Analysis of the architecture and function of the nuclear DNA replication apparatus in *Trypanosoma brucei*

Calvin Tiengwe
BSc, MSc, MRes

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

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

October, 2010

Abstract

DNA replication is central to the propagation of life and initiates by the designation of genome sequences as origins, where synthesis of a copy of the genetic material begins once per cell division. Despite considerable progress in understanding mitochondrial (kinetoplast) DNA replication in kinetoplastid parasites, little is known about nuclear DNA replication. The mechanism and machinery of DNA replication initiation is well-conserved among characterised eukaryotes. The six protein origin recognition complex (ORC, Orc1-Orc6), Cdc6, and Cdt1 are recruited sequentially to DNA, and once bound, they load the replicative helicase (MCM, a heterohexamer; subunits Mcm2-7) to form a pre-replicative complex (pre-RC) at potential origins of replication. The largest subunit of ORC, Orc1, is related in sequence to Cdc6, indicative of derivation from a common ancestor. Such an ancestral molecule appears still to function in archaea. These prokaryotes lack Cdc6 and possess a protein named Orc1/Cdc6, which appears to provide all ORC functions, since orthologues of Orcs2-6 are absent. In addition to this, archaeal orthologues of Cdt1 have not been clearly described, though potentially related factor, named WhiP (winged helix initiator protein), has been found. Comparative genome analysis of *Trypanosoma brucei* and related trypanosomatids (*Leishmania major* and *Trypanosoma cruzi*) revealed, remarkably, only a single ORC protein that is equally related to eukaryotic Orc1 and Cdc6 (named here TbORC1/CDC6). In addition, no clear homologue of Cdt1 was found. These observations have been interpreted as suggesting that origin designation in trypanosomatids, although eukaryotic, may be archaeal-like, raising numerous mechanistic and evolutionary questions.

To test this hypothesis, and to dissect the process of nuclear DNA replication, a number of experiments are described in this thesis. We used RNA interference (RNAi) to demonstrate that knockdown of *TbORC1/CDC6* in procyclic form (PCF) *T. brucei* cells inhibits nuclear DNA synthesis, as revealed by cell cycle analysis and a BrdU incorporation assay.

Immunofluorescence and GFP-tagging showed that in procyclic form (PCF) cells TbORC1/CDC6 is a nuclear protein. In PCF cells, based on the evidence gathered, we confirm that TbORC1/CDC6 acts in nuclear DNA synthesis. In contrast, RNAi knockdown of *TbORC1/CDC6* in bloodstream form (BSF) *T. brucei* cells resulted in the rapid accumulation of cells with more than two nuclei and two kinetoplasts, indicating a deregulation of the cell cycle, which is then followed by cell rapid cell death. This RNAi result provides greater evidence that TbORC1/CDC6 provides an essential function in the parasite, since RNAi depletion of TbORC1/CDC6 in PCF cells has a less pronounced effect on growth. Nevertheless, attempts to generate TbORC1/CDC6 null mutants failed in PCF cells, consistent with an essential role in this life cycle stage also.

To study the molecular interactors of TbORC1/CDC6, we performed immunoprecipitation analyses. From this, we have identified one protein (gene ID, Tb927.10.13380) that acts as a component of the *T. brucei* pre-replicative machinery, and suggest that this is a previously unidentified orthologue of Orc4. We also indentified a further protein (gene ID, Tb927.10.7980) that may also act in *T. brucei* DNA replication, but whose identity and function are unclear. TbORC1/CDC6 appears not to interact directly with the TbMCM helicase (for which orthologues of all subunits can be identified), consistent with previous observations from a number of eukaryotic organisms, and contrary to reports in some archaeal species. MCM subunits in *T. brucei* form at least one subcomplex (TbMCM2/4/6/7) homologous to that previously observed for human, yeast, *Drosophila, Xenopus* and mouse MCM proteins. Taken together, these data appears to refute the hypothesis that the DNA replication pre-RC machinery in *T. brucei* is analogous to archaea. Rather, we propose that TbORC contains at least two components, TbORC1/CDC6 and Tb927.10.13380, more analogous to the eukaryotic model, suggesting that origin designation is not carried out by a single protein.

To identify potential replication origin sequences, we performed chromatin immunoprecipitation with functional, epitope-tagged TbORC1/CDC6 in PCF

cells and, using a high-resolution tiling array (NimbleGen) for *T. brucei,* we have mapped TbORC1/CDC6 binding sites along all the megabase chromosomes in the genome. Analyses of chromosomes 1-10 showed that 278 binding sites are sparsely located within the core of chromosomes, of which 114 loci (40%) co-localise with probable RNA Polymerase II transcription start sites, perhaps consistent with an origin function. In addition, a further 330 binding sites are present as high density clusters in subtelomeric *VSG* arrays, and 81 binding sites are associated with sub-telomeric elements, perhaps consistent with a non-origin function. Consistent with these results, RNAi knockdown of *TbORC1/CDC6* led to derepression of metacyclic *Variant Surface Glycoprotein* (*VSG*) genes, suggesting that TbORC1/CDC6 plays a role in the epigenetic silencing at *VSG* expression sites in PCF *T. brucei*. Similar analysis of *VSG* expression in BSF cells, and of BSF *VSG*s in PCF cells, was less conclusive, perhaps suggesting differential functions of TbORC1/CDC6 in different life cycle stages or at different *VSG* expression sites. These analyses shed new light on the architecture and potential function of TbORC1/CDC6 in *T. brucei* nuclear DNA replication in general, as well as a potential association between replication and antigenic variation in *T. brucei*.

Table of Contents

List of Tables

List of Figures

List of Appendices

"*Nullius in verba*"

"On the authority of no one"

Motto of the Royal Society

Acknowledgements

I would like to express my sincere appreciation to my supervisors Dr. Richard McCulloch and Prof. Dave Barry for proposing an exciting topic for a PhD, unrelenting guidance and encouraging remarks at every stage during this project.

Special thanks also go to Dr. Michal Swiderski, Dr. Chris Stockdale and Dr. Subhash Verma for the initial lab training and suggestions which have been very significant towards the success of this project. I will never forget all the discussions with Michal from which brilliant ideas were always hatched, as well his constant challenges which were a source of inspiration. I am also grateful to Dr. Lucio Marcello who contributed greatly in the analysis of the microarray data, and all the discussions which drove the project forward.

The friendly working environment from all staff at the Wellcome Trust Centre for Molecular Parasitology will always remain in my mind. I would not have had this life changing opportunity without a recommendation from Dr. Olwyn Byron, my external examiner at the MSc. Course in Nottingham and I say many thanks to her.

Finally, I am greatly indebted to the Wellcome Trust Fund, the Wellcome Four Year Programme at Glasgow University and the Directors of the programme (Prof. Bill Cushley, Prof. Darren Monckton, and Dr. Olwyn Byron for their guidance throughout the course. I am indebted to the Scottish Overseas Research Students Awards Scheme for part sponsorship.

I am also indebted to my entire family for their moral support and to my sister Delphine Tiengwe for all her support in Glasgow, especially throughout the write-up period.

Author's declaration

I declare that this thesis and the results presented within are entirely my own work except where otherwise stated. No part of this thesis has been previously submitted for a degree at any university.

_____

Calvin Tiengwe

List of Abbreviations

ACS	ARS consensus sequence
ADP	adenosine diphosphate
ARS	Autonomous replicating sequence
ATP	adenosine triphosphate
BES	bloodstream VSG expression site
BLE	bleomycin resistance protein gene
BrdU	5'-bromodeoxyuridine
BrdUTP	5'-bromodeoxyuridine Triphosphate
BSA	bovine serum albumin
BSD	blasticidin–S-deaminase gene
BSF	bloodstream form
BVSG	bloodstream variant surface glycoprotein
Cdc6	cell division cycle 6 protein
CDS	coding sequence
ChIP	chromatin immunoprecipitation
cDNA	complementary DNA
Cdt1	Cdc10-dependent transcript 1
co-IP	co-immunoprecipitation
DAPI	4', 6-diamidino-2-phenylindole
dH20	distilled water
DIC	differential interference contrast
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dsDNA	double-stranded DNA
DTT	1, 4-dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ESAG	expression site associated gene
ESB	expression site body
FITC	fluorescein isothiocyanate
gDNA	genomic DNA
GPI	glycosylphosphatidylinositol
HML	Hidden MAT Right
HMR	Hidden MAT Left
HSV-1	herpes simplex virus-type 1
HYG	hygromycin phosphotransferase gene
kb	kilobase-pairs

kDa	kilo-dalton
kDNA	kinetoplast DNA
LB	Luria-Bertani
Mb	megabase-pairs
MCM	minichromosome maintenance
MES	metacyclic VSG expression site
mRNA	messenger RNA
MVSG	metacyclic variant surface glycoprotein
NEO	neomycin phosphotransferase gene
ORC	origin recognition complex
ORF	open reading frame
OriC	origin of chromosomal replication
PARP	procyclic acidic repetitive protein
PBS	phosphate buffered saline
PCF	procyclic form
PCR	polymerase chain reaction
Pol I	RNA polymerase I large subunit
pol II	RNA polymerase II
PUR	puromycin-N-acetyltransferase gene
RNA	ribonucleic acid
RNAi	RNA interference
RT	reverse transcriptase
RT-PCR	reverse transcription polymerase chain reaction
SCC	sodium citrate with sodium chloride
SDS	sodium dodecyl sulphate
ssDNA	single-stranded DNA
SSR	strand switch region
T7RNAP	T7 RNA polymerase
TAE	Tris-acetate-EDTA
TBE	Tris / borate / EDTA buffer
TE	Tris-EDTA
Tet	Tetracycline
TetR	Tet repressor
TREU	Trypanosomiasis Research Edinburgh University
UV	ultraviolet
VSG	variant surface glycoprotein

1 General Introduction

1.1 Introduction to *Trypanosoma brucei*

1.1.1 *Phylogeny of Trypanosoma brucei*

Taxonomically, the species *Trypanosoma brucei* belongs to the eukaryotic supergroup *Excavata,* most of which are unicellular, flagellated organisms (Adl *et al*. 2005; Simpson & Roger 2004). The Excavate supergroup includes many other parasitic protozoan genera: *Giardia, Trichomonas, Leishmania* and *Trypansonoma* - Figure 1-1 (Dacks, Walker, & Field 2008). This new level classification of these eukaryotic organisms is based on a combination of molecular sequence data and morphological traits (Adl *et al*. 2005; Simpson & Roger 2004) and is based on what has been termed an "egalitarian" view of eukaryotic evolution - Figure 1-1 (Dacks *et al*. 2008). It suggests that parasitic protozoans such as *Giardia, Trichomonas, Leishmania* and *Trypansonoma* are not "early-diverging" eukaryotes, as they are frequently described (Dacks & Doolittle 2001; Sogin *et al*. 1989), but rather share a common ancestor and have specialised and adapted to their various niches (Dacks *et al*. 2008). Many studies consider there to be the five eukaryotic supergroups described in Figure 1-1, while others argue that that the Excavata should be considered as two distinct groups, perhaps without a common origin: the Jakobids, Euglenozoans (including *T. brucei*) and Heteroloboseans; and the Parabasalids (including *Trichomonas vaginalis*), the Oxymonads and the Diplomonads (including *Giardia lamblia*) (Fritz-Laylin *et al*. 2010). The sub-taxomic ranks for *T. brucei* are thus: phylum, *Euglenozoa*; order, *Kinetoplastida*; family, *Trypanosomatidae*; and Genus, *Trypanosoma*.

A characteristic feature of members of the order *Kinetoplastida* is the presence of a mitochondrial-like organelle known as the kinetoplast, within which is a concatenated DNA network that assumes the role of more typical mitochondrial DNA present in higher eukaryotes (Liu *et al*. 2005; Maslov, Podtipaev, & Lukes 2001). The family *Trypanosomatidae* has three closely related species known as the TriTryps: *Trypanosoma cruzi, Leishmania major* and *Trypanosoma brucei*. The genomes of each of these organisms, plus some further species, have been sequenced (Ivens *et al*. 2005; Berriman *et al*. 2005; El Sayed *et al*. 2005a) and show a high degree of gene conservation and gene synteny (Ghedin *et al*. 2004).

The trypanosomatids (derived from *Trypansomatidae*) are responsible for a variety of diseases of both human and veterinary importance. These diseases include Chagas' disease (caused by *T. cruzi*), Leishmaniasis (caused by several species of the genus *Leishmania*), Human African Trypanosomiasis (HAT) in humans and nagana in domestic livestock (both caused by several species of the genus *Trypanosoma*)(Garcia *et al*. 2006; Barrett *et al*. 2003). The most important *Trypanosoma* species for humans is *T. brucei*. Two sub-species, *T. brucei rhodesiense* and *T. brucei gambiense,* infect humans in central and western Africa, and in eastern and southern Africa, respectively. Another subspecies, *T. brucei brucei,* is not human infective but infects livestock, causing huge economic misery. HAT disease is invariably fatal if untreated and is a menace to over 60 million people inhabiting over 36 sub-Saharan African countries (Fevre *et al*. 2006; Garcia *et al*. 2006; Barrett 1999).

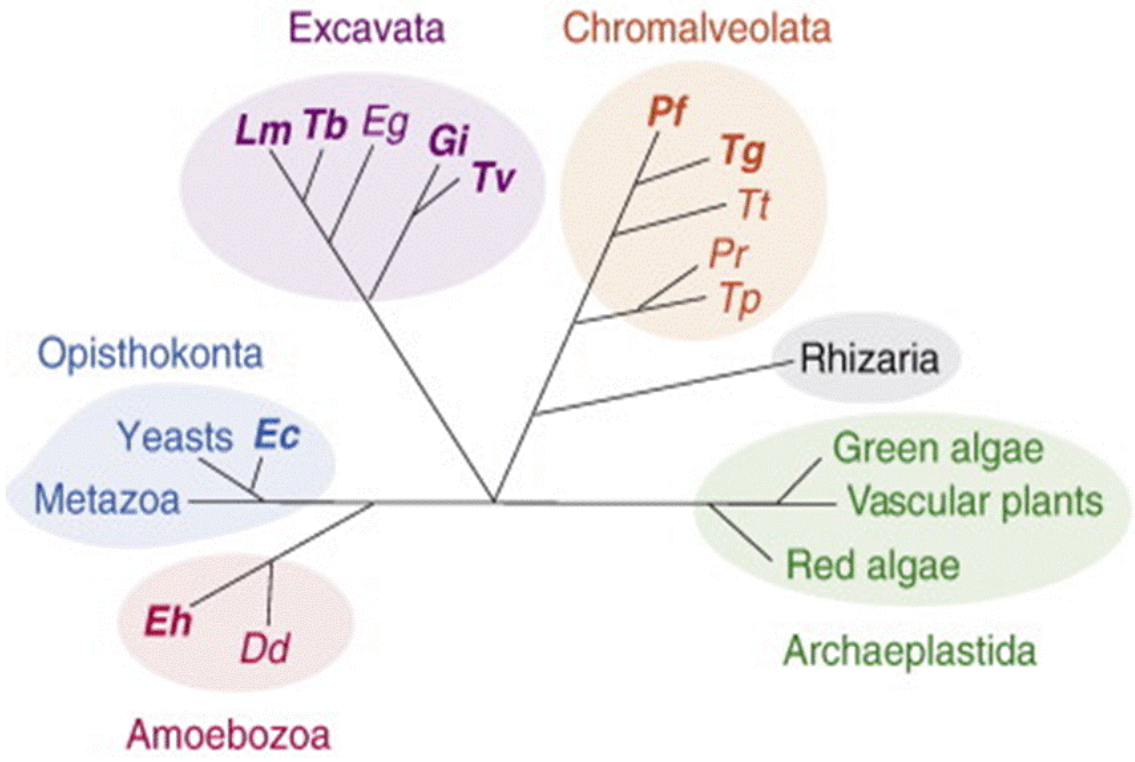


Figure 1-1 - The egalitarian view of eukaryotic phylogeny

A phylogenetic tree showing the classification of eukaryotes into six supergroups: Amoebozoa, Opisthokonta, Excavata, Chromalveolata, Rhizaria, Archaeplastida. The tree, which is not rooted, is generated according the new level of classification by (Adl *et al*. 2005; Simpson & Roger 2004); Lm = *Leishmania major*, Tb = *Trypanosoma brucei*, Gi = *Giardia intestinalis*, Tv = *Trichomonas vaginalis*, Eh = *Entamoeba histolytica*, and Dd = *Dictyostelium discoideum*, Ec = *Encephalitozoon cuniculi*, Pf = *Plasmodium falciparum*, Tt = *Tetrahymena thermophila,* Tp = *Thalassiosira pseudonana. T. brucei* belongs to the supergroup Excavata. Taken from (Dacks *et al*. 2008)

1.1.2 Trypanosomiasis and disease control

The two human infective *Trypanosoma* subspecies: *T. b. rhodesiense* and *T. b. gambiense,* cause different variants of HAT disease or African sleeping sickness. While *T. b. rhodesiense* causes an acute infection, *T. b. gambiense* causes a more chronic infection (Barrett *et al.* 2003). In acute infections, the parasites reside extracellularly in the lymphatic and vascular systems of the mammalian host and are more readily detected in blood, lymph, or tissue aspirates (Barrett *et al.* 2003). However, the infection is characterised by a rapid transition to the central nervous system and death occurs within weeks or months (Brun *et al*. 2010). In chronic infections which can last around three years (Brun *et al.* 2010), the parasites cross the blood brain barrier and reside in the central nervous system where they cause motor and sensory disorders leading to alterations in sleep/wake patterns and seizures; hence the term sleeping sickness (Barrett *et al.* 2003). If untreated, HAT disease is invariably fatal.

African sleeping sickness is among the most neglected diseases in the world. Because it affects the world's poorest countries there is a lack of incentive from major pharmaceutical companies for investment in the development of new drugs (Barrett *et al*. 2007; Maudlin 2006; Ehrenberg & Ault 2005). It is a major public health problem and is a strong impediment to the socio-economic development of affected countries. Current control measures for trypanosome infections rely primarily on prevention of transmission and chemotherapy (Barrett *et al.* 2003). However, this is hindered by an increasing problem of drug resistance, slow progress in prevention approaches (such as vector and reservoir host controls), the lack of a preventive vaccine, limited access to essential medicines, lack of surveillance, and lack of clinical and diagnostic expertise especially in regions affected by conflict (Hotez *et al*. 2007; Ehrenberg & Ault 2005). Other than being limited in supply, current drugs also suffer drawbacks of high levels of toxicity and an increasing incidence of treatment failures (Barrett *et al.* 2007), and hence there is a necessity for the identification and characterisation of new drug targets.

Trypanosomatids display considerable divergence, both structurally and functionally, in basic biological processes. This may bring hope that novel

therapies can be developed that exploit these processes, as they show differences between the parasites and their hosts. Examples of such divergence in trypanosomatids include the kDNA structure, RNA editing in the mitochondrion, and nuclear polygenic transcription (see Section 4.1.1). Another biological process that is central and essential to the propagation of life is DNA replication. With the *T. brucei* genome sequence now available (Berriman *et al.* 2005), an understanding of DNA replication and its associated characteristics can provide insights into both fundamentally conserved and fundamentally different aspects of this process. Indeed, unique aspects of the *T. brucei* DNA replication machinery may differ sufficiently from its host (see Section 4.1.1) to represent targets for future development of novel chemotherapeutic agents.

