, K. R. (2010). A novel phosphatase cascade regulates differentiation in Trypanosoma brucei via a glycosomal signaling pathway. *Genes Dev,* **24**, 1306-16.

Szoor, B., Wilson, J., McElhinney, H., Tabernero, L. & Matthews, K. R. (2006). Protein tyrosine phosphatase TbPTP1: A molecular switch controlling life cycle differentiation in trypanosomes. *J Cell Biol,* **175**, 293-303.

Takara, T. J. & Bell, S. P. (2011). Multiple Cdt1 molecules act at each origin to load replication-competent Mcm2-7 helicases. *EMBO J,* **30**, 4885-96.

Tamar, S. & Papadopoulou, B. (2001). A telomere-mediated chromosome fragmentation approach to assess mitotic stability and ploidy alterations of Leishmania chromosomes. *J Biol Chem,* **276**, 11662-73.

Tanaka, S. & Araki, H. (2013). Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harb Perspect Biol,* **5**, a010371.

Tanaka, S. & Diffley, J. F. (2002). Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nat Cell Biol,* **4**, 198-207.

Tanaka, S., Nakato, R., Katou, Y., Shirahige, K. & Araki, H. (2011). Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Curr Biol,* **21**, 2055-63.

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y. & Araki, H. (2007). CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature,* **445**, 328-32.

Tatsumi, Y., Tsurimoto, T., Shirahige, K., Yoshikawa, H. & Obuse, C. (2000). Association of human origin recognition complex 1 with chromatin DNA and nuclease-resistant nuclear structures. *J Biol Chem,* **275**, 5904-10.

Tetley, L., Turner, C. M., Barry, J. D., Crowe, J. S. & Vickerman, K. (1987). Onset of expression of the variant surface glycoproteins of Trypanosoma brucei in the tsetse fly studied using immunoelectron microscopy. *J Cell Sci,* **87 (Pt 2)**, 363-72.

Tetley, L. & 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**, 1-19.

Thomas, S., Green, A., Sturm, N. R., Campbell, D. A. & Myler, P. J. (2009). Histone acetylations mark origins of polycistronic transcription in Leishmania major. *BMC Genomics,* **10**, 152.

Tiengwe, C. (2010). *Analysis of the architecture and function of the nuclear DNA replication apparatus in Trypanosoma brucei*. PhD PhD, University of Glasgow.

Tiengwe, C., Marcello, L., Farr, H., Dickens, N., Kelly, S., Swiderski, M., Vaughan, D., Gull, K., Barry, J. D., Bell, S. D. & McCulloch, R. (2012a). Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of *Trypanosoma brucei*. *Cell Rep,* **2**, 185-97.

Tiengwe, C., Marcello, L., Farr, H., Gadelha, C., Burchmore, R., Barry, J. D., Bell, S. D. & McCulloch, R. (2012b). Identification of ORC1/CDC6-interacting factors in *Trypanosoma brucei* reveals critical features of origin recognition complex architecture. *PLoS One,* **7**, e32674.

Tiengwe, C., Marques, C. A. & McCulloch, R. (2013). Nuclear DNA replication initiation in kinetoplastid parasites: new insights into an ancient process. *Trends Parasitol*.

Trenaman, A., Hartley, C., Prorocic, M., Passos-Silva, D. G., van den Hoek, M., Nechyporuk-Zloy, V., Machado, C. R. & McCulloch, R. (2013). Trypanosoma brucei BRCA2 acts in a life cycle-specific genome stability process and dictates BRC repeat number-dependent RAD51 subnuclear dynamics. *Nucleic Acids Res,* **41**, 943-60.

Trenaman, A. L. (2012). *Trypanosoma brucei BRCA2 in the regulation of genome stability and DNA repair*. PhD PhD, University of Glasgow.

Tu, X. & Wang, C. C. (2004). The involvement of two cdc2-related kinases (CRKs) in Trypanosoma brucei cell cycle regulation and the distinctive stage-specific phenotypes caused by CRK3 depletion. *J Biol Chem,* **279**, 20519-28.

Turner, C. M., Melville, S. E. & Tait, A. (1997). A proposal for karyotype nomenclature in Trypanosoma brucei. *Parasitol Today,* **13**, 5-6.

Ubeda, J. M., Raymond, F., Mukherjee, A., Plourde, M., Gingras, H., Roy, G., Lapointe, A., Leprohon, P., Papadopoulou, B., Corbeil, J. & Ouellette, M. (2014). Genome-wide stochastic adaptive DNA amplification at direct and inverted DNA repeats in the parasite Leishmania. *PLoS Biol,* **12**, e1001868.

Unnikrishnan, A., Gafken, P. R. & Tsukiyama, T. (2010). Dynamic changes in histone acetylation regulate origins of DNA replication. *Nat Struct Mol Biol,* **17**, 430-7.

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

Urbaniak, M. D., Martin, D. M. & Ferguson, M. A. (2013). Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of Trypanosoma brucei. *J Proteome Res,* **12**, 2233-44.