1.1.3 Life Cycle of Trypanosoma brucei

T. brucei has a complex digenetic life cycle (Figure 1-2), being transmitted between mammalian hosts by an insect vector (*Glossina* spp; commonly known as the tsetse fly)(Gull 2002). When the mammal-infective form of the parasite, known as the metacyclic trypomastigote, is introduced into its mammalian host by the bite of an infected tsetse fly, they multiply extravascularly at the site of the bite, and establish the infection. From there they migrate into the bloodstream and tissue fluids, and eventually find their way into the cerebrospinal fluid in chronic infections (Tetley & Vickerman 1985). In the bloodstream, trypomastigotes differentiate from the non-dividing metacyclic form cells into the proliferative long slender form. Both these life cycle forms express a variant surface glycoprotein (VSG) coat on the cell surface, which acts as 'shield' to prevent the immune system from recognising invariant surface antigens. Changes in the VSG coat in the longer slender cells provide for immune evasion, a process termed antigenic variation. Throughout infection in the bloodstream, the long-slender form cells differentiate into the non-dividing short 'stumpy' form (Tetley *et al*. 1987). This latter form is capable of infecting the tsetse fly. During a blood meal, short 'stumpy' trypomastigotes are ingested into the fly gut, where they differentiate into replicative procyclic trypomastigotes and then into the epimastigote form in the salivary glands. During the first of these differentiation processes the major outer membrane protein of the parasite changes from VSG to a distinct glycoprotein coat,

composed of procyclic acidic repetitive proteins (PARPs) or procyclins (Barry & McCulloch 2001; Roditi *et al*. 1998). In the salivary gland of the tsetse fly the replicative epimastigote cells differentiate into the mammal-infective metacyclic form cells, which are not attached to the gland wall and are capable of re-infecting during a blood meal, thereby closing the cycle (Barry *et al*. 1998). During this differentiation, the surface coat is changed back to VSG from procyclin, in preparation for confronting the mammal's immune system.

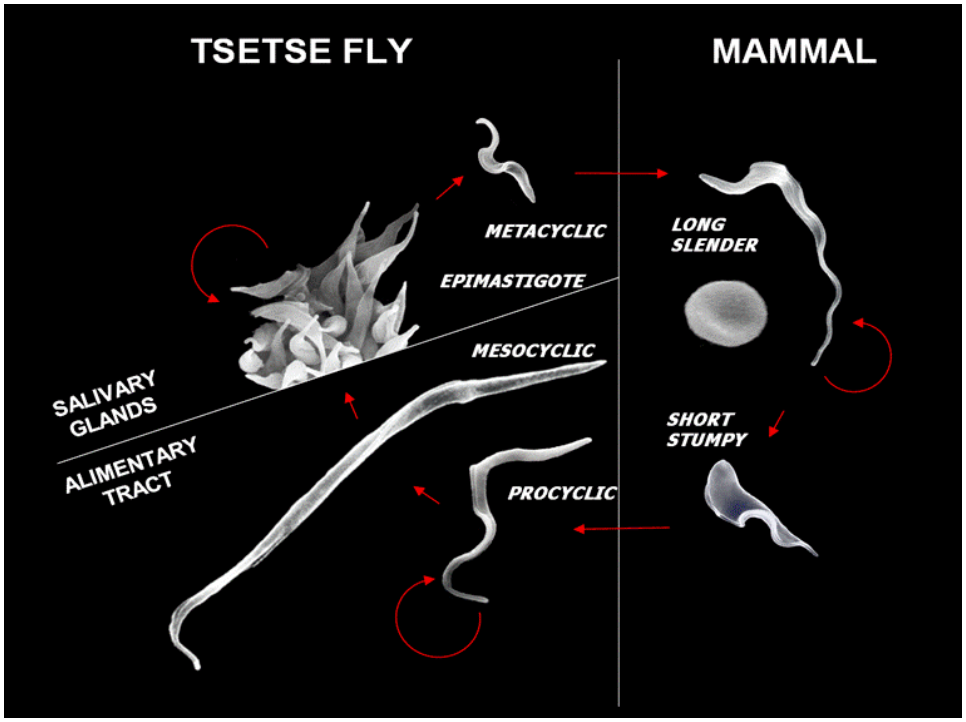


Figure 1-2 – The Life cycle of *Trypanosoma brucei*

T. brucei proliferative and non-proliferative stages in mammalian and tsetse fly hosts are shown as scanning electron micrographs; shown to scale. Circular arrows represent replicative stages, whereas straight arrows represent differentiation and progression through the life cycle. Taken from (Barry & McCulloch 2001)

1.2 Cellular and genome organisation of *Trypanosoma brucei*

1.2.1 Trypanosoma brucei cell structure

T. brucei has an elongated shape within which are single copy organelles (one kinetoplast, one Golgi, and one nucleus) precisely positioned between the posterior end and the centre of the cell (Figure 1-3). At the posterior end of the cell is an invagination of the plasma membrane known as the flagellar pocket,

since this is where the flagellum emerges (Matthews 2005). The single mitochondrion that *T. brucei* possesses runs the length of the cell and contains a unique matrix of concatenated DNA, termed kinetoplast DNA (kDNA) (Shapiro & Englund 1995). The kDNA consists of an interlocked network of large and small DNA rings known as maxicircles and minicircles, respectively. The kDNA maxicircles range in size from 20–40 kb, encode ribosomal RNA and several mitochondrial proteins and are present in tens of copies. The minicircles range in size from 0.5–2.9 kb, encode guide RNAs that regulate the RNA editing machinery for the formation of functional maxicircle mRNA, and are present in thousands of copies (Morris *et al*. 2001; Shapiro & Englund 1995). kDNA replication and segregation is independent of nuclear DNA replication and mitosis, although the events are coordinated (Woodward & Gull 1990).

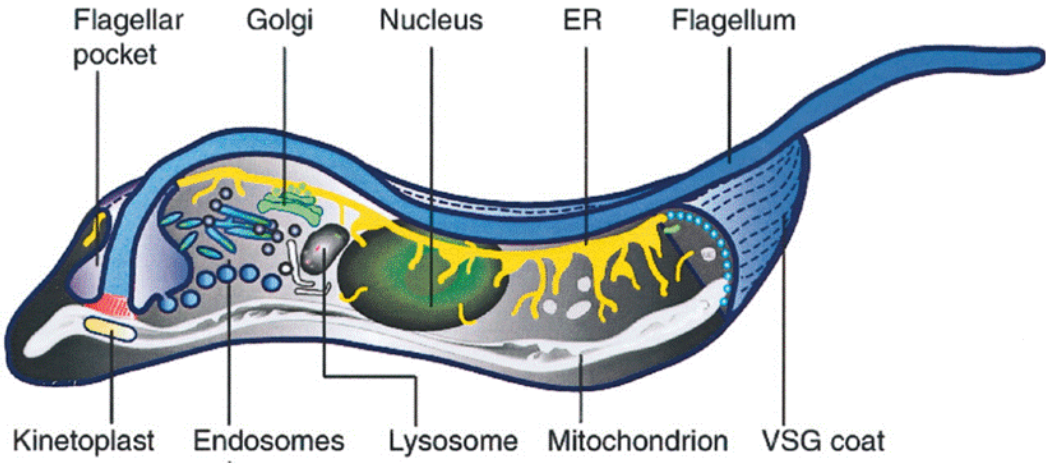


Figure 1-3 - Cartoon of a *T. brucei* cell showing single copy organelles
Image taken from (Grunfelder *et al*. 2003)

1.2.2 The nuclear genome structure

The nuclear genome comprises 11 diploid megabase-sized chromosomes (0.9-6 megabase in length), several intermediate-sized chromosomes (~200-900 kb in size) and approximately 100 mini-chromosomes (~50-150 kb in size) (El Sayed *et al*. 2000; Melville *et al*. 1998). The near-completed genome sequence of *T. brucei* strain TREU (Trypanosomiasis Research Edinburgh University) strain 927/4 has been published (Berriman *et al*. 2005). Sequence analysis of the *T. brucei* megabase nuclear chromosomes revealed a haploid DNA content of ~26-

megabases, containing 9068 predicted genes. The genetic repertoire includes ~900 pseudogenes, primarily derived from variable surface glycoprotein genes (*VSGs*), and ~1700 further *T. brucei*-specific genes (Berriman *et al.* 2005). These figures are exclusive of some of the subtelomeric ends of the megabase chromosomes, the intermediate-sized chromosomes and the minichromosomes, whose genetic content are yet to be published.

In general, a *T. brucei* megabase chromosome can be sub-structured into three main parts: a central core that harbours most of the 'house-keeping' genes, a proximal subtelomeric region that harbours tandem repeated units of *VSG* arrays and a distal subtelomeric region that harbours *VSG* expression sites (Hertz-Fowler, Renauld, & Berriman 2007). The house-keeping genes within the central core have a highly unconventional structural organisation. The genes are arranged in closely packed clusters in which the ORFs face the same direction and are transcribed by RNA Polymerase II into, it appears, a single pre-mRNA from which individual mRNAs are generated by coupled trans-splicing and polyadenylation (Martinez-Calvillo *et al*. 2010; Clayton 2002). These so-called directional gene clusters are reminiscent of bacterial operons, although the co-transcribed genes are not normally functionally related. Adjacent directional gene clusters can be arranged head to head (convergent) or tail to tail (divergent), and are separated by gene-free regions called strand switch regions (SSRs) (Hall *et al*. 2003; El Sayed *et al*. 2003; Myler *et al*. 1999). Available evidence suggests that: (a) no consensus motifs exist for sequences present at SSRs (Hall *et al.* 2003), and (b) transcription initiation occurs bi-directionally in divergent strand switch regions. As a result, divergent SSRs represent transcription start sites in which no promoters have been described, though they are sites of accumulation of histone variants (H2AZ, H2BV, and H4K10ac). Transcription termination is thought to occur at convergent SSRs (Siegel *et al*. 2009). In *T. cruzi,* as well as in *T. brucei,* some SSRs have repetitive DNA sequences, which have been shown to act as centromeres functioning to provide mitotic stability to chromosomes (Obado *et al*. 2007; Obado *et al*. 2005).

Due to their repetitive nature, the subtelomeric ends of some megabase chromosomes of the genome strain are still being assembled, and when this information becomes available, more light will be shed on their organisation (Renauld, Kelly, & Horn 2007). However, available information suggest that 55 %

of the size of *T. brucei* megabase chromosomes is devoted to subtelomeres, harbouring 10 % of the total predicted gene repertoire (Hertz-Fowler *et al.* 2007). The majority of this subtelomeric repertiore are *VSG* and *VSG*-associated genes, involved in evasion of host immune destruction by antigenic variation (Marcello & Barry 2007; Barry & McCulloch 2001). The minichromosomes also harbour a repertoire of *VSGs*, though these all appear to be directly telomeric. Minichromosomes have a characteristic central core region of 177-bp repetitive sequences arranged in large palindromes (Wickstead, Ersfeld, & Gull 2004) and this is shared with intermediate chromosomes. *VSGs* in the minichromosomes and in the subtelomeric *VSG* arrays in the megabase chromosomes are not expressed *in-situ*. Instead, they can only become transcribed if they are translocated by recombination to *VSG* expression sites (see Section 1.3) in the megabase and intermediate chromosomes (Taylor & Rudenko 2006; Barry & McCulloch 2001).

1.3 Antigenic variation in *T. brucei*

The biological relationship of host-pathogen interactions has intrigued scientists over the years in trying to unravel the evasion mechanisms pathogens utilise to overcome host innate and/or adaptive immunity. From simple prokaryotes to more complex single-celled eukaryotes, different strategies have been adopted by pathogens to ensure survival within a host organism. For example, Gram-positive bacteria such as *Streptococcus pneumoniae* have a thick peptidoglycan cell wall layer, which prevents complement-mediated lysis by the membrane-attack complex (Joiner *et al*. 1983). Alternatively, Gram-negative bacteria, such as members of the genus *Salmonella,* employ a mechanism known as phase variation, where they randomly switch phenotypes by varying the expression of specific antigenic factors either in an “on-state” or in an “off-state” (Tamura *et al*. 1988). Another mechanism by which pathogens such as viruses evade host immune attack is by a process known as molecular mimicry, whereby a protein is expressed that closely mimics the structure and function of a host protein (Zhao *et al*. 1998). An example is an epitope expressed on the coat protein of the herpes simplex virus-type 1 (HSV-1) KOS strain, which is analogous to corneal antigens expressed in a murine mouse model (Zhao *et al.* 1998).

Unicellular parasites such as *T. cruzi* and *Leishmania,* which are close relatives of *T. brucei,* escape host-immune surveillance by inhabiting intracellular environments, and hence become inaccessible (Sibley 2004). On the other hand, *T. brucei* proliferates extracellularly in the blood and tissue fluids of a mammalian host, where it is subject to continuous immune attack from host defence molecules such as antibodies. To thrive successfully in these environments and to subsequently establish a chronic infection the parasites utilise a protective mechanism known as antigenic variation (Pays, Vanhamme, & Perez-Morga 2004; Barry & McCulloch 2001). Antigenic variation involves continuously altering exposed surface antigens, thereby displaying to the host immunologically distinct antigens, meaning that the adaptive immune response cannot eradicate the infecting population (Morrison, Marcello, & McCulloch 2009). In *T. brucei*, the variant antigen is VSG, which forms a protective 'coat' consisting of densely packed molecules (~5 x 10^6 dimers per cell) that protects invariant surface molecules from immune recognition. VSG is anchored to the cell membrane through a glycosylphosphatidylinositol (GPI) anchor (Masterson *et al*. 1989). The ability to continuously maintain the process of antigenic variation depends on the huge repertoire (~1500 *VSG*s) of silent *VSG* genes present in the subtelomeres of all *T. brucei* chromosomes (Marcello & Barry 2007; Berriman *et al*. 2005). At any one time, a trypanosome cell can only express a single VSG molecule on its surface, and this expression occurs from specialised sites at the distal subtelomeric ends of megabase or intermediate size chromosomes known as VSG expression sites (Becker *et al*. 2004).

The life cycle of *T. brucei* (see above) involves distinct periods in either an insect host or a mammalian host. Prior to infection of a mammalian host, the expression of VSGs commences in the metacyclic form cells in the insect salivary glands and these VSGs are referred to as metacyclic VSGs (MVSGs), while in the bloodstream form (BSF) the expressed VSGs are referred to as bloodstream VSGs (BVSGs) (Barry & McCulloch 2001). *MVSG* genes are therefore expressed from metacyclic expression sites (MES), while *BVSG* genes are expressed from bloodstream expression sites (BES). The structure and organisation of these *VSG* expression sites are distinct and is discussed in Section 3.8.6. Fourteen distinct BESs have been characterised in the BSF stage of the *T. brucei* strain Lister 427 (Hertz-Fowler *et al*. 2008). Of these, only one can be transcriptionally active at

any one time in a trypanosome cell. Fewer MES have been described in detail, but again only one is actively transcribed at a time (Ginger *et al*. 2002). The process of singular BES or MES transcription is not yet fully understood. Accumulating evidence suggest that epigenetic phenomena may be involved in maintaining transcriptionally silent chromatin at inactive BES, while little is known about MES (Molloy 2010; Rudenko 2010; Stockdale *et al*. 2008). Section 3.8.6 discusses published work on the transcriptional regulation of BES in the context of new evidence we provide of how both BES and MES activation may be regulated.

1.4 The *T. brucei* cell cycle

The mammalian cell division cycle consists of four temporally distinct phases: G_0/G_1, S, G_2, and M. G_0/G_1 is a gap phase marked by synthesis of proteins, nucleotides and nutrients that are required in S phase, where replication of DNA occurs. There is also significant protein synthesis in the G_2 phase, another gap phase during which the cell prepares to enter mitosis (M-phase) or, in specific cases, meiosis. The M-phase is characterised by nuclear and cytoplasmic division. *T. brucei* broadly follows a typical mammalian cell cycle, although it is characterised by unique features and unusual complexities compared with higher eukaryotes (Mckean 2003). For example, whereas in mammalian cells mitochondrial DNA is replicated throughout the cell cycle, kDNA replication in *T. brucei* occurs in S-phase and division of the kDNA network precedes, but overlaps, with nuclear division (Mckean 2003; Woodward & Gull 1990). The *T. brucei* cell cycle has been extensively studied in procyclic form (PCF) cells (Figure 1-4). In this life cycle stage, the cell cycle takes 8.5 hours to complete (compared to ~6 hours in the BSF) and can be followed chronologically by the duplication of specific organelles in the following order: basal body, golgi, endoplasmic reticulum (ER), flagellum, kinetoplast, and nucleus (though, again, kinetoplast division overlaps with nuclear division) (Hammarton, Wickstead, & Mckean 2007). The replicated nucleus then undergoes mitosis, which is followed by the division of the cell into two daughter cells in a process known as cytokinesis, which commences prior to the completion of mitosis (Hammarton *et al*. 2007).

As mentioned above, kDNA and nuclear DNA replication are distinct steps in the cell cycle and, although they overlap, the duration of kDNA replication and segregation (~3 hrs) is slightly shorter than that of nuclear DNA replication and nuclear mitosis (4 hrs) - Figure 1-4 (Hammarton *et al*. 2007). This time difference has been exploited to study the progress of cell cycle events in both PCF and BSF cells in *T. brucei*, as shown in Figure 1-4 (Hammarton *et al*. 2005; Hammarton *et al*. 2003). This is described in detail in Section 3.3.5.

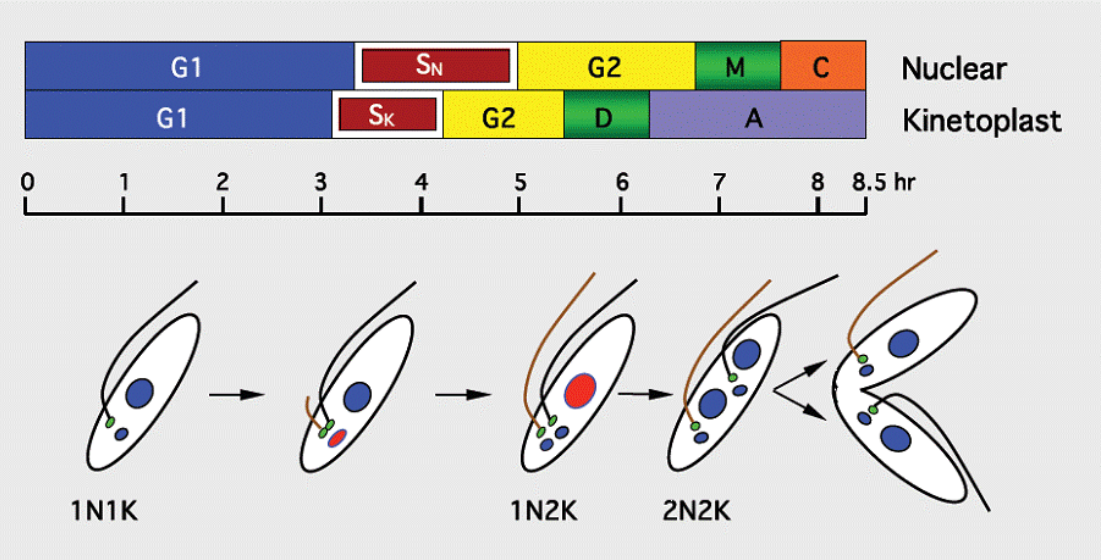


Figure 1-4 – The cell division cycle in procyclic form *T. brucei*

Upper panel: Relative timings of the nuclear and kinetoplast division cycles during the cell cycle phases: G_1 and G_2 represent gap phases, S_N and S_K represent nuclear and kinetoplast DNA synthesis, respectively, M represents Mitosis, C represents cytokinesis, D represents kinetoplast division and A represents kinetoplast segregation. Lower panel: The trypanosome cell is depicted by a white oval; basal body is indicated by a green-shaded circle; flagellum is indicated by black line that emerges from basal body; kinetoplast is indicated by smaller blue-shaded circle; and nucleus is indicated by larger blue-shaded circle. Arrow heads indicate direction of progression of cell cycle; the onset of S_K is indicated by smaller red-shaded oval while the onset of S_N is indicated by red-shaded circle. The organelle configuration of nucleus and kinetoplast at each phase of the cell cycle is denoted below: 1N1K for one nucleus and one kinetoplast, 1N2K for one nucleus and two kinetoplasts, and 2N2K for two nuclei and two kinetoplasts. Adapted from (Hammarton *et al*. 2007)

The molecular machinery of kDNA replication and the mechanisms involved in copying this complex genome have been widely studied in *T. brucei* and are discussed in Section 3.1. Much of the *T. brucei* kDNA replication machinery appears to be substantially distinct from that used in nuclear DNA replication (e.g. specialised DNA polymerases and DNA ligases). In stark contrast to this, and despite the central role of this process, nuclear DNA replication still remains a very poorly studied process in *T. brucei*, or indeed in any trypanosomatid, and is the subject of this PhD project.

1.5 Origins of DNA replication

The blueprint of every cell in an organism is its DNA. The information encoded in DNA must be faithfully duplicated before it is passed from parent(s) to progeny to ensure the propagation of life. Despite the central nature of this process, DNA replication mechanisms show considerable variation in organisation in prokaryotes and eukaryotes (Robinson & Bell 2005). The foundation for our understanding of DNA replication regulation began in 1963 when Jacob, Brenner and Cuzin published a paper "On the regulation of DNA replication in bacteria" and in it proposed the replicon model (Jacob, Brenner, & Cuzin 1963). This model proposed that DNA replication is closely similar to gene transcription, whereby its initiation involves interactions between *trans*-acting regulators, known as initiator factors, and *cis*-acting elements in the genome, known as the replicator sequences (replicons), that determine the initiation of DNA replication (Jacob *et al*. 1963). Over the years this model has proved to be accurate, and much has been learned about the structure, function and organisation of the initiator factors and replicator sequences (elements that determine the origin of DNA replication) in both simple and complex model organisms [for reviews see (Masai *et al*. 2010; Sun & Kong 2010; Mott & Berger 2007; Robinson & Bell 2005)]. Below, we discuss DNA replication sequences (origins) in bacteria, archaea and eukaryotes and their associated initiator factors

1.5.1 Bacterial origins

In the bacterium *Escherichia coli,* DNA replication of the single, circular chromosome initiates at a specific locus known as *Ori*gin of *C*hromosomal replication (*oriC*), and the replication from this site proceeds bidirectionally (Bird *et al*. 1972). By integrating Phage Mu-I into the different chromosomal locations in *E. coli* K-12, establishing a synchronous cycle of replication of the genome, and *in vivo* radialobelling the genomic DNA with radioactive thymidine and then using DNA-DNA hybridisation to determine the rate of duplication of the integrated markers in the genome, Bird *et al* (1972) were able to experimentally map the positions of the origin and terminus of replication in *E. coli* (Bird *et al*. 1972). They reported that the origin was located near the *dnaA-ilv* locus present at 74 min on the genetic map of the chromosome (Bird *et al*.

1972). Over time, biochemical and genetic studies have been carried out to further characterise this locus. Yasuda and Hirota (1977) were able to clone the locus into a plasmid, pSYR211, and showed that this molecule was only able to replicate in the presence of the origin sequence (Yasuda & Hirota 1977). This study refined the location of the origin to 83 min on the genetic map, and later on, a minimum region of 245 bp was shown to be conserved in five enteric bacteria (Zyskind *et al*. 1983).

The origin region (O*riC*) has been partitioned into sub-domains known as DnaA boxes or R-boxes, which promote DNA-initiator protein (DnaA) interaction (Fuller, Funnell, & Kornberg 1984). The exact number of DnaA boxes varies between different bacterial species (Yoshikawa & Ogasawara 1991). In *E. coli,* for example, there are five DnaA boxes (numbered R1-R5; Figure 1-5) that contain a highly conserved consensus 9 bp sequence; 5' - TTATCCACA - 3' (Fuller *et al*. 1984). Thirteen DnaA boxes are found in *Thermus thermophilus,* fifteen in *Bacillus subtilis,* and nineteen in *Streptomyces lividans* (Messer 2002). Interspersed within the DnaA boxes are additional DnaA binding sites termed I-sites (I1, I2, and I3; Figure 1-5); consensus, 5'-AGATCT (McGarry *et al*. 2004; Sekimizu, Bramhill, & Kornberg 1987). Located between DnaA box R1 and R5 is a single binding site for Integration Host Factor (IHF; Figure 1-5), which is necessary for facilitating DnaA-mediated unwinding of *OriC* by introducing a bend into the DNA backbone (Polaczek *et al*. 1997). Immediately adjacent to DnaA box R2 is a binding site for Factor for Inversion Stimulation (Fis), a protein that inhibits DNA melting at *OriC* (Wold, Crooke, & Skarstad 1996). Upstream of DnaA box R1 are three upstream 13-mer AT-rich repeat elements that facilitate melting of duplex DNA [Figure 1-5; (McGarry *et al*. 2004)]. The concerted interactions of DnaA-ATP to the DnaA boxes and I-sites, and of IHF to its binding site, cause the AT-rich sequence upstream to become negatively supercoiled, facilitating unzipping of local DNA (Speck & Messer 2001; Crooke *et al*. 1991). Several other proteins are recruited via the polymerisation of DnaA, thereby initiating DNA replication (Fang, Davey, & O'Donnell 1999) - see Section 1.6.1.

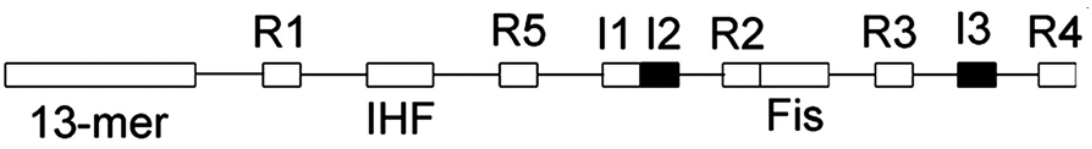


Figure 1-5 – Architecture of the bacterial origin of chromosomal replication (*OriC*)

Schematic representation of DNA sequence elements present at *E. coli OriC* showing five DnaA boxes depicted as by boxes [R1-R5 (M)], interspersed by I-sites depicted by boxes (I1, I2, and I3); a site bound by integration host factor (IHF) is depicted by IHF; a 13-mer duplex unwinding element is depicted by 13-mer; and a site bound by FIs protein is depicted by Fis. Taken from (McGarry *et al*. 2004)

1.5.2 Eukaryotic origins

Unlike most prokaryotes, which have circular chromosomes, in eukaryotes chromosomes are linear, larger in size and replication fork progression is slower. To ensure that the entire genome is duplicated within the time allocated during the cell cycle (termed S-phase), initiation of DNA replication requires multiple origins of replication on each chromosome (Kelly & Stillman 2006). In the well-studied single-celled eukaryote, *Saccharomyces cerevisiae*, initiation of DNA replication involves the recognition (by the Origin Recognition Complex, ORC; see below) of a consensus 11 bp AT-rich sequence within an ~200 bp region known as an autonomously replicating sequence (ARS) (Weinreich, Debeer, & Fox 2004; Newlon 1996; Bell & Stillman 1992). Origin sequences in budding yeast are named ARS because of their capacity to confer stability to extrachromosomal DNA molecules and direct their autonomous replication (Hsiao & Carbon 1979). Within the ARS consensus sequence are four distinct regions: an ARS-consensus element (ACS, or A element) and three B elements (B1, B2, and B3) - Figure 1-6A (Marahrens & Stillman 1992). The A element consists of a highly conserved motif (5'- T/ATTTAYRTTTT/A -3', where Y corresponds to a cytosine or thymine base and R to adenine or guanine base) that is absolutely required for recognition by ORC *in vitro*, and mutation within the sequence leads to loss of origin function (Theis & Newlon 1997; Marahrens & Stillman 1992). Although B elements constitute more divergent motifs, collectively they are also required for origin recognition as mutations in their sequences abrogates ORC recognition (Rao & Stillman 1995). Collectively, A and B elements act co-operatively to provide a

landing pad for interaction with specific replication proteins to facilitate the initiation of DNA replication (Bell & Stillman 1992).

This strict requirement for a well conserved DNA sequence element to function as an origin of replication is recognisable only in the budding yeast *S. cerevisiae*. In *Schizosaccharomyces pombe,* clusters of asymmetric AT-rich intergenic sequences with a size range of 0.5 - 1 kb function as budding yeast ACS element analogues - Figure 1-6B (Clyne & Kelly 1995; Dubey *et al*. 1994). Like their budding yeast counterparts, fission yeast asymmetric AT-rich clusters are also able to confer extrachromosomal stability to plasmids and are characterised by A residues on one strand and Ts on the complementary strand (Okuno *et al*. 1999). This sequence, however, displays a degree of heterogeneity in different locations compared with the very well defined budding yeast ARS sequences (Segurado, de Luis, & Antequera 2003; Dubey *et al*. 1994).

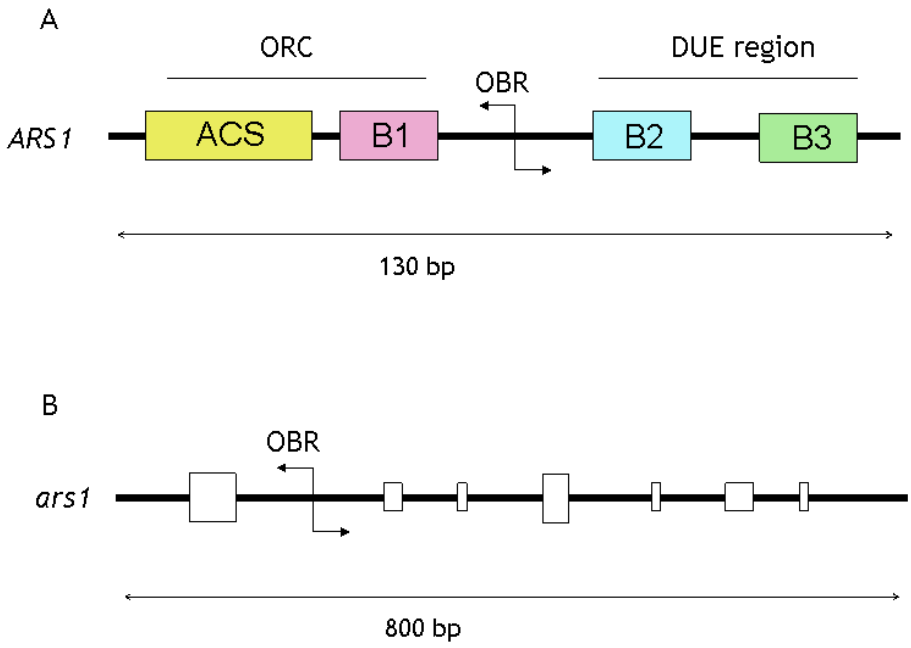


Figure 1-6 – Budding and fission yeast origin of replication map

(A) The first origin locus to be characterised in budding yeast, denoted by *ARS1,* is shown; it comprises an 11-bp consensus element (ACS) and less conserved B elements downstream (B1, B2, B3). ORC high-affinity binding sites are denoted by a horizontal line above ACS and B1; bidirectional replication site is denoted by OBR and AT-rich duplex unwinding elements are denoted by B2 and B3. The approximate size of the locus is shown beneath. (B) In fission yeast, the origin locus *ars1* is shown; the chromosome is depicted by a black horizontal line, white boxes indicate AT-rich ORC binding sites; box sizes reflect affinity of binding sites (the bigger the box, the higher the binding affinity). Figure adapted from (Aladjem, Falaschi, & Kowalski 2006)

In metazoans, it has been shown that replication origins are rarely defined sequence elements, but rather are organised within a broad initiation zone from which there are a number of initiation origins, one or a few of which are capable of firing within S-phase (Aladjem *et al*. 2006). Within the proposed broad initiation zones, in *Xenopus* egg extracts, initiation site selection is dependent on the presence of closely clustered AT-rich asymmetric sequences (Stanojcic *et al*. 2008), which have also been reported for mammalian origins such as the CHO *Dihydrofolate reductase* (*DHFR*) gene, human *ß-globin* locus and Lamin B2 origins - Figure 1-7 (Aladjem *et al*. 2006). Apart from AT-rich clusters, in HeLa cells, ORC preferentially localises with GC-rich tracks and is found sparsely in GC-poor regions; such GC-rich tracks tend to coincide with CpG islands in the genome (Cadoret *et al*. 2008). The AT-rich eukaryotic origins tend to replicate late and/or are inefficient initiation sites, while GC-rich regions are early replicating regions (MacAlpine & Bell 2005). Examining DNA sequence element requirements from higher eukaryotes therefore indicates that there is more flexibility in selection of origins of replication.

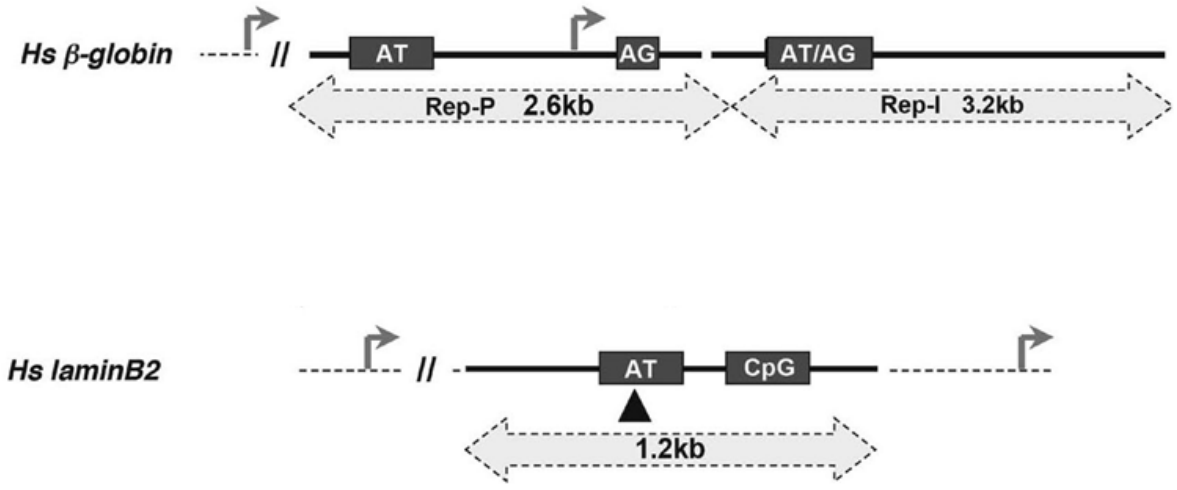


Figure 1-7 – Architecture of the two metazoan chromosomal origins of replication

DNA sequence elements present at the human *β-globin* origin locus (*Hs β-globin)* and the human *Lamin B2* origin locus *(Hs laminB2).* Locations of AT-rich origin loci are indicated by "AT"; asymmetric purine-pyrimidine tracks are indicated by "AG"; and CpG islands are indicated by "CpG"; promoters are indication by arrows with arrow heads pointing in the direction of transcription; bulk arrows indicate an "initiation zone", while filled triangle indicates a precise initiation site, and sizes of initiation zones are indicated within arrows; Taken from (Aladjem *et al*. 2006)

1.5.3 Archaeal origins

Beyond bacterial and eukaryotic systems for understanding the molecular mechanisms that define sequence elements for selection as origins of DNA replication, the archaeal DNA replication machinery has recently emerged as an interesting model system to understand this vital process (Kelman & Kelman 2003). It appears that archaea utilise a blend of both eukaryotic and bacterial systems (Robinson & Bell 2005; Kelman & Kelman 2003). While initiator factors in archaea appear to be a somewhat simplified eukaryotic machine, the sequence elements appear to be well-defined (Robinson *et al*. 2004). In the archaeon *Sulfolobus solfataricus*, Robinson *et al* (2004) identified two origins of replication in its single chromosome and demonstrated that Orc1/Cdc6 protein (see below) is able to bind a motif within the origin sequence termed an origin recognition box (ORB) element, which is also conserved at the positions of characterised origins in a number of *Pyrococcus* species: *P. abyssi, P.furiosus,* and *P. horikosii* (Robinson *et al*. 2004; Matsunaga *et al*. 2003). Based on this finding they proposed that ORB-like consensus motifs (5'-GTTCCAGTGGAAAC-AAA----GGGGG-3' for Crenarchaeal species, and 5'-GTTCCAGTGGAAAC-AAA----GGGGG-3' for Euryarchaeal species) may be a characteristic of archaeal origins (Robinson *et al*. 2004). Subsequently, by measuring the copy numbers of specific markers located close to origin loci relative to markers located at termination sites (a technique known as marker frequency analysis), Lundgren and colleagues were able to show that two *Sulfolobus* species (*S. acidocaldarius* and *S. solfataricus*) actually have three origins of replication, two of which located in close proximity to the locus of the initiator protein gene (*orc1/cdc6),* while no replication-associated gene appeared to be in the vicinity of the third origin (Lundgren *et al*. 2004). Further work by Robinson *et al* (2007) afterwards identified a distantly related eukaryotic Cdt1-like factor, which they named the winged-helix initiator protein (WhiP) (Robinson & Bell 2007). Using two-dimensional gel electrophoresis, they later showed that the third archaeal origin was located in the vicinity of the *whip* gene locus - Figure 1-8 (Duggin, McCallum, & Bell 2008). Like eukaryotes, and unlike bacteria, the multiple origin paradigm seems to be conserved in archaea (Duggin *et al*. 2008; Robinson & Bell 2007; Lundgren *et al*. 2004).