Urwyler, S., Studer, E., Renggli, C. K. & Roditi, I. (2007). A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of Trypanosoma brucei. *Mol Microbiol,* **63**, 218-28.

Urwyler, S., Vassella, E., Van Den Abbeele, J., Renggli, C. K., Blundell, P., Barry, J. D. & Roditi, I. (2005). Expression of procyclin mRNAs during cyclical transmission of Trypanosoma brucei. *PLoS Pathog,* **1**, e22.

van Deursen, F., Sengupta, S., De Piccoli, G., Sanchez-Diaz, A. & Labib, K. (2012). Mcm10 associates with the loaded DNA helicase at replication origins and defines a novel step in its activation. *Embo j,* **31**, 2195-206.

van Leeuwen, F., Wijsman, E. R., Kieft, R., van der Marel, G. A., van Boom, J. H. & Borst, P. (1997). Localization of the modified base J in telomeric VSG gene expression sites of Trypanosoma brucei. *Genes Dev,* **11**, 3232-41.

van Luenen, H. G., Farris, C., Jan, S., Genest, P. A., Tripathi, P., Velds, A., Kerkhoven, R. M., Nieuwland, M., Haydock, A., Ramasamy, G., Vainio, S., Heidebrecht, T., Perrakis, A., Pagie, L., van Steensel, B., Myler, P. J. & Borst, P. (2012). Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. *Cell,* **150**, 909-21.

Vanhamme, L., Berberof, M., Le Ray, D. & Pays, E. (1995). Stimuli of differentiation regulate RNA elongation in the transcription units for the major stage-specific antigens of Trypanosoma brucei. *Nucleic Acids Res,* **23**, 1862-9.

Vassella, E., Acosta-Serrano, A., Studer, E., Lee, S. H., Englund, P. T. & Roditi, I. (2001). Multiple procyclin isoforms are expressed differentially during the development of insect forms of Trypanosoma brucei. *J Mol Biol,* **312**, 597-607.

Vassella, E., Den Abbeele, J. V., Butikofer, P., Renggli, C. K., Furger, A., Brun, R. & Roditi, I. (2000). A major surface glycoprotein of trypanosoma brucei is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. *Genes Dev,* **14**, 615-26.

Vassella, E., Reuner, B., Yutzy, B. & Boshart, M. (1997). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J Cell Sci,* **110 (Pt 21)**, 2661-71.

Vaughan, S. & Gull, K. (2008). The structural mechanics of cell division in Trypanosoma brucei. *Biochem Soc Trans,* **36**, 421-4.

Vickerman, K. (1976). *The diversity of the kinetoplastid flagellates.* London: Academic Press.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. *Br Med Bull,* **41**, 105-14.

Vijayraghavan, S. & Schwacha, A. (2012). The eukaryotic Mcm2-7 replicative helicase. *Subcell Biochem,* **62**, 113-34.

Walters, A. D. & Chong, J. P. (2010). An archaeal order with multiple minichromosome maintenance genes. *Microbiology,* **156**, 1405-14.

Wang, B., Feng, L., Hu, Y., Huang, S. H., Reynolds, C. P., Wu, L. & Jong, A. Y. (1999). The essential role of *Saccharomyces cerevisiae* CDC6 nucleotide-binding site in cell growth, DNA synthesis, and Orc1 association. *J Biol Chem,* **274**, 8291-8.

Wang, G., Klein, M. G., Tokonzaba, E., Zhang, Y., Holden, L. G. & Chen, X. S. (2008). The structure of a DnaB-family replicative helicase and its interactions with primase. *Nat Struct Mol Biol,* **15**, 94-100.

Wang, Z. & Englund, P. T. (2001). RNA interference of a trypanosome topoisomerase II causes progressive loss of mitochondrial DNA. *Embo j,* **20**, 4674-83.

Wang, Z., Morris, J. C., Drew, M. E. & Englund, P. T. (2000). Inhibition of Trypanosoma brucei gene expression by RNA interference using an integratable vector with opposing T7 promoters. *J Biol Chem,* **275**, 40174-9.

Wardleworth, B. N., Russell, R. J., Bell, S. D., Taylor, G. L. & White, M. F. (2002). Structure of Alba: an archaeal chromatin protein modulated by acetylation. *Embo j,* **21**, 4654-62.

Watanabe, Y., Fujiyama, A., Ichiba, Y., Hattori, M., Yada, T., Sakaki, Y. & Ikemura, T. (2002). Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions. *Hum Mol Genet,* **11**, 13-21.

Weatherly, D. B., Boehlke, C. & Tarleton, R. L. (2009). Chromosome level assembly of the hybrid Trypanosoma cruzi genome. *BMC Genomics,* **10**, 255.

Wei, Z., Liu, C., Wu, X., Xu, N., Zhou, B., Liang, C. & Zhu, G. (2010). Characterization and structure determination of the Cdt1 binding domain of human minichromosome maintenance (Mcm) 6. *J Biol Chem,* **285**, 12469-73.

Werner, F. & Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol,* **9**, 85-98.