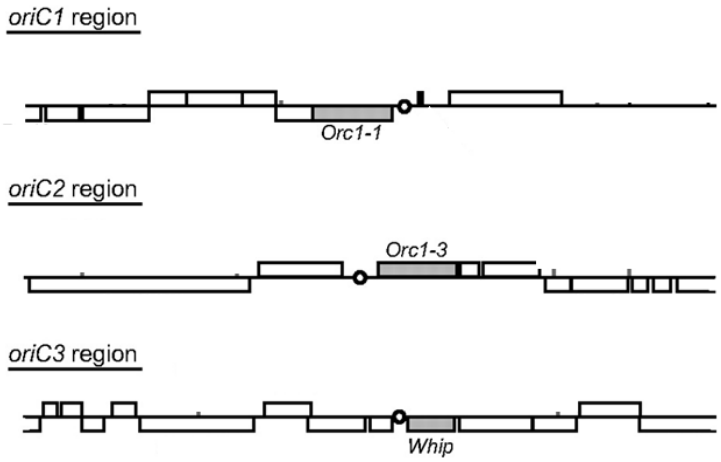


Figure 1-8 – Diagram of *S. acidocaldarius* origins of replication

Map of three *S. acidocaldarius* origin loci indicated by *OriC1* region, *OriC2* region, and *OriC3* region. Horizontal line depicts segments of chromosomes; square boxes above and below the horizontal line depict direction of transcription; circles indicate the predicted location of origins, which are located either upstream of an initiator factor gene (*Orc1-1*) or downstream of initiator factor genes (*Orc1*-3 and *Whip);* grey shaded regions indicate loci for initiator factors (Orc1-1, Orc1-3, and Whip). (Duggin *et al*. 2008)

1.5.4 Trypanosomatid origins

Unlike the much-studied mitochondrial genome replication, nuclear DNA replication initiation in *T. brucei* remains a largely unexplored subject (El Sayed *et al*. 2005a). In fact, the latter is not only true for *T. brucei* but also true for its close relatives, *L. major* and *T. cruzi* (El Sayed *et al*. 2005a), and for protists in general. The first attempt geared towards finding origins of nuclear DNA replication in *T. brucei* was made by Patnaik and colleagues (Patnaik, Kulkarni, & Cross 1993). In this study, random DNA segments from *T. brucei* chromosomes were cloned into bacterial circular plasmids and tested for their ability to confer autonomous replication on these extrachromosomal DNA molecules, which are otherwise unable to survive as an episome (Patnaik *et al*. 1993). The study identified two plasmids (pT13-11 and pT13-41) that were able to be maintained as single copy episomes in PCF cells, and one plasmid (pT11-bs) that was extrachromosomally stable in BSF cells. The region that conferred their extrachromosomal stability was attributed to a large inserted gene fragment, which they termed the plasmid maintenance sequence (PMS), and deletion of this region abrogated the ability of the plasmids to survive as an episome (Patnaik *et al*. 1993). Further work by the same group later characterised pT13-11 in an attempt to identify the specific region of the genome that played a role

in its ability to survive as an episome (Patnaik, Fang, & Cross 1994). Here, they showed that an interaction between the PMS and a 108 bp region encompassing the *PARP* gene promoter region were both necessary for stable extrachromosomal existence. If the *PARP* gene region was replaced with a *rRNA* promoter region, the ability of pT13-11 to survive as an episome was incapacitated, suggesting that this region played a role in successful propagation of the plasmid in PCF *T. brucei* (Patnaik *et al*. 1994). Although the authors failed to physically map the specific location of the origin(s) either within the plasmids or in their natural chromosomal context, an important finding from both studies (Patnaik *et al*. 1994; Patnaik *et al*. 1993) was the demonstration of a possible link between the ability of the *PARP* gene promoter region to confer episome stability and its additional role in the same plasmid to drive the transcription of a selectable marker, thereby supporting a close relationship between DNA replication and transcription regulatory elements in *T. brucei;* a phenomenon that is now emerging as a paradigm in higher eukaryotes, including yeasts and humans (Maric & Prioleau 2010; Hiratani *et al*. 2009; Zappulla, Sternglanz, & Leatherwood 2002).

Although the focus of this thesis is on nuclear DNA replication in *T. brucei,* it is worth mentioning that prior to the work by Patnaik and colleagues (1993, 1994), the lab of P.T. Englund had reported the presence of a conserved region (100 – 200 bp in length) in mitochondrial (kDNA) minicircles that contained a replication origin in *T. equiperdum* (Ntambi *et al*. 1986). Within this region is a consensus dodecameric sequence 5'-GGGTTGGTGTA-3' [known as the universal minicircle sequence (UMS)] and a hexameric sequence 5'-ACGCCC-3' that are both present in all minicircles examined from virtually all trypanosomatid species (*T. brucei, T. Equiperdum, L. Tarentolae, T.lewisi, T. cruzi, and Crithidia fasciculata*) (Ray 1989). The UMS is associated to minicircle replication initiation in the leading strand while the hexameric sequence is the initiation site for the first okazaki fragment (Morris *et al*. 2001). Unlike minicircles, much less is known about maxicircle replication initiation in *T. brucei* (Morris *et al*. 2001).

In parallel with the work by Patnaik and colleagues (1993, 1994) and the work by the lab of P.T Englund described above, another laboratory identified a 1 kb mitochondrial kDNA minicircle fragment that permitted autonomous replication

of a plasmid (pTbo-1) in the nuclei of PCF *T. brucei* cells (Metzenberg & Agabian 1994). This 1 kb minicircle kDNA fragment had within it the consensus UMS motif, and was maintained as a supercoiled concatemer of approximately nine monomeric units (Metzenberg & Agabian 1994). Although pTbo-1 was maintained stably as a concatemer in the nucleus under hygromycin selection, it was lost in the absence of drug selection, suggesting that the minicircle DNA fragment alone was necessary but not sufficient for plasmid maintenance. Also worthy of note is the fact that the organisation (concatenated multimers) of pTbo-1 in the nucleus is very reminiscent of minicircle organisation in the mitochondrion of *T. brucei* (Liu *et al*. 2005). This may imply that the presence of the UMS within the 1 kb mitochondrial DNA fragments could impose some mitochondrial properties in the nucleus, although its maintenance is likely to be driven by nuclear-specific DNA replication proteins, since the UMS binding proteins that have been shown to bind the UMS are not localised to the nucleus (Milman *et al*. 2007). It had been proposed in two earlier papers (Papadopoulou, Roy, & Ouellette 1994; Patnaik *et al*. 1993) that the existence of episomes as multimers (rather than as single copy molecules) suggests that multiple weak or adventitious origins are likely to fire from the plasmid within the cell cycle, which might therefore facilitate the random selection of an origin within the plasmid sequence. This proposition lends support to the notion that pTbo-1's stability is unlikely to be driven by a single bona fide origin of DNA replication. However, if it does, could this mitochondrial DNA fragment function as a nuclear DNA replication origin? This remains to be investigated, though it seems unlikely that mitochondrial and nuclear chromosome would naturally share DNA replication origin sequences.

Whereas in *T. brucei* both the *PARP* promoter gene region and the PMS are required for extrachromosomal stability (Patnaik *et al*. 1994), in related trypanosomatids such as *L. major* and *T. cruzi,* bacterial plasmids entirely deficient in trypanosomatid-derived sequences can be maintained as stable episomes (Kelly 1995; Papadopoulou *et al*. 1994). These bacterial plasmids could only replicate autonomously in *T. brucei* if they had undergone rearrangements or acquired sequences from the genome by a process of recombination (Patnaik 1997). These findings led to the assertion that in *L. major* and *T. cruzi,* but not in *T. brucei,* the initiation of DNA replication is likely to occur at loci that do not

have specific sequence requirements (Kelly 1995; Papadopoulou *et al*. 1994; Patnaik *et al*. 1993).

Apart from studying autonomously replicating circular plasmids in trypanosomatid species, other studies have focused on autonomously replicating linear DNA molecules, otherwise known as artificial linear mini-chromosomes (Patnaik *et al*. 1996; Lee, E, & Axelrod 1995). In both studies (Patnaik *et al*. 1996; Lee *et al*. 1995), the authors used constructs bearing the PMS and the *PARP* gene promoter (derived from pT13-11 described above), but this time added telomeric repeats and subtelomeric *T. brucei*-derived sequences at the ends of the constructs. The linear constructs, when transfected into PCF *T. brucei* cells, could be maintained as stable extrachromosomal elements only after extensive rearrangements with additional sequences present at subtelomeric and telomeric loci (Patnaik *et al*. 1996; Lee *et al*. 1995). The authors asserted that the additional sequences which were located were most likely of minichromosomal origin acquired through recombination events, although they were not fully characterised (Patnaik *et al*. 1996; Lee *et al*. 1995). As these additional sequences remain unidentified and uncharacterised, it still remains unclear whether they conferred the ARS activity to these artificial linear molecules, or whether the sequences will display ARS activity in a natural chromosomal context.

Prior to experiments to investigate the stability of artificial linear DNA molecules in *T. brucei,* the structure and DNA nucleotide sequence of naturally occurring minichromosomes had been studied (Weiden *et al*. 1991). Using sucrose gradient fractionation, minichromosomes were isolated from *T. brucei* cells and visualised by electron microscopy (EM). The authors reported the identification of a variety of putative replication intermediates in the minichromosomes, including various bubble and replication fork structures. A more remarkable observation was the identification of a single replication bubble on all minichromosomes analysed, suggesting the presence of a single origin of DNA replication in each molecule (Weiden *et al*. 1991). For five molecules (with a size range of 47.6 kb to 110.5 kb) the analysed bubble length ranged from 7.2 kb to 18.2 kb, giving a mean position of the replication bubble a distance of ~ 35 % from one end of the molecules (Weiden *et al*. 1991). In the same study, structural analyses of minichromosomes also revealed that most of their content comprises AT-rich sequence, which was subsequently shown to be

organised as repeats (termed 177 bp repeats) reminiscent of viral, or perhaps even yeast, origins (Wickstead *et al.* 2004) and found also in intermediate chromosomes. Whether these AT-rich repeats are present on *T. brucei* megabase chromosomes and behave in the same way as observed in the minichromosomes remains to be investigated.

In an attempt to define the structure and organisation of centromeres in *T. cruzi* megabase chromosomes, telomere-associated chromosome fragmentation studies mapped centromeric DNA for chromosome (Chr) 3 to a 16 kb GC-rich region that separates two directional gene clusters (the SSR discussed above) (Obado *et al.* 2005). In a follow-up study by the same group, etoposide-mediated Topoisomerase II cleavage mapping demonstrated that an analogous 40 kb SSR region on Chr. 1 functions as a centromere in *T. cruzi* (Obado *et al.* 2007). Using the same method in *T. brucei,* centromeres were also mapped to SSRs (Obado *et al.* 2007). However, the latter were composed of arrays of AT-rich repeats, as opposed to the GC-rich sequences observed for *T. cruzi* (Obado *et al.* 2007; Obado *et al.* 2005). In both *T. cruzi* and *T. brucei,* these centromeres contained degenerate retrotransposon elements that conferred mitotic stability to the chromosomes (Obado *et al.* 2007; Obado *et al.* 2005). However, the lack of sequence conservation of the similarly mapped SSRs in both *T. cruzi* and *T. brucei* does not allow us to state if the regions also act as replication origins. This claim is supported by evidence from the SSR on Chr. 1 of *L. major,* which has been shown not to be essential for mitotic stability of the chromosome (Dubessay *et al*. 2002).

To summarise, for all the attempts that have been made to define sequences that confer replicative stability to extrachromosomal DNA molecules, no clear consensus sequences have been reported. In addition, no studies in trypanosomatids (*T. cruzi, T. brucei* and *L. major*) have directly examined the presence of replication origins in the megabase chromosomes, although these chromosomes contain most of the parasite's genetic content (Berriman *et al*. 2005). Here, we have interrogated the *T. brucei* genome in a search for TbORC1/CDC6 binding sites (see below), with the ultimate goal of delineating DNA replication origins in all 11 megabase chromosomes.

1.6 The DNA replication initiation machinery; a synopsis

DNA replication is central to the propagation of life and initiates by the designation of genome sequences as origins, where synthesis of a copy of the genetic material begins once per cell division. In all organisms studied (bacteria, archaea and eukaryotes), DNA replication requires a chronological sequence of events: the definition of replication start site, termed an origin of DNA replication, by binding of specific factor(s) (see below); the recruitment of regulatory factors (see below), which then guide the recruitment of a replicative helicase complex to locally unwind the double helical structure of DNA at the origin locus; the recruitment of primases to provide a 3'-OH group for extension of nascent strand DNA by DNA polymerases; and finally termination of leading strand replication and filling-in of okazaki fragments on the lagging strand (these final steps need also be co-ordinated with replication of the linear chromosome ends (telomeres) via telomerase).

The establishment of a competent start site is known as the initiation step and this is the key point of DNA replication regulation, which determines which origins are fired, as well as the rate of replication (Sclafani & Holzen 2007). For example, in *Drosophila* and *Xenopus* embryos where cells are dividing more rapidly, origins occur and fire randomly without a defined initiation point, whereas in somatic cells origins becomes more defined and are fired distinctly (Sasaki *et al*. 1999; Hyrien, Maric, & Mechali 1995). Since this work only investigates DNA replication initiation events in *T. brucei*, we only focus on the molecular machinery used for the initiation process and its associated regulatory events in bacteria, archaea and eukaryotes. We will then consider our current understanding of the conservation of this process in trypanosomatids. Although the process of DNA replication is well conserved among characterised eukaryotes, it is surprising that there is considerable diversity in the modes of recognition and regulation used to ensure the correct duplication of genomes in bacteria, archaea and eukaryotes (Kawakami & Katayama 2010), given the essential nature of this process.

1.6.1 DNA replication initiation factors in bacteria

In bacteria DNA replication initiates from a single well-characterised origin of replication in its circular chromosome termed *OriC* (Fuller *et al.* 1984). Pioneering *in vitro* work on DNA replication initiation in *Escherichia coli* in the Kornberg laboratory shed light on our understanding of the process in bacteria; for comprehensive review see (Mott & Berger 2007; Messer 2002). The fundamentals of the process are conserved throughout the bacterial kingdom, and here we focus on the molecular players involved in DNA replication initiation in *E. coli*.

In *E. coli* the main initiator protein of DNA replication is DnaA and the initiation process begins with the recognition and binding of DnaA-ATP to five 9-bp repeat elements known as DnaA boxes, followed by a DnaA-ATP interaction with a second class of DNA sequences termed I-sites, which are interspersed among the DNA boxes (McGarry *et al.* 2004; Schaper & Messer 1995). Once bound to the DnaA boxes and I-sites, DnaA-ATP binds a further 13-bp AT-rich repeat element known as DUE (*D*NA *u*nwinding *e*lement) located upstream of the DnaA boxes (Speck, Weigel, & Messer 1999; Bramhill & Kornberg 1988). These interactions of DnaA induce cooperative loading of more DnaA monomers, which facilitates a conformational distortion at the intrinsically less stable AT-rich DUE, thereby causing it to melt (Speck & Messer 2001). A consequence of the melting event is the facilitated loading of two hexamers of the replicative helicase DnaB, mediated by its loader, DnaC (Carr & Kaguni 2001). The presence of all three factors (DnaA, DnaB and DnaC) licenses *OriC* for recruitment of downstream factors: single-stranded DNA binding proteins (SSBPs), which bind each parent DNA template to prevent premature re-annealing; a primase (DnaG), which synthesises 10-12 nucleotide RNA primers for priming of each template indicating the end of initiation and the onset of elongation; and finally elongation by DNA polymerases (DNA polymerase-III holoenzyme) (Fang *et al.* 1999).

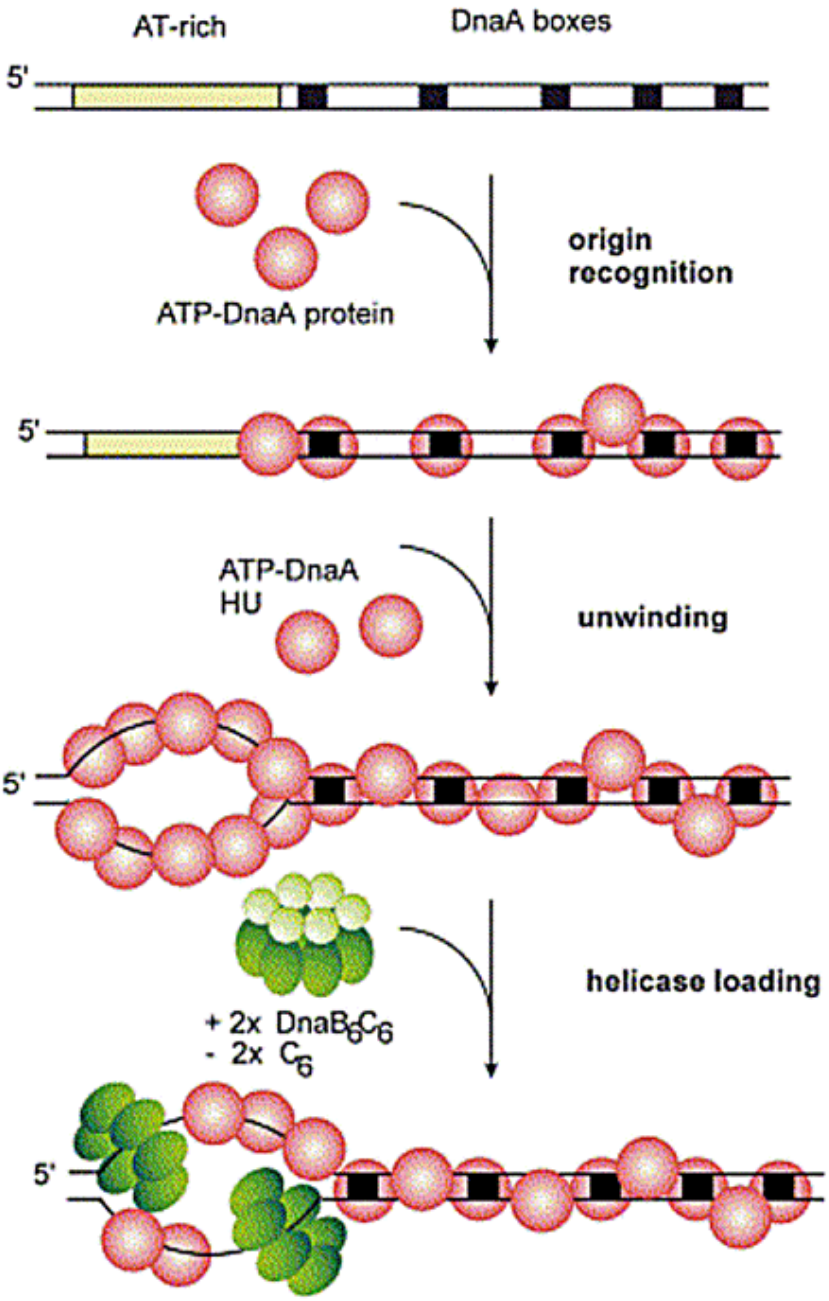


Figure 1-9 – Model of DNA replication initiation in bacteria

The assembly of DnaA, DnaB and DnaC at *OriC: OriC* is represented by parallel horizontal lines; black boxes represent DnaA boxes; I-sites are interspersed between DnaA boxes; yellow shaded area represents 13-mer AT-rich DUE; six white balls represented DnaC which loads DnaB represented by green shaded ovals; Upon loading of DnaB onto melted DUE, DnaC is released from DnaB. Figure taken from (Messer, 2002)

Structurally, based on homology sequence alignment and functional biochemical characterisation, four domains have been identified in DnaA that are crucial for its replication function: an N-terminal domain (amino acid residues 1-86) essential for its interaction with DnaB and homo-oligomerisation (Weigel *et al*. 1999; Messer *et al*. 1999); a linker domain (amino acid residues 87-134), which is

variable in sequence among bacterial species and appears not to be essential for DnaA function (Messer *et al*. 1999); a AAA^+ ATPase domain (amino acid residues 135-373), which confers its ATPase activity necessary for unwinding of the DUE element (Sekimizu *et al*. 1987); and a C-terminal helix-turn-helix domain, which specifically binds *OriC* (Roth & Messer 1995). The ability of DnaA to bind and hydrolyse ATP serves as a key regulatory step of DNA replication initiation in *E. coli*. Although the energy from ATP hydrolysis is not required for co-operative DnaA-*OriC* interactions, as a non-hydrolysable ATP analogue does not abrogate these interactions, the allosteric modifications induced by the binding event has been shown to be essential for melting the DUE (McGarry *et al*. 2004; Sekimizu *et al*. 1987). The helicase loader DnaC is also belongs to the family of AAA^+ ATPases (Iyer *et al*. 2004), and here, hydrolysis of ATP is essential for DnaC to be released from DnaB after it has been loaded onto DnaA (Fang *et al*. 1999). To limit replication to occur only once per cell cycle, at least two other modes of regulation have been described, among others (see below).

One mode of regulating re-initiation of DNA replication in *E. coli* involves a protein known as SeqA. This protein co-operatively and preferentially binds hemimethylated GATC repeats (the recognition sequence of Dam methyltransferase) interspersed among the DnaA-boxes at *OriC*, thereby sequestering the newly replicated *OriC* from DnaA and preventing re-initiation of replication (Nievera *et al*. 2006). A second system used by *E. coli* to control DNA replication initiation occurs by titration of available DnaA by a locus called the *datA* locus (Ogawa *et al*. 2002). This locus has a high affinity for DnaA and null mutants have been shown to undergo several rounds of re-replication, while additional copies leads to incomplete replication, suggesting that it is necessary to adjust the concentration of available DnaA (Kitagawa *et al*. 1998).

1.6.2 DNA replication initiation factors in eukaryotes

In contrast with the bacterial DNA replication initiation machinery, the initiation of DNA replication in all eukaryotes examined to date, including a number of yeast species, *Drosophila*, *Xenopus* and mammals (including humans) requires the binding of at least four different proteins: a six protein complex known as the origin recognition complex (ORC; composed of factors named Orc1-Orc6),

Cdc6, Cdt1, and the Mcm2-7 helicase complex (Bell & Dutta 2002). Once ORC has bound, Cdc6 is recruited onto the ORC-origin complex, prompting the recruitment of the replicative helicase via a preformed heptameric complex composed of Cdt1 bound to the Mcm2-7 complex (Tsakraklides & Bell 2010). ORC, Cdc6, Cdt1 and Mcm2-7 together form a larger complex known as the pre-replication complex (pre-RC) at the origin of replication (Figure 1-10). Pre-RC formation and activation is tightly regulated by the cell so that only pre-existing pre-RCs formed during the G_1 phase are activated in S-phase (Bell & Dutta 2002). S-phase-promoting cyclin-dependent kinases facilitate loading of additional initiator factors (MCM10, Cdc45 and GINS) to ensure that the cell only replicates its DNA once per cell cycle. Each component of the pre-RC seems to play a well-defined role in this regulation and activation process (see Section 1.6.3). Despite this mechanistic similarity between bacteria and eukaryotes, no evidence has suggested that the proteins involved are conserved.

The precise way that ORC designates origins shows remarkable plasticity, meaning that it is not feasible to predict replication origins by sequence alone. For example, the budding yeast *S. cerevisiae* uses a limited set of origins each with a conserved DNA sequence element, while in metazoans, no consensus DNA sequence has been found to function at DNA replication initiation sites (see Section 1.5).

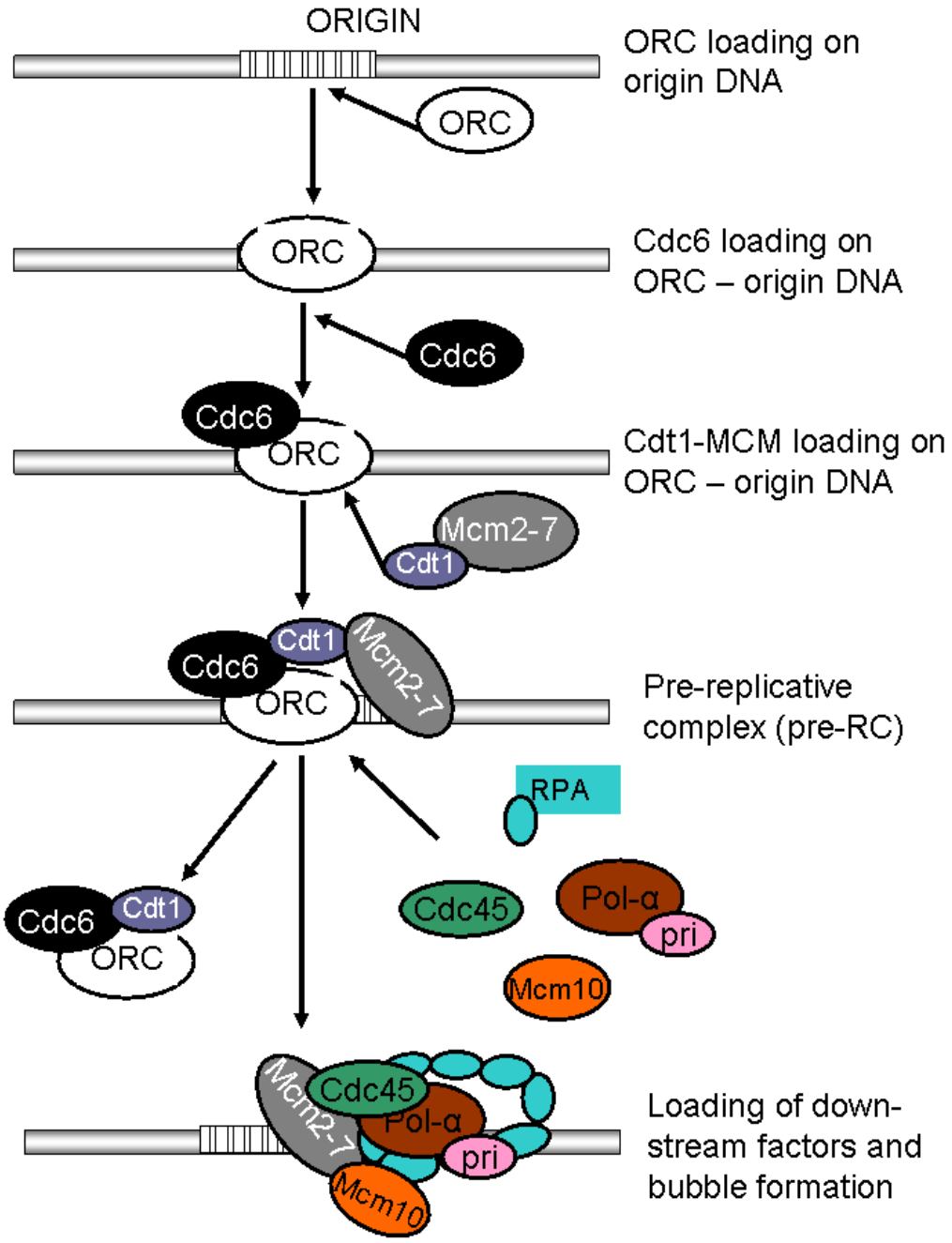


Figure 1-10 – Model of Pre-RC assembly as currently understood in budding yeast

The assembly of the pre-RC complex in eukaryotes requires the recruitment of ORC (Orc1-6) on to a specific locus of a chromosome (horizontal grey line) known as an origin of replication (striped box). Origin-bound ORC recruits Cdc6 and this DNA-protein complex then triggers helicase loading in the presence of a pre-formed Cdt1-Mcm2-7 complex. Binding of regulatory factors (Mcm10, replication factor A (RPA), Cdc45, GINS, and cyclin-dependent kinases) following helicase loading triggers the release of ORC, Cdc6 and Cdt1 from origin DNA. Binding of Cdc45 promotes binding of primases (Pri) and DNA polymerase- α (Pol- α), and initiation of DNA replication begins with the formation of a replication bubble. Notice that some MCM subunits are shown to bind origin DNA. Model is adapted from (Francon, Maiorano, & Mechali 1999) and includes modification based on recent *in vitro* work by (Tsakraklides & Bell 2010)

1.6.3 Components of eukaryotic DNA replication initiation machinery

1.6.3.1 The origin recognition complex (ORC) proteins

The origin recognition complex (ORC) is a six component protein consisting of five related proteins (Orc1-Orc5) that share a similar modular domain architecture (Figure 1-11), and another protein that lacks this structural organisation; Orc6. ORC proteins were first purified thanks to their ability to specifically recognise and bind ARS sequences in *S. cerevisiae;* DNA sequences present at *S. cerevisiae* origins of replication (Bell & Stillman 1992). Initially, the well-characterised budding yeast origin locus *ARS1* was used as 'bait' to affinity purify proteins that were specifically bound to origins from a yeast nuclear extract (Bell & Stillman 1992). Using glycerol gradient centrifugation coupled with SDS-PAGE on the purified nuclear extracts, a multi-protein complex was identified that specifically recognised ARS sequences. This multi-protein complex was shown to protect *ARS1* from DNaseI nuclease activity *in vivo*. Similar experiments using *ARS307* and *ARS121* suggested a universal role of ORC to bind yeast origin sequences in an ATP-dependent manner (Bell & Stillman 1992). Over the years, greater progress has been made to understand the evolutionary conservation of ORC proteins across species, its role in DNA replication, as well as its functions beyond DNA replication (Duncker, Chesnokov, & McConkey 2009).

As in budding yeasts, all six subunits have been shown to be conserved in other eukaryotic species: *Drosophila melanogaster, Xenopus laevis, Mus musculus, Arabidopsis thaliana* and *Homo sapiens* (Dutta & Bell 2006). Orc1 through to Orc6 are thus numbered (in descending order of their molecular weight) based on their migration on SDS-PAGE, with Orc1 being the largest subunit and Orc6 being the smallest subunit. The assembly of ORC in *S. cerevisiae* occurs in a stepwise fashion, with specific interactions occurring between different subunits in the following order: Orc2 and Orc3 interact, Orc4 and Orc5 interact, and the dimeric Orc4-Orc5 is capable of interacting with dimeric Orc2-Orc3 and separately with Orc1 with to form a pentameric Orc1-Orc5 (Chen *et al*. 2008). Pentameric Orc1-Orc5 is fully functional in *S. cerevisiae* (Lee & Bell 1997),

although Orc6 has been shown to interact with this complex to form the ORC hexamer (Chen, de Vries, & Bell 2007). In humans, Baculovirus expression of recombinant ORC proteins in insect cells and subsequent purification revealed that Orc2 interacts with Orc3, which then recruits Orc4-Orc5 to form a core complex. This core complex of Orc2-Orc5 then recruits Orc1 and Orc6 (Dhar, Delmolino, & Dutta 2001), consistent with the budding yeast model (Chen *et al.* 2008).

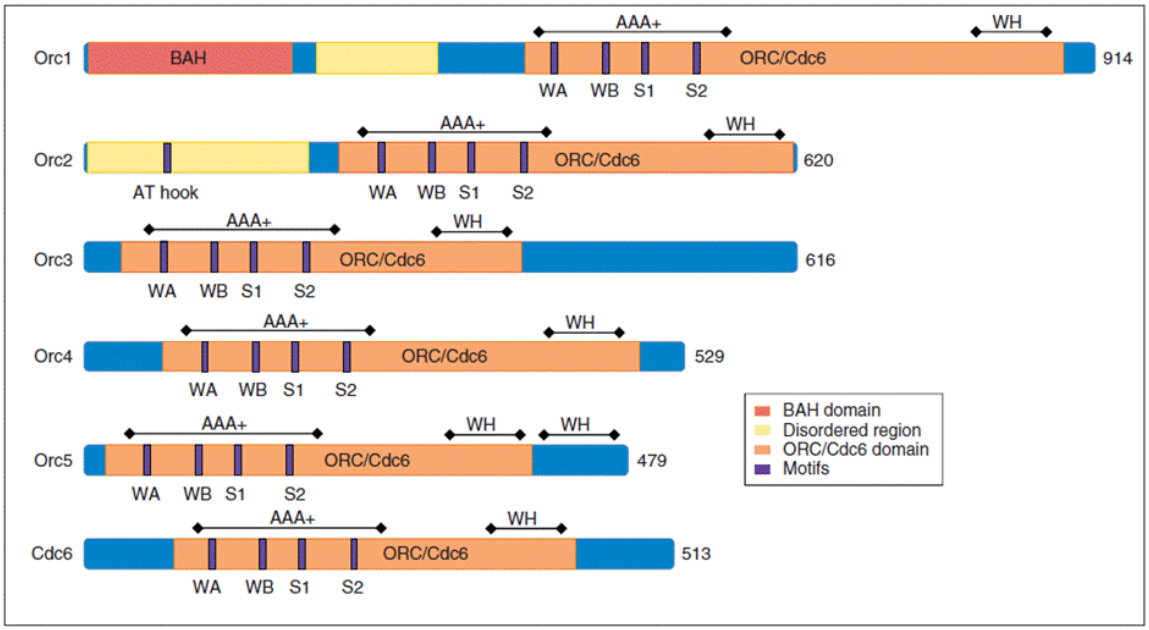


Figure 1-11 – Structural domain organisation of Orc1-Orc5 and Cdc6 from *S. cerevisiae*

Orc1, Orc4, Orc5, and Cdc6 each contain an AAA^+ domain indicated by AAA+ (orange; ORC/Cdc6). Orc2 and Orc3 are shown to have the AAA^+ domain but lack the associated ATPase activity. Amino acid sequence motifs within the AAA+ domain: Walker A (WA), Walker B (WB), Sensor-1 (S1) and Sensor-2 (S2) are shown. ORC and Cdc6 posses a winged-helix domain (WH), involved in DNA binding. Orc1 contains an additional BAH (bromo-adjacent homology) domain (pink); Orc1 and Orc2 have regions of disorder (yellow); a DNA-binding AT-hook motif (here PRKRGRPRK) is identified in *S. cerevisiae* Orc2, and several of these have also been identified in disordered regions in *S. pombe* Orc4.The number of amino acids for each protein is indicated at the right. Taken from (Duncker *et al.* 2009)

All ORC subunits, with the exception of Orc6, have been shown to associate specifically with the A and B1 elements (~50 bp in length) of *ARS* at replication origins in *S. cerevisiae* (Lee & Bell 1997; Rao & Stillman 1995). Later, based on secondary structure conservation, it was proposed that that the interaction between the Orc1-5 complex with DNA occurs via a winged-helix DNA binding domain, which is present either at the C-terminus or internally (Speck *et al.* 2005) of each ORC subunit. Apart from the budding yeast paradigm, where the

ORC-DNA interaction is dependent on the presence conserved ARS motifs, its interactions with origin DNA in other eukaryotic systems is less well defined. In *S.pombe,* for example, under mild salt extraction conditions (0.3 M NaCl) Orc4 remains associated with chromatin while the other five subunits (Orc1, Orc2, Orc3, Orc5, and Orc6) could be purified from the extract (Moon *et al*. 1999). Further biochemical analysis of the Orc4 subunit revealed the presence of nine AT-hook motifs at its N-terminal domain that preferentially recognised the high AT-rich clusters present at *S. pombe* origins (Lee *et al*. 2001), and it was later shown that this Orc4-specific interaction was the key event that directed the recruitment of the other five subunits (Kong & DePamphilis 2001b). In *D.melanogaster* and *X. laevis* embryos, where cells are dividing more rapidly, DNA sequence recognition by ORC proteins occurs randomly, whereas in somatic cells ORC's DNA-sequence recognition becomes more defined (Sasaki *et al.* 1999; Hyrien *et al.* 1995), although any consensus is yet to be identified.

A functional and stable ORC-DNA complex is a prerequisite to carry on and complete pre-RC assembly. Mutational and biochemical analyses show that there is considerable variability in the interplay and involvement of ORC subunits in ATP binding and hydrolysis in different organisms. Structurally, three of the six ORC subunits (Orc1, Orc4, and Orc5) belong to the AAA^+ family of ATPases, based on consensus Walker A and B motifs in their primary amino acid sequences (Speck *et al.* 2005; Neuwald *et al*. 1999). This suggests that ORC function, like DnaA in bacteria, is tightly regulated by binding and/or hydrolysis of ATP, reminiscent of a common evolutionary origin (Iyer *et al.* 2004). Studies in *S. cerevisiae* have shown that Orc2 and Orc3 lack the key conserved α-helix and ß-sheet secondary structure elements of the ATP binding site, and thus appear to be more distant relatives of the AAA^+ family of ATPases (Speck *et al.* 2005). In *S. cerevisiae* the stability of the ORC complex appears to be mediated by the Orc5 ATP-binding motif, since a mutation in the Walker A motif (K43E) leads to overexpression of the Orc4 subunit and a slower growth rate at 37 °C due to ORC dissociation (Takahashi *et al*. 2004). In another study by Klemm *at al* (1997), a mutation in the Walker A motif (K485T) of Orc1 led to impaired ORC binding to *ARS1,* suggesting that ATP-binding to the Orc1 subunit is essential for ARS interaction (Klemm, Austin, & Bell 1997). The role of the Walker B motif has been established in *S. cerevisiae* by mutating a key conserved Aspartate to

Tyrosine (D569Y) in Orc1 (Klemm & Bell 2001). Overexpression of this dominant negative mutant allele is lethal and the D569Y mutant in Orc1 cannot hydrolyse ATP, suggesting an important role of Orc1-ATP binding and hydrolysis (Klemm & Bell 2001; Klemm *et al.* 1997). Mutations of a conserved arginine residue in the so-called "arginine finger" in *S. cerevisiae* Orc4 have been shown to impair loading of the replicative helicase (Mcm2-7) (Bowers *et al*. 2004), suggesting an essential role for the Orc4 subunit in pre-RC assembly. Although Orc6 lacks the aforementioned functionally conserved AAA^+ domains (Speck *et al.* 2005) and shows no affinity for ARS sequences, it still remains essential for cell viability in budding yeasts (Lee & Bell 1997) and is required for recruitment of the downstream factor Cdc6 in *S. cerevisiae,* the next essential step in pre-RC assembly (Chen *et al.* 2007). This scenario is contrary to observations in *D. melanogaster,* where all six subunits are required for origin recognition and DNA replication (Chesnokov, Remus, & Botchan 2001).

The subcellular localisation of ORC, and its eventual fate post-replication initiation, still remains debatable from eukaryotic DNA replication studies. Depending on the visualisation technique, the results obtained from localising ORC proteins not only do not agree for different cell cycle stages and across species, but also depend on the specific subunit of ORC that is being localised. For example, in *D. melanogaster* embryos, imaging of a green fluorescent protein (GFP)-tagged Orc2 revealed a temporal association with chromatin, where it is excluded from chromosomes until anaphase (when mitotic cyclin activity is depleted)(Baldinger & Gossen 2009). Earlier studies in *D. melanogaster* larval neuroblasts, which used anti-Orc2 antibody and indirect immunofluorescence (Loupart, Krause, & Heck 2000), and studies on *X. laevis* Orc1 and Orc2 by the same method, showed that these subunits are also excluded from metaphase chromatin (Romanowski *et al*. 1996). In Hela cells, using chromatin fractionation experiments, human Orc1 appears to be dissociated from chromatin during S-phase to M-phase transition (Kreitz *et al*. 2001). It has been suggested that in Chinese hamster Ovary (CHO) cells Orc1 and not Orc2 is targeted for degradation via a 26S proteasome ubiquitination pathway (Li & DePamphilis 2002). Together these findings suggest that Orc1 does not permanently associate with chromatin throughout the cell cycle in these organisms. On the contrary, in other studies in CHO cells (McNairn *et al*. 2005)

and *D. melagonaster* embryos (Pak *et al*. 1997), Orc1 and Orc2 subunits localised to chromatin throughout the cell cycle, including metaphase. In *S. cerevisiae* imaging YFP-tagged Orc6 and Myc-tagged Orc2 showed that ORC localised to punctate subnuclear foci throughout the cell cycle, consistent with a role in DNA replication (Semple *et al*. 2006; Pasero *et al*. 1999).