Wheeler, R. J., Gluenz, E. & Gull, K. (2011). The cell cycle of Leishmania: morphogenetic events and their implications for parasite biology. *Mol Microbiol,* **79**, 647-62.

Wheeler, R. J., Gluenz, E. & Gull, K. (2013a). The limits on trypanosomatid morphological diversity. *PLoS One,* **8**, e79581.

Wheeler, R. J., Scheumann, N., Wickstead, B., Gull, K. & Vaughan, S. (2013b). Cytokinesis in Trypanosoma brucei differs between bloodstream and tsetse trypomastigote forms: implications for microtubule-based morphogenesis and mutant analysis. *Mol Microbiol,* **90**, 1339-55.

WHO (2012). *Accelarating work to overcome the global impact of negleted tropical diseases: a roadmap for implementation.* Geneva, Switzerland: World Health Organization.

Wickstead, B., Ersfeld, K. & Gull, K. (2004). The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. *Genome Res,* **14**, 1014-24.

Wigley, D. B. (2009). ORC proteins: marking the start. *Curr Opin Struct Biol,* **19**, 72-8.

Wirtz, E., Leal, S., Ochatt, C. & Cross, G. A. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. *Mol Biochem Parasitol,* **99**, 89-101.

Wohlschlegel, J. A., Dwyer, B. T., Dhar, S. K., Cvetic, C., Walter, J. C. & Dutta, A. (2000). Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. *Science,* **290**, 2309-12.

Woodward, A. M., Gohler, T., Luciani, M. G., Oehlmann, M., Ge, X., Gartner, A., Jackson, D. A. & Blow, J. J. (2006). Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. *J Cell Biol,* **173**, 673-83.

Woodward, R. & 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)**, 49-57.

Wright, J. R., Siegel, T. N. & Cross, G. A. (2010). Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in Trypanosoma brucei. *Mol Biochem Parasitol,* **172**, 141-4.

Yamazaki, S., Hayano, M. & Masai, H. (2013). Replication timing regulation of eukaryotic replicons: Rif1 as a global regulator of replication timing. *Trends Genet,* **29**, 449-60.

Yanagi, K., Mizuno, T., You, Z. & Hanaoka, F. (2002). Mouse geminin inhibits not only Cdt1-MCM6 interactions but also a novel intrinsic Cdt1 DNA binding activity. *J Biol Chem,* **277**, 40871-80.

Yardimci, H., Loveland, A. B., Habuchi, S., van Oijen, A. M. & Walter, J. C. (2010). Uncoupling of sister replisomes during eukaryotic DNA replication. *Mol Cell,* **40**, 834-40.

Yeeles, J. T., Deegan, T. D., Janska, A., Early, A. & Diffley, J. F. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature*.

Yu, Z., Feng, D. & Liang, C. (2004). Pairwise interactions of the six human MCM protein subunits. *J Mol Biol,* **340**, 1197-206.

Yuan, G. C., Liu, Y. J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J. & Rando, O. J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. *Science,* **309**, 626-30.

Yuan, H., Liu, X. P., Han, Z., Allers, T., Hou, J. L. & Liu, J. H. (2013). RecJ-like protein from Pyrococcus furiosus has 3'-5' exonuclease activity on RNA: implications for proofreading of 3'-mismatched RNA primers in DNA replication. *Nucleic Acids Res,* **41**, 5817-26.

Zawilak-Pawlik, A., Kois, A., Majka, J., Jakimowicz, D., Smulczyk-Krawczyszyn, A., Messer, W. & Zakrzewska-Czerwinska, J. (2005). Architecture of bacterial replication initiation complexes: orisomes from four unrelated bacteria. *Biochem J,* **389**, 471-81.

Zegerman, P. & Diffley, J. F. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature,* **445**, 281-5.

Zeiner, G. M., Sturm, N. R. & Campbell, D. A. (2003). Exportin 1 mediates nuclear export of the kinetoplastid spliced leader RNA. *Eukaryot Cell,* **2**, 222-30.

Zhang, J., Yu, L., Wu, X., Zou, L., Sou, K. K., Wei, Z., Cheng, X., Zhu, G. & Liang, C. (2010). The interacting domains of hCdt1 and hMcm6 involved in the chromatin loading of the MCM complex in human cells. *Cell Cycle,* **9**, 4848-57.

Zhang, Y. (2008a). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics,* **9**, 40.

Zhang, Y. (2008b). Progress and challenges in protein structure prediction. *Curr Opin Struct Biol,* **18**, 342-8.

Zhao, H., Halicka, H. D., Li, J., Biela, E., Berniak, K., Dobrucki, J. & Darzynkiewicz, Z. (2013). DNA damage signaling, impairment of cell cycle progression, and apoptosis triggered by 5-ethynyl-2'-deoxyuridine incorporated into DNA. *Cytometry A,* **83**, 979-88.

Zisimopoulou, P., Staib, C., Nanda, I., Schmid, M. & Grummt, F. (1998). Mouse homolog of the yeast origin recognition complex subunit ORC1 and chromosomal localization of the cognate mouse gene Orc1. *Mol Gen Genet,* **260**, 295-9.