Using polyclonal anti-Orc6 antiserum, localisation of human Orc6 to kinetochores and to a reticular-like structure around the cell periphery in a cleavage furrow has been reported (Prasanth, Prasanth, & Stillman 2002). In another report, human Orc2 subunit localises to centrosomes and centromeres during mitosis (Prasanth *et al*. 2004). More recently, in human cells Orc1 has been shown to be involved in regulating centriole copy number, which is consistent with a localisation at centromeres (Hemerly *et al*. 2009). These results are indicative of a role for ORC in chromosome segregation and hence cell cycle progression in human cells. Given that individual ORC subunits display differential subcellular localisation in different species, further work is therefore needed to reach a consensus of the localisation of the ORC. However, in all of these studies, nuclear localisation seems to be common across species.

Apart from the other non-replication roles of ORC that have been reported from sub-cellular localisation studies, the first non-replication role of ORC proteins to be reported arose from the discovery that ORC in *S. cerevisiae* protected all four silent mating-type loci, *HMR-E, HMR-I, HML-E, and HML-I*, from DNaseI digestion in foot-printing experiments, suggesting that it mediates the establishment of transcriptionally silent chromatin (Bell, Kobayashi, & Stillman 1993). Further experimental work on the transcriptional regulatory role of ORC showed that deletion of an N-terminal region (235 amino acids) of Orc1 from *S. cerevisiae* derepresses the *HMR* locus (Bell *et al*. 1995). In addition, *S. cerevisiae* Orc1 was shown to contain a Bromo-adjacent homology (BAH) and the full length protein was 50 % identical and 63 % similar to Silent Information Regulator 3 (Sir3), a protein that has been shown to play a role in transcriptional silencing, confirming that the N-terminus of Orc1 was required for epigenetic silencing at this locus (Bell *et al.* 1995). This finding is supported by subsequent work, which showed that Orc1 interacts with Sir1 (Triolo & Sternglanz 1996). Here, the authors propose a model for the establishment and spreading of heterochromatin at the *HM* loci orchestrated by an interaction between Orc1 with Sir1 (Triolo &

Sternglanz 1996). This interaction recruits Sir4, Sir3 and Rap1 in a complex that interacts with histone H3 and H4 tails, leading to chromatin compaction (Triolo & Sternglanz 1996). It is now well established that apart from a direct interaction of Orc1 with Sir1, Orc1 also interacts with the transcription factors Rap1 and Abf1, and after recruitment of Sir3, Sir4 and Sir2, the latter deacetylates adjacent histone tails, facilitating the methylation of H3 at lysine 79 by Dot1, among other histone modifications [reviewed in (Burgess, Guy, & Zhang 2009)]. While the aforementioned are true for *S. cerevisiae*, Orc1's role in transcriptional regulation has also been described in a number of eukaryotes, for example *D. melanogaster* and humans.

In *Drosophila* and human cells, Orc2 has been shown to co-localise with the Heterochromatin Protein 1 (HP1) (Prasanth *et al.* 2004; Pak *et al.* 1997), while in *Xenopus* and humans, Orc1 co-localises with HP1 (Lidonnici *et al*. 2004; Romanowski *et al.* 1996). Both these findings support a role for ORC in heterochromatin formation and hence epigenetic silencing, though what genes may be targeted is unknown. It has been proposed that a possible rationale for ORC's interaction with HP1 is to facilitate progression of a replication fork through heterochromatin (Mechali 2010). This hypothesis is supported by the observation that architectural remodelling of chromatin mediated by ISW1 facilitates progression of a replication fork (Mechali 2010; Collins *et al*. 2002). Since the discovery of ORC's role in silencing, studies on ORC proteins have established its involvement in ribosome biogenesis, sister chromatid cohesion, cytokinesis, among others [see (Sasaki & Gilbert 2007) for a comprehensive review].

How is ORC recruited to chromatin? Like ORC's localisation, this subject still remains controversial. Several mechanisms have been suggested: a non-coding RNA-dependent mechanism in *Tetrahymena thermophila* (Donti *et al*. 2009; Mohammad *et al*. 2007) and in yeasts and mammalian cells (Deng *et al*. 2009); and via interactions with other proteins, for example EBNA1 in Epstein-Barr virus (Norseen *et al*. 2008). This topic is discussed in more depth in Section 4.10.6 where preliminary interactions studies of *T. brucei* ORC1/CDC6-interacting proteins suggest a possible RNA-binding protein-dependent recruitment of TbORC1/CDC6 to chromatin in trypanosomatids.

ATP-dependent ORC recognition and binding to origins of replication serves to provide a platform to load a functional pre-RC, which involves the recruitment of downstream factors: Cdc6, Cdt1 and the minichromosome maintenance proteins (Mcm2-Mcm7). These are discussed below.

1.6.3.2 Cell division cycle 6 (Cdc6) protein

The Cell division cycle 6 (Cdc6) protein was first identified in *S. cerevisiae* as a factor essential for re-initiation of DNA synthesis (Hartwell 1976). At non-permissive temperatures, Hartwell (1976) showed that *S. cerevisiae* mutants in the *cdc6* gene were only able to complete a single round of replication and were unable to re-initiate a second round, suggesting an essential role of the protein in DNA replication (Hartwell 1976). In *S. cerevisiae,* Cdc6 has been shown be cell cycle dependent; it is transcribed only in late G_1 and early S-phase (Piatti, Lengauer, & Nasmyth 1995), after which it is rapidly degraded by a Cdc4/34/53 ubiquitin-mediated pathway (Drury, Perkins, & Diffley 1997). Beyond cell cycle dependent regulation of Cdc6 availability, its activity is also regulated by ATP binding and hydrolysis. Cdc6 (named Cdc18 in *S. pombe*) is closely related to Orc1 in primary amino sequence - Figure 1-11 (Duncker *et al.* 2009), possessing the characteristic Walker A and B motifs typical of the AAA^+ family of ATPases, which have been shown to be involved in ATP binding or hydrolysis,thereby regulating Cdc6 activity (Speck & Stillman 2007; Iyer *et al.* 2004). In *S. cerevisiae,* mutation of the Cdc6 Walker A motif (K114A) is lethal. However, the mutant is still able to bind chromatin, but prevents Mcm2-7 recruitment, suggesting that ATP hydrolysis by the Walker A motif mediates structural changes that lead to helicase binding (Weinreich, Liang, & Stillman 1999). In the same study a double alanine mutation in the Walker B motif [DE(223,244)AA] was fully functional (Weinreich *et al.* 1999). In humans (Herbig, Marlar, & Fanning 1999), mutations in the Walker A motif (K208A) and Walker B motif (E285Q) have been described, and similar mutations in *Xenopus* (Frolova *et al.* 2002)[Walker A mutation (K202E) and Walker B mutation (E277G)], all supporting an essential role of ATP binding and hydrolysis by Cdc6 in these organisms. Based on primary sequence homology comparison, the C-terminus of Cdc6 is predicted to have a winged helix-turn-helix domain capable of binding DNA, as observed for Orcs1-5 (Duncker *et al.* 2009), although evidence that this domain directly binds DNA is still lacking (Kawakami & Katayama 2010). Rather, a KRKK motif

(amino acid residues 29-32) present in *S. cerevisiae* Cdc6 has been shown by site-directed mutagenesis to be essential for DNA recognition *in vivo* and *in vitro* (Feng *et al*. 2000). The collective function of ORC and Cdc6 is to load a pre-formed heptameric complex of Cdt1 and Mcm2-Mcm7 complex onto origin DNA (Tsakraklides & Bell 2010).

1.6.3.3 Cdc10-dependent transcript 1 (Cdt1)

A third factor involved in the formation of pre-RC is Cdt1, essential for recruitment of the MCM helicase (Tsakraklides & Bell 2010; Chen *et al*. 2007). Cdt1 was originally identified in *S. pombe* in a screen to isolate novel targets of the *Cdc10* gene, which encodes a cell cycle regulated DNA binding protein (Hofmann & Beach 1994). Like Cdc6 (Cdc18 for *S. pombe*), Cdt1 in *S. pombe* is transcribed in a cell cycle dependent manner (by Cdc10-dependent transcriptional regulation), with expression peaking at G_1/S transition and absent from the rest of the cell cycle (Nishitani *et al*. 2000; Hofmann & Beach 1994). Homologues of Cdt1 exist in higher eukaryotes, including *Drosophila* (named double parked or Dup) (Whittaker, Royzman, & Orr-Weaver 2000), *Xenopus* (named XCDT1) (Maiorano, Moreau, & Mechali 2000), humans (Wohlschlegel *et al*. 2000), and *S. cerevisiae* (also known as Tah11/Sid2) (Tanaka & Diffley 2002). Sequence similarity between Cdt1 eukaryotic homologues is very low, and hence a single BLAST search using *S. pombe* Cdt1 alone was insufficient to confidently identify a budding yeast homologue (Tanaka & Diffley 2002). Rather, a three-way multiple sequence alignment of a weak candidate from BLAST searches showed that Cdt1 from budding yeast and fission yeast were 32 % similar and 18 % identical in amino acid sequence, while both proteins were only 12 % identical to the *Xenopus* homologue (Tanaka & Diffley 2002).

The primary structure of Cdt1 lacks motifs associated with enzymatic properties. Thus, Cdt1 has been suggested to function as an adaptor protein that holds the MCM helicase and the ORC-Cdc6 complex together (Tada 2007). Functionally, Cdt1 has a poorly conserved N-terminal consensus motif with a phosphorylation site possibly recognised by CDK for regulation of its activity (Senga *et al*. 2006; Sugimoto *et al*. 2004). Like Cdc6, in mammalian cells Cdt1 is degraded at late S-phase and G_2, when CDK activity is high and where it is phosphorylated and degraded by an Skp2-dependent mechanism (Sugimoto *et al*. 2004). To prevent

re-initiation of DNA-synthesis, the activity of Cdt1 is negatively regulated by an interacting factor (Geminin), whose expression increases at late S and G_2 phase of the cell cycle, binding Cdt1 and blocking its interaction with MCM, thereby preventing re-licensing (Tada *et al*. 2001; Wohlschlegel *et al*. 2000; McGarry & Kirschner 1998). Although Cdt1 has been shown to interact with Cdc6 in mammalian and yeast cells (Cook, Chasse, & Nevins 2004; Nishitani *et al*. 2000), there is still controversy in the literature as to whether Cdt1 binds directly to chromatin, or is able to trigger the recruitment of Mcm2-7 with or without Cdc6 in *Drosophila* and *Xenopus* extracts (Maiorano *et al*. 2000; Whittaker *et al*. 2000). However, more recent studies in *Xenopus* suggest that Cdt1 associates with chromatin and only in the presence of Cdc6 can it mediate loading of the helicase Mcm2-7 for DNA licensing to occur (Tsuyama *et al*. 2005). Other studies have also confirmed that all three components (ORC, Cdc6 and Cdt1) are essential for loading of the helicase to form the pre-RC (Randell *et al*. 2006). Even more recently, in *S. cerevisiae,* Cdt1 has been shown to play a direct role in recruiting the MCM complex via its interaction with specific motifs (N-terminus residues 1–185 and C-teriminus residues 270–435) of Orc6 (Chen *et al*. 2007).

1.6.3.4 Minichromosome maintenance (MCM) proteins

The Mini Chromosome Maintenance (MCM) proteins were first identified in *S. cerevisiae* mutants which were defective in the maintenance of mitotically stable circular plasmids with a functional ARS and centromeric sequence (Maine, Sinha, & Tye 1984). Subsequently, it was independently shown that mutations in *mcm2* affected the replication of a plasmid but not its segregation (Maiti & Sinha 1992). In all eukaryotes studied so far, the prime candidate for duplex DNA unwinding during initiation and progression of DNA replication is a complex composed of six closely related paralogues, named Mcm2 through Mcm7 (Mcm2-Mcm7) (Liu, Richards, & Aves 2009b; Tye 1999). Functional MCM in eukaryotes forms a hetero-hexameric complex involving all six proteins and this organisation is conserved in all eukaryotes examined (Bochman & Schwacha 2009; Liu *et al*. 2009b; Forsburg 2004). However, various MCM subcomplexes have been detected, suggesting that formation of the hexameric complex occurs in a stepwise fashion; for a comprehensive review see (Bochman & Schwacha 2009). Using a chromatographic fractionation assay in *Xenopus laevis* egg extracts, it

has been shown that Mcm4, Mcm6 and Mcm7 are first assembled into a stable hexameric complex (two trimers), which then binds two Mcm2 homodimers to form two tetramers of Mcm2/4/6/7. Each of the Mcm2/4/6/7 tetramers then interacts with a heterodimer of Mcm3/5 to form the complete heterohexameric complex (Prokhorova & Blow 2000). In human cells, a stable trimeric subcomplex of Mcm4/6/7, with a loose interaction to Mcm2, forms the same tetramer observed in *Xenopus* (Schulte *et al*. 1996; Musahl *et al*. 1995). Mcm2/4/6/7 and Mcm3/5 subcomplexes have also been described in yeasts (both budding and fission), mouse, *Drosophila*, human cells and *Xenopus*, suggesting the mechanism of hexamer formation is also conserved (Schwacha & Bell 2001; Prokhorova & Blow 2000; Ishimi 1997; Schulte *et al*. 1996; Musahl *et al*. 1995; Kimura *et al*. 1995). *In vivo* and *in vitro* studies from various eukaryotes have demonstrated that the dimeric subcomplex comprised of Mcm4/6/7 forms a ring-shaped structure, with a high level of 3′ → 5′ DNA-dependent helicase activity, and is capable of binding single-stranded DNA (Sato *et al*. 2000). Addition of Mcm2, Mcm3, and Mcm5 modulates this helicase activity, suggesting that Mcm4/6/7 is the functional helicase subcomplex while Mcm2, Mcm3, and Mcm5 play a regulatory role within the heterohexamer (Kanter, Bruck, & Kaplan 2008; Schwacha & Bell 2001; Lee & Hurwitz 2000; Sato *et al*. 2000; Schulte *et al*. 1996).

It has been suggested that MCM is loaded onto ORC and Cdc6 as a hexameric ring composed of Mcm2-Mcm7 and this loading event, mediated by Cdt1, is regulated by hydrolysis of ATP by Cdc6 (Randell *et al*. 2006; Machida, Hamlin, & Dutta 2005). *In vitro* experiments using *S. cerevisiae* support this model and further propose that Mcm3 and Mcm5 are stabilised in the hexameric complex by Cdt1 (Tsakraklides & Bell 2010). A stable interaction between ORC and MCM has not previously been described in unmodified eukaryotes; only when Cdt1 is tethered to Orc1-5 has this interaction been described (Chen *et al*. 2007). In *S. cerevisiae,* once the MCM complex is recruited onto chromatin-bound ORC and Cdc6, ORC and Cdc6 can be washed from the complex by high salt, suggesting that MCM is the key licensing factor that is needed to prime pre-RC for DNA elongation (Donovan *et al*. 1997). This licensing is regulated by post-translation modifications, such as phosphorylation of bound MCM subunits [for example the Mcm4 subunit (Sheu & Stillman 2010)] by Dbf4–Cdc7 kinase (DDK, Dbf4-dependent protein kinase). The regulatory pathways that are involved in

replication licensing post pre-RC formation is beyond the scope of this thesis, and hence are not covered.

1.7 DNA replication initiation in archaea

In archaea the DNA replication initiation machinery is fundamentally related to that of Eukarya, indicative of their descent from a common (and presumably archaeal-like) ancestral machinery (Dionne *et al*. 2003). Eukaryotes typically have large numbers of linear chromosomes, each of which possess multiple replication origins (see above). In contrast, in archaea with circular chromosomes, some have single replication origins [e.g. *Pyrococcus abyssi;* (Matsunaga *et al*. 2001)] and others have three origins [e.g. *S. solfataricus* and *Sulfolobus acidocaldarius;* (Lundgren *et al*. 2004)]. In both systems, licensing of a replication initiation site involves the loading of a series of different factors. ORC in all archaea studied (except three methanogenic archaeal species where the protein is not been identified) is not a multi-component complex as observed in eukaryotes (see above), but is instead composed of a protein that is related to both eukaryotic Orc1 and Cdc6 (referred to as Orc1/Cdc6) - Figure 1-12 (Barry & Bell 2006). Some archaeal species encode a single Orc1/Cdc6 protein (e.g members of the *Pyrococcus spp*), while others encode greater numbers: for example, two for *Aeropyrum pernix* and *Methanobacterium thermautotrophicum*; three for *Sulfolobus Spp;* five for *Natronomonas pharaonis*; seventeen for *Haloarcula marismortui;* and none for *Methanococcus jannaschii, M. maripaludis, and M. kandleri* (Barry & Bell 2006). It remains unknown what the initiator protein is in these latter organisms that do not encode an ORC or a Cdc6 homologue (Barry & Bell 2006).

Archaeal Orc1/Cdc6s that have been characterised bind specifically to origins of replication *in vivo* and *in vitro* and, in some cases, as observed in *Solfolobus spp* (similar to bacteria for *dnaA* gene locus and *OriC*), the origin is located in close proximity to the *orc1/cdc6* gene (Duggin *et al*. 2008; Duggin & Bell 2006; Matsunaga *et al*. 2001). More recently, the structure of Orc1/Cdc6 from *Sulfolobus solfataricus* bound to a replication origin has been determined to 3.4 angstrom–resolution, showing DNA distortion and protein conformation changes that are likely to allow the recruitment of further factors, suggesting they

designate replication origins in a manner related to their eukaryotic orthologues (Dueber *et al*. 2007). Like eukaryotic ORC proteins, archaeal Orc1/Cdc6 proteins have a C-terminal winged helix domain, which is conserved for a number of archaeal species (Singleton *et al*. 2004) and functions as a DNA binding domain, as observed from structural and mutational analysis in *Pyrobaculum aerophilium* (Liu *et al*. 2000). The N-terminal domain contains an AAA^+ ATPase domain indicative of a regulatory role imposed by ATP binding and hydrolysis; a phenomenon that is observed as structural conformational changes induced when Orc1/Cdc6-2 (the second Orc1/Cdc6 protein) of *A. pernix* is crystallised with ADP (Singleton *et al*. 2004). The downstream factors in archaea that constitute the pre-replication complex also appear simpler than those in eukaryotes.

Cdt1 is a critical factor in eukaryotic pre-replication complex formation, functioning with ORC and Cdc6 to recruit the MCM helicase (see above). Very recently, a factor in archaea, named WhiP (*w*inged *h*elix *i*nitiator *p*rotein), has been identified that bears 34 % sequence similarity and 21 % identity to the C-terminal region of *S. cerevisiae* Cdt1 (Robinson & Bell 2007). Further bioinformatic analyses and alignment with four other archaeal homologues revealed the presence of two highly conserved N-terminal and C-terminal regions which consist of a winged helix-turn-helix (wHTH) DNA binding domain reminiscent of the RepA plasmid initiator protein from the bacterium *Pseudomonas* (perhaps suggesting the likely evolutionary descent of Cdt1) (Robinson & Bell 2007). As mentioned above, *S. solfataricus* has three origins, two of which are located upstream of the *orc1/cdc6-1* and *orc1/cdc6-3* genes. Following identification of the WhiP protein, Duggin *et al* (2008) found that the third origin was located in close proximity to *whip* locus (Duggin *et al*. 2008).

The presumptive DNA unwinding element, MCM, observed to be a heterohexameric protein in eukaryotes, is a homohexamer in archaea. Unlike eukaryotic MCMs, all archaea have at least a single MCM homologue (for example in *Pyrococcus spp* and *Solfolobus spp*), although some have more than one MCM homologue (for example two in *Methanopyrus kandleri* and three in *Haloarcula marismortui*) (Barry & Bell 2006). Structural studies on archaeal MCM proteins have produced diverse results, depending on the species studied. For example, *in vitro* characterisation of *S. solfataricus* MCM showed that the protein forms

homohexamers in solution (Carpentieri *et al*. 2002), while electron microscopic reconstruction studies of *Methanobacterium thermoautotrophicus* have shown a ring-shaped dodecameric complex (Kasiviswanathan *et al*. 2004; Chong *et al*. 2000). Like eukaryotic MCM, the archaeal MCM proteins also possess 3′ → 5′ DNA-dependent helicase activity and bind both single-stranded DNA and double-stranded DNA (De Felice *et al*. 2003; Chong *et al*. 2000). Direct interactions have been reported for archaeal MCM and Orc1/Cdc6 in *S. solfataricus* by gel filtration analyses (De Felice *et al*. 2003) and in *M. thermautotrophicus* by yeast two-hybrid analysis (Shin *et al*. 2003), and by pull-down and co-immunoprecipitation experiments (Kasiviswanathan, Shin, & Kelman 2005). From the results above, it was suggested that Orc1/Cdc6 may function as a helicase loader in the absence of a clear Cdt1 archaeal homologue (De Felice *et al*. 2003).

1.8 DNA replication initiation in *T. brucei*

In *T. brucei* previous work on DNA replication have examined sequences likely to function as replication origins (see above) rather than consider initiator factors that are required to carry out the process. With the publication of the genome sequence for *T. brucei* in 2005 (Berriman *et al*. 2005), it is now possible to use web-based algorithms to predict the function of *T. brucei* genes based on amino acid sequence homology with those of other organisms. Surprisingly, based on *in silico* searches, kinetoplastids (though eukaryotes) appear to possess an archaeal-like, simplified machinery for origin designation. Notably, *T. brucei*, *T. cruzi* and *Leishmania* contain a single identifiable ORC homologue, which appears to be related to both eukaryotic Orc1 and Cdc6 and is thus predicted to provide the functions of both Orc1 and Cdc6 (El Sayed *et al*. 2005b), as reported for archaea (Robinson & Bell 2005). Given this, the protein has been named ORC1/CDC6 (named hereafter TbORC1/CDC6 for the *T. brucei* protein), following the archaeal nomenclature.

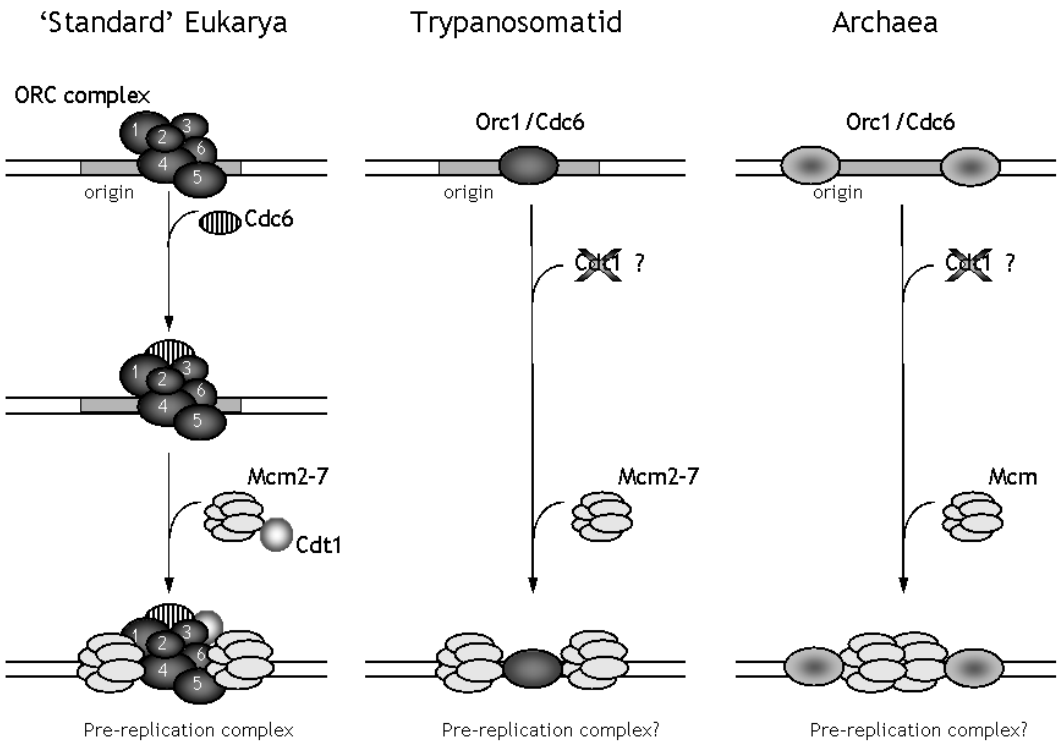


Figure 1-12 – Comparison of the DNA replication initiation factors in Eukarya, Trypanosomatids, and Archaea

A comparison of the identified factors that generate the DNA pre-replication complex, and their presumptive order of action, in those eukaryotes that have been studied ('standard'), trypanosomatids (e.g. *T. brucei*) and in the archaea. Homologues of eukaryotic Cdt1 have not been identified in any archaea or in the trypanosomatids (cross); the replicative helicase, MCM, is a hexamer composed of 6 distinct polypeptides in eukaryotes, which are conserved in the trypanosomatids, whilst it is normally a homomeric double or single hexamer in the archaea

In support of this nomenclature, *T. brucei* ORC1/CDC6 appears to be as closely related to its archaeal relatives as to eukaryotic Orc1 and Cdc6 proteins (Figure 1-13A), suggesting it assumes both roles. It also lacks the N-terminal extensions found in the Orc1 and Cdc6 proteins of other eukaryotes, instead possessing the N-terminal AAA^+ domain and C-terminal winged-helix-turn-helix architecture characteristic of the archaeal proteins (Liu *et al.* 2000) (Figure 1-13B).

In addition to the finding of only ORC1/CDC6, no clear homologue of Cdt1 could be found by BLAST searches of the kinetoplastid genomes. Though this may not be surprising, given the sequence divergence of Cdt1 in those eukaryotes in which it has been characterised, this may lend support to the idea that the very early steps of replication intitiation in *T. brucei* is archaeal-like. Recently, what appeared to be a distant homologue of archaeal WhiP was identified by bioinformatic searches of the kinetoplastid genomes (gene ID Tb927.6.5070,

referred to here as TbWhiP/CDT1). However, TbWhiP/CDT1 functional analysis by RNAi and BrdU incorporation to assess *in vivo* the effect of knockdown on DNA replication showed that this protein is unlikely to be involved in DNA replication in *T. brucei*. These data, being negative, are not presented in this thesis, and the protein is not considered further, except for in one experiment (see Section 3.3.6).

The absence of homologues of Orc2-Orc6 has not been described for any other eukaryotic organism, raising numerous mechanistic and evolutionary questions. If, indeed, these eukaryotic parasites utilise an archaeal-like, streamlined machinery (Figure 1-12) for replication origin licensing, this would be wholly novel for a eukaryote. However, this is complicated by the fact that the MCM helicase in kinetoplastids looks conventionally heterohexameric, with all six components readily identified bioinformatically. Nevertheless, a recent publication (Godoy *et al*. 2009) has begun to examine this question, and showed that *T. brucei* ORC1/CDC6 can complement an *S. cerevisiae cdc6* mutant, providing evidence for CDC6-related activity and support for hypothesis that the trypanosome pre-replication machinery may be similar to the archaeal system.

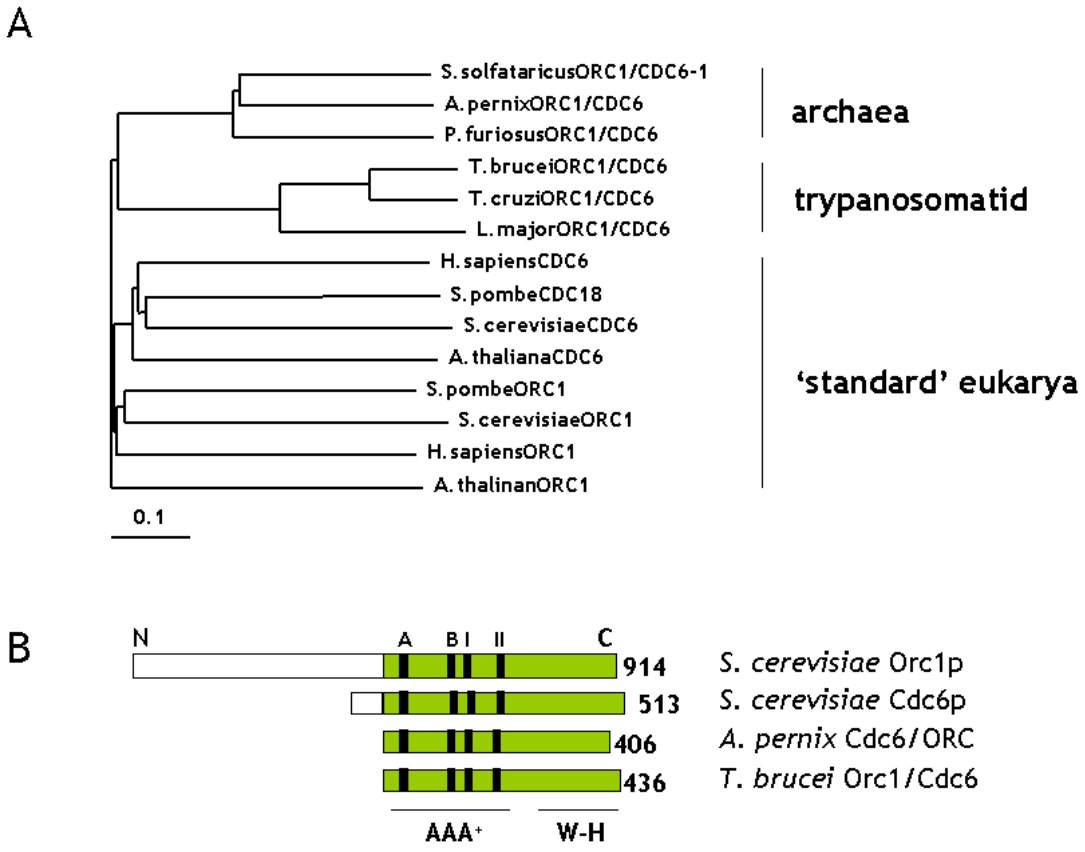


Figure 1-13 – A phylogenetic comparison of complete Orc1 and Cdc6-related proteins

A phylogenetic comparison of complete Orc1 and Cdc6-related proteins in those eukaryotes that have been studied ('standard'), trypanosomatids and in the archaea; the length corresponding to 10 amino acid (aa) changes per 100 residues is indicated. **B.** A diagram of sequence homology between Orc1 and Cdc6 proteins from *S. cerevisiae*, and Orc1/Cdc6 proteins from *A. pernix* and *T. brucei* (numbers indicate lengths in aa). Regions of homology are boxed grey; Walker A and B boxes, which define the AAA+ domain along with sensor motifs I and II, are shown as black bars and the the C-terminal winged-helix domain (W-H) involved in DNA binding is indicated. N-terminal sequence not conserved in trypanosomatid or archaeal proteins is in white

1.9 Project aims

To the best of our knowledge, at the outset of this project, the process of nuclear DNA replication initiation in *T. brucei* and related kinetoplastids (*L. major* and *T. cruzi*) had not been studied. Our overall objective was therefore to elucidate the mechanism and machinery of nuclear DNA replication in *T. brucei*, and, as you will see, we have begun to develop a clearer picture of the process than was previously understood.

1.9.1 Specific aims

This project had three specific aims:

1. To characterise TbORC1/CDC6 function in *T. brucei* (Chapter 3)
2. To identify TbORC1/CDC6 molecular partners (Chapter 4)
3. To identify and dissect replication origins in *T. brucei* (Chapter 5)

To address Aim 1, Chapter 3 describes RNAi analysis of *T. brucei* ORC1/CDC6 function that was performed prior to the publication of Godoy *et al* (2009), and will discuss similarities and differences in the findings. Using the *TbORC1/CDC6* RNAi lines, non-replication functions of TbORC1/CDC6 are also presented.

To address Aim 2, Chapter 4 describes the generation of functional epitope-tagged TbORC1/CDC6, and its application in co-immunoprecipitation experiments coupled with mass spectrometry to uncover TbORC1/CDC6 binding partners.

To address Aim 3, Chapter 5 describes the application of the functional epitope-tagged TbORC1/CDC6 to identify TbORC1/CDC6 binding sites by chromatin immunoprecipitation coupled with microarray technology.

2 Material and Methods

2.1 Trypanosome Strains, Cell Culture, & Genetic Modification

2.1.1 T. brucei strains

In this thesis, three strains of *T. brucei* and their genetically modified derivatives were used: Lister427, TREU (Trypanosomiasis Research Edinburgh University) 927, and EATRO (East African Trypanosomiasis Research Organisation) 795. For RNA interference (RNAi), *T. brucei* procyclic form strain 427 pLew29-pLew13, bloodstream form strain 427 pLew90-pLEW13, or procyclic form 795 pLew13-pLew29 were used. The pLew90-pLEW13 or pLew29-pLew13 parasite lines developed by Wirtz *et al* (Wirtz *et al*. 1999) constitutively co-expresses T7 RNA polymerase (T7RNAP) and the Tet repressor (TetR). Integration of the T7RNAP and TetR constructs were selected, respectively, by neomycin (G418) and hygromycin B selectable markers. Where necessary, selection of RNAi parasite lines expressing the T7RNAP, TetR and the RNAi construct was achieved by addition of hygromycin B, G418, and zeocin (for procyclic form cells) or phleomycin (for bloodstream form cells). Apart from RNAi experiments, the TREU927 wild type cell line was used for all other experiments except where specific genetically modified lines are indicated to have been used.

2.1.2 Procyclic-form cell culture

The term “procyclic form” (PCF) refers to the replicative tsetse fly midgut stage of *T. brucei*. *In vitro,* PCF *T. brucei* cultures were grown at 27 °C in a derivative of semi-defined medium (SDM-79, see recipe below) (Brun 1992), supplemented with 10 % (v/v) heat inactivated foetal calf serum (FCS) and 0.2 % (v/v) haemin [diluted from a 2.5 mg.mL^{-1} in NaOH (Gibco)]. Typically, PCF cells were maintained by a 1:10 dilution every 3 – 4 days of a culture that was usually at a density of ~8 x 10^{6} cells.

SDM79 recipe (5 L)

One sachet SDM79	Cat# 07490916N, Gibco
NaHCO3 (BDH Ltd)	24 mM final concentration
Distilled H2O	up to to 5 L
pH adjusted with 5 M NaOH	7.3

2.1.3 Bloodstream-form cell culture

'Bloodstream form' (BSF) refers to the replicative stage of *T. brucei* that resides extracellularly in the blood and tissue fluids of mammals. *In vitro,* BSF cultures were grown in petri dishes at 37°C, 5 % CO_2 in a humified incubator (Heraeus Instruments). BSF cells were grown in HMI-9 medium supplemented with 20 % (v/v) heat inactivated FCS; see recipe below (Hirumi & Hirumi 1994). Except where stated, BSF cells were harvested at densities of 3 – 5 x 10^6 cells.mL^{-1} for all experiments or for cryo-preservation. BSF cultures were maintained by a 1:100 dilution of a culture of density 3 – 5 x 10^6 cells.mL^{-1} every 3 – 4 days.

HMI-9 recipe (5 L)

One sachet HMI9 powder (Gibco)	Cat# 07490915N, Gibco
$NaHCO_3$ (BDH Ltd)	36 mM final concentration
ß-mercaptoethanol (Sigma)	1.4 %
Penicillin/Streptomycin (Cat# 15140, Gibco)	50 mL
Distilled H2O	up to 5 L

2.1.4 Routine growth and maintenance of trypanosomes

Cell density was determined microscopically using an improved Neubauer haemocytometer (Weber Scientific). In a 10 µl aliquot, the number of parasites under a 1 mm square area was counted. This was repeated for three independent 1 mm squares and the average of three counts was multiplied by 10^4 to get the number of cells per millilitre of culture. For long-term preservation, *T. brucei* cells were cryo-frozen in liquid nitrogen with unique identification numbers, known locally as "stabilate numbers". Log-phase PCF or BSF cells were supplemented with 10 % glycerol in SDM79 or HMI-9 medium, respectively, aliquoted in 1 ml volumes into cryo-vials, wrapped with cotton wool and frozen overnight at – 80 C. After the overnight period, the labelled cryo-vials were transferred to liquid nitrogen at a specific location and electronically recorded into a database. Recovery of stabilates was achieved by gradually thawing to room temperature (RT) and supplementing with 5 ml of warm medium without any drug selection. After a few days (2 – 3 days) in culture, the medium was supplemented with drugs and lines re-tested for any previous modification before further analyses.

2.1.5 Stable transfection of T. brucei procyclic form cells

All centrifugation steps were performed at 600 x g. PCF cells at a density of 5-10 x 10^6 cells.mL^{-1} were centrifuged for 10 min at RT and the supernatant removed and preserved for use as "conditioned medium". ~ 2.5 x 10^7 cells were resuspended in 0.5 ml of ice-cold Zimmerman medium ('ZMnoG'; 132 mM NaCl, 8 mM KCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 0.5 mM $MgAc_2$, and 0.06 mM $CaAc_2$, pH 7.5), known as electroporation buffer. 10 µg of linearised DNA (prepared as described below), in a maximum volume of 10 ul of sterile double-distilled water, was added to the cells resuspended in ZMnoG. After mixing to homogenise, two pulses were delivered using a Bio-Rad Gene Pulser II set at 1.5 kV and 25 µF capacitance. After electroporation, the delivered voltage was usually ~ 1.40 to 1.60 kV and a time constant of ~ 0.30 to 0.50 ms was recorded. The cells were then transferred into 10 ml of pre-warmed SDM79 medium and incubated at 27 ºC overnight (usually 12 – 24 hrs) to allow recovery before being subjected to drug selection. After overnight recovery, cells were selected in

medium supplemented with antibiotics at concentrations described below (Table 2-1).

Table 2-1 - Antibiotics used for selection of genetically modified parasites

Antibiotics	Stock concentration	Final concentration
Phleomycin (Cayla)	20 $mg.mL^{-1}$	2.5 $\mu g.mL^{-1}$ (BSF)
Zeocin (Invitrogen)	100 $mg.mL^{-1}$	10 $\mu g.mL^{-1}$ (PCF)
Hygromycin B (CALBIOCHEM)	50 $mg.mL^{-1}$	5 $\mu g.mL^{-1}$ (BSF) 50 $\mu g.mL^{-1}$ (PCF)
Neomycin or G418 (SIGMA)	50 $mg.mL^{-1}$	2.5 $\mu g.mL^{-1}$ (BSF) 10 $\mu g.mL^{-1}$ (PCF)
Blastidicin (CALBIOCHEM)	10 $mg.mL^{-1}$	10 $\mu g.mL^{-1}$ (BSF) 15 $\mu g.mL^{-1}$ (PCF)
Puromycin (CALBIOCHEM)	10 $mg.mL^{-1}$	0.2 $\mu g.mL^{-1}$ (BSF) 1 $\mu g.mL^{-1}$ (PCF)
Tetracycline (CALBIOCHEM)	5 $mg.mL^{-1}$	1-2 $\mu g.mL^{-1}$ (BSF) 1-2 $\mu g.mL^{-1}$ (PCF)

"Conditioned medium" was prepared by adding 10 % (v/v) FCS, 15 % (v/v) sterile filtered medium from a dense culture (up to 10^7 cells/ml), and 75 % (v/v) SDM79 medium supplemented with appropriate selection antibiotics. 100 ul and 1 ml of the population of electroporated and recovered parasites was added separately to 20 ml of conditioned medium and distributed in all welsl of a 96 well plate (175 µl in each well). The growth of antibiotic-resistant transformants was monitored by looking for wells containing growing cells 10 – 14 days later. The transformants were considered to be clonal if growing cells arose in 10 or less wells of the 96 well plates. This procedure was used for all PCF transfections except for antibiotic selection of clones which depended on the resistance marker present in the vectors used which are shown in Table 2-2 and the drug concentrations used are shown in Table 2-1.

DNA to be transfected was linearised using appropriate restriction enzymes as described in Section 2.3.3; except that this time the reaction was scaled up to a final reaction volume of 300 µl containing ~ 100 µg of DNA, the reaction mixture was incubated overnight, ethanol precipitated as described in Section 2.2.3 and resuspended in sterile double-distilled H_20. The DNA concentration was measured using a Nanodrop spectorphotometer (ThermoScientific).

2.1.6 Stable transfection of T. brucei bloodstream form cells

DNA preparation and all steps up to electroporation were essentially the same as described for PCF cells, except for the following. Transformation used the AMAXA Nucleofactor kit, optimised for Human T-cells according to the manufacturer's protocol (Cat# VPA-1002, Amaxa Biosystems). For each transfection, 1-2 x 10^6 cells.mL^{-1} BSF cells were used with 10 µg of DNA. After electroporation, the cells were serially diluted 1:10 in three consecutive tubes containing 30mL, 27mL, and 27 mL of HMI-9, without antibiotic. Each dilution was then distributed in 24 well plates (1.0 ml per well) and incubated at 37 ^{0}C for 6 – 12 hrs for recovery. After recovery, 1 ml of HMI-9 medium containing double the concentration of antibiotic used for selection of transformants (Table 2-1) was added to each well and incubated for 7 – 10 days at 37 ^{0}C. Plates in which at most 6 out of 24 wells contained growing *T. brucei* cells were considered clones and were used for further experiments.

2.2 General methods for DNA analysis

2.2.1 Primer design

A necessary first step in primer design for PCR in plasmid construction was retrieval of DNA sequences from genome sequence databases. DNA sequences were retrieved from TriTryDB (http://tritrypdb.org/tritrypdb/) using gene IDs for any gene of interest. All oligonucleotides used were designed using the primer design software supplied by Invitrogen® with the vectorNTI® package. Suitable restriction enzyme recognition sites were appended to the 5' end of each primer, along with a further four G residues at the 5' terminus. Also, restriction sites within the target gene were avoided and caution was taken to ensure in-frame fusion of the gene for any tags that were necessary in the final vector. All oligonucleotides were synthesised by Eurofins MWG Operon® (Germany). The sequences of oligonucleotides, destination vectors, and restriction sites used are shown in Table 2-2.

2.2.2 DNA Isolation

Plasmid purification from *Escherichia coli* and DNA extraction from agarose gels used the Qiagen Miniprep kit and the Qiagen Gel Extraction kit, respectively, following the manufacturer's protocols. Genomic DNA (gDNA) isolation from *T. brucei* cells was carried out using the Qiagen DNeasy Kit according to manufacturer's instructions.

2.2.3 Ethanol precipitation

To purify and concentrate DNA from solution, three volumes of ice cold ethanol (Fisher) was added to the DNA solution and, subsequently, sodium acetate (CH_3COONa, pH 5.2) (BDH) was added to a final concentration of 0.3 M. The mixture was incubated at - 80 ^oC for 30 minutes and then centrifuged at 20,000 x *g* for 30 minutes at 4 ^oC. The supernatant was carefully decanted and the pellet was washed twice with 1 mL 70 % ethanol. After the washes, the pellet was air dried and re-suspended in Tris-EDTA (TE) buffer (see recipe below) or distilled water depending on the subsequent application.

TE buffer

Tris-HCl 10 mM (pH 7.5)

EDTA 1 mM

2.3 Molecular cloning

2.3.1 Polymerase chain reaction (PCR)

A typical 50 ul PCR reaction was constituted as follows: 1 U of either *Taq* (New England Biolabs, NEB) or Phusion (NEB), 2.5 µl (10 µM stock) of each primer, the required buffer concentration (manufacturer's recommendation), 4 µl (2.5 mM stock) dNTPs, 10-50 ng of template DNA and distilled water (dH_20) to a final volume of 50 µl per reaction. DNA from *T. brucei* strain TREU927 was used as

DNA template for all cloning of ORFs and gene fragments, except for cloning of inserts for RNAi experiments where gDNA of the Lister427 strain was used. PCR amplification was carried out using a TC-3000 thermocycler (Techne), under the following PCR conditions: denaturation at 94 ^{o}C for 5 minutes, followed by 25 - 30 cycles of denaturation at 94 ^{o}C for 1 minute, annealing of primers at 55 ^{o}C for 1 minute and extension at 72 ^{o}C for 1 minute per kilobase (kb) amplified. A final extension time of 10 minutes at 72 ^{o}C was included at the end to ensure complete PCR amplification. After PCR cycling, samples were checked by running a 1 % agarose gel and the PCR product was cleaned from primer dimers and contaminating salt and enzymes using the PCR purification kit from Qiagen® according to manufacturer's instructions.

2.3.2 Agarose Gel Electrophoresis

DNA was separated on 1 % (w/v) agarose (Invitrogen) gels in 1X Tris Acetate-EDTA (TAE) buffer (40 mM Tris-HCl, 1 mM EDTA, pH 8.0, 0.11 % Acetic acid) containing 0.2 $\mu g.mL^{-1}$ ethidium bromide (Sigma) or 1:20,000 dilution of SYBRSafe (Invitrogen). Using a Bio-Rad PowerPac Mini, agarose gels were run at 100 – 120 V for 30 – 45 min depending on the sizes of the DNA fragments resolved. Visualisation was done with a U.V. transilluminator (Bio-Rad) and digitised using QuantityOne software v4.6 (Bio-Rad). A commercial 1 Kb size marker (NEB) was used to determine the size of DNA fragments resolved by agarose gel electrophoresis.

2.3.3 Restriction endonuclease digestion

All restriction enzyme digests for cloning were carried out as follows: e.g. 1 – 2 µl (typically 10 – 20 units. μl^{-1}) of appropriate enzyme (NEB), 1 - 10 µg total DNA, the required NEB buffer at the appropriate concentration (10X stock provided), and dH_20 to a final volume of 20 µl per reaction. In cases where DNA was being prepared for transfection of trypanosome cells, a total reaction volume of 300 µl was prepared, with buffer and enzymes scaled up appropriately. A standard 20 µl reaction was incubated at the optimum temperature required for the particular enzyme for 1 – 2 h. For DNA to be used

in transfection, incubation was performed overnight at the manufacturer's specified temperature of the restriction enzyme.

2.3.4 DNA ligation

To insert digested PCR products or DNA fragments into plasmids, DNA ligation was carried out. A typical 20 µl reaction contained the following: 2.5 units of T4-DNA ligase (NEB), molar ratios of vector to insert was typically 1:3 (typically 50 ng linearised dephosporylated plasmid vector to 150 ng digested insert DNA), 2 µl of 10X ligation buffer (NEB provided) and dH_20 to a final volume of 20 µl. The reaction mix was incubated at RT for 2 h prior to bacterial transformation.

2.3.5 Bacterial transformation

E. coli XL1-Blue MRF [Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)173 *endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F´ proAB *lacI*^{q}ZΔ*M15*]; Stratagene] competent cells were used for plasmid propagation. For ligation reactions, 5 µl was mixed with a 50 – 100 µl aliquot chemically competent *E. coli* XL1-Blue, purchased commercially from Stratagene. The DNA/*E. coli* mix was first incubated on ice for 30 minutes and then heat shocked at 42 ^oC for 40 seconds. 750 µl of LB (see recipe below) was added to the cells and the mix was incubated for 45 minutes at 37 ^oC in a shaker. After incubation, 200 µl of cells were plated onto antibiotic selective LB agar plates (see recipe below) with appropriate drugs for positive selection of plasmids (Section 2.4.1). LB agar plates were incubated at 37 ^oC overnight. Colonies were picked, screened by colony PCR (described below) and plasmids from positive colonies were sent for sequencing by the University of Dundee DNA sequencing service. Primers used for sequencing were usually the same primers used to amplify the gene from gDNA. After sequencing, contigs were assembled using the ContigExpress tool in Vector NTI Suite 10 (Invitrogen) or CLC genomics (CLC Bio).

Luria Bertani (LB) Medium		**LB Medium + Agar = LB Agar**	
Tryptone	10 g/L	LB Medium	
Yeast extract	5 g/L	Bacto-Agar	15 g/L
NaCl	10 g/L		

2.4 Selection of bacterial transformants

2.4.1 Drug selection

Positive colonies selected on LB agar plates were carried out at concentrations shown below:

Antibiotics	Stock concentration	Final concentration
Ampicillin (SIGMA)	100 $mg.mL^{-1}$	100 $\mu g.mL^{-1}$
Kanamycin (CALBIOCHEM)	30 $mg.mL^{-1}$	30 $\mu g.mL^{-1}$
Chloramphenicol (CALBIOCHEM)	34 $mg.mL^{-1}$	34 $\mu g.mL^{-1}$

2.4.2 Colony PCR

The PCR mix for bacterial colony screening and conditions were the same as the standard PCR described in Section 2.3.1, except for the following. Template DNA was obtained from a single colony picked from the LB agar plate and resuspended in 20 µl of dH_2O. Twenty-four colonies were picked for each round of screening and primers used for PCR-amplifying DNA fragments from gDNA were used for colony PCRs. After PCR, a 1 % (w/v) agarose gel was run alongside a positive control; where gDNA had been used as the template for the PCR and dH_2O as a negative control. Positive clones identified were grown overnight in LB

medium, plasmids were purified from overnight cultures and sent for sequencing.

2.5 Southern blot, hybridisation and detection

The NEBlot Kit protocol was used for Southern blot analysis (Cat No. N1500S, NEB). 5 - 10 µg of DNA was digested (as described in Section 2.3.3) overnight in a total volume of 100 µl with appropriate units of enzyme(s). After digestion, the DNA was ethanol precipitated, resuspended in a final volume of 20 µl, and loaded onto a 0.8 % agarose gel. The gel was run overnight at 30 V (350 m/s) in 1X TAE. The gel was then photographed under UV light alongside a ruler to allow calculation of the size of fragments after hybridisation and detection. After electrophoresis, the gel was incubated in 0.25 M HCl for 10 min with gentle agitation. The DNA in the gel was then denatured by soaking the gel in a denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 30 min. The gel was neutralised by soaking first in a neutralising solution (1 M Tris-HCl pH 8.0, 1.5 M NaCl, 1mM EDTA) for 30 min, and then into 20X SSC solution (3 M NaCl + 0.3 M Na Citrate). The DNA was transferred overnight to a nylon membrane (Hybond-N+, Amersham) by capillary blotting. After the transfer, the blotting apparatus was dismantled; the membrane was rinsed in 20X SSC for 5min, air-dried and UV-crosslinked twice to bind DNA to membrane (UV Stratalinker 2400, Stratagene).

PCR-amplification for generation of DNA probes used primers CT_OL9/CT_OL10 for the 5' UTR probe and primers CT_OL47/CT_OL48 for the TbORC1/CDC6 ORF probe. PCR-amplification was carried out as described in Section 3.1. The DNA probes were radiolabelled using the random priming method according to manufacturer's protocol (NEBlot protocol). After labelling, probes were purified by gel filtration on Sephadex® G-50, Spin Column Elutips® (GE Healthcare) to reduce non-specific background and limit exposure of lab personnel to high levels of radioactivity during hybridization experiments. Following purification, labelled DNA probes were prepared for hybridization as follows: denatured by heating in boiling H_2O bath 95–100°C for 5 minutes and quickly placed in ice bath for 5 minutes. The denatured DNA was either used directly in hybridisation experiment or stored at –20°C.

After crosslinking the DNA to membrane, the blot was pre-hybridised in 50 ml (to cover membrane) pre-warmed Church-Gilbert solution (0.342 M Na2HPO4, 0.158 M NaH2PO4.2H2O, 0.257 M SDS and 1 mM EDTA) for 2 h at 60 ºC. Then, 100ng of radiolabelled, purified, and denatured DNA probe was added and hybridisation was performed at 60 °C overnight in pre-warmed Church-Gilbert solution. After hybridisation, post hybridisation washes were carried as follows: the blot was washed twice at 60°C, initially for 5 min, then again for 15 min with: (a) 50 ml of 2X SSC + 0.1% SDS and (b) 50 ml 0.2X SSC + 0.1% SDS. The blot was then sealed in transparent cling film and visualised by exposure to an x-ray film in an autoradiography cassette at - 70°C for 15 min to 7 days.

2.6 RNA interference and analysis of samples

2.6.1 RNA isolation and analysis

2.6.1.1 Generation of cDNA and semi quantitative RT-PCR

Total RNA was prepared according to the manufacturer's protocol from the Qiagen® RNeasy mini kit. A total of 2 – 4 x 10^7 cells from cultures with densities between 1 - 2 x 10^6 cells.mL^{-1} (BSF) or 3 - 4 x 10^6 cells.mL^{-1} (PCF) was used for each RNA preparation. To digest any contaminating DNA, the RNA preparation was treated on column with RNase-free DNase as described by the manufacturer (Qiagen®).

For first strand cDNA synthesis, the SuperScript™ First-Strand Synthesis System for RT-PCR (Invitrogen) was used as described by the manufacturer. 1 µg of total RNA was used with 50 µM oligo (dT) primers, as per manufacturer's instructions. Identical reactions were set up for each cDNA sample with reverse transcriptase (RT) or without RT (to act as a control for gDNA contamination). PCRs were set up according to manufacturer's instructions (Invitogen) using 1 µl of cDNA/reaction and primers for *TbORC1/CDC6* (CT_OL3/CT_OL4, Table 2-2). The conditions for PCR were: 1 cycle of 95 °C for 5 min, either 26 or 30 cycles of 95 °C for 30 s, 55 °C for 30 s, 72 °C for 60 s, and a final 1 cycle of 72 °C for 10 min. *RNA Polymerase I largest subunit (Tb927.8.5090)* PCR-amplified using primers CT_OL5/CT_OL6 (Table 2-2) as a loading control.

2.6.1.2 Real-Time PCR

To quantify levels of mRNA knockdown, primers for real-time PCR of TbORC1/CDC6 (CT_OL7/CT_OL8, Table 2-2) were designed using Primer Express® software (Applied Biosystems). The SYBR® Green PCR Master Mix (Applied Biosystems) was used for PCR amplification in 96 well plates. A master mix for 30 reactions was made according to the following recipe: 12.5 µl of SYBR mix (provided in kit), 1.0 µl of each primer (300 nM stock), 9.5 µl of dH_2O, and 1.0 µl cDNA (from Section 3.8). *GPI8* primers (CT_OL27/CT_OL28) were used as an internal control, and dH_2O was used for non-template primer controls. The ABI Prism 7000 thermocycler conditions for all reactions were 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. Amplification plots generated were used for quantification of mRNA levels as described in the Applied Biosystems manual.

2.6.2 4, 6 - Diamidino-2-phenylindole (DAPI) Staining

For DAPI staining of *T. brucei* DNA, 5×10^5 cells were centrifuged at 600 x g, washed twice in PBS and resuspended in 100 µl of PBS. 50 µl was spread on a glass microscope slide, air-dried, and fixed for 5 min in methanol at RT. The slides were then removed from the methanol, which was allowed to evaporate at RT, and DAPI with Vectashield (VectorLabs) was added to the slide and spread by the addition of a coverslip. Slides were sealed with nail varnish and examined under UV light on a Zeiss Axioplan microscope. Images were captured using a Hamamatsu ORCA-ER digital camera and Openlab version 3.0.3 software was used for processing images.

2.6.3 Fluorescent Activated Cell Sorting (FACS) Analysis

For FACS analysis, 10^6 cells were pelleted by centrifugation at 600 x g, washed once in PBS, and resuspended in 70% methanol, 30% PBS. The cells were incubated at 4 °C overnight. After fixation overnight, cells were washed in 10 ml of ice cold PBS, resuspended in 1 ml of PBS containing 10 µg ml^{-1} propidium iodide (SIGMA) and 10 µg ml^{-1} RNase A (SIGMA), and incubated at 37 °C for 45 min. FACS was performed with a Becton Dickinson FACSCalibur using detector FL2-A and an AmpGain value of 1.75.

2.6.4 In vivo Bromo-2-deoxyuridine (BrdU) labelling, detection and quantitation

After RNAi for 96 hrs, 10^8 PCF *T. brucei* cells were labelled with 50 µM BrdU and 50 µM 2'-deoxycytidine in SDM-79 and incubated at 27°C for 60 mins. After incubation, the cells were harvested by centrifugation at 600 x g for 10 minutes, and total DNA was extracted using the Qiagen DNeasy Kit according to manufacturer's instructions. DNA samples were then incubated for 1 hr at 37 °C with 33 µg.mL^{-1} RNase A (Sigma R4642). The amount of purified DNA was determined using a NanoDrop (Thermo Scientific) and 2 µg of total DNA was incubated with 10 volumes of 0.4 N NaOH solution for 30 min at RT and kept on ice to prevent re-annealing. The DNA solution, kept on ice, was then neutralised with an equal volume of 1 M Tris·HCl (pH 6.8). The single-stranded, neutralized DNA was next dot-blotted (50 ng in 5 µl) onto a nitrocellulose membrane (Amersham) and allowed to air dry. The DNA was fixed twice on the membrane using an ultraviolet cross-linker Stratalinker (Stratagene). After the fixation step, the membrane was incubated with mouse anti-BrdU monoclonal antibody (1:2,000 dilution, B2531, Sigma) in buffer containing TBST (20 mM Tris·HCl, pH 7.6, 136 mM NaCl, and 0.05% Tween 20) containing 1 % non-fat milk for 1 hr at RT. After incubation with the primary antibody, the membrane was washed with TBS-T three times for 10 mins each at RT, and then incubated with horseradish peroxidase-conjugated anti-mouse IgG antibody (1:5,000 dilution) for 1 h at RT. After incubation with the secondary antibody, the membrane was washed three times again with TBS-T for 20 min at RT. Detection of the BrdU signal used an enhanced chemiluminescence detection system using the QuantityOne software (BioRad). Dot intensity on the membrane was quantified by measuring chemiluminescent signal relative a neutral spot on the membrane (background) using the same software.

2.7 Protein analysis

2.7.1 Preparation of whole cell T. brucei extracts

To prepare whole cell extracts from *T. brucei*, 10^8 cells were pelleted at 2000 x *g* for 10 minutes and washed twice in ice cold Phosphate buffered Saline (PBS).

The cells were then re-centrifuged and resuspended in 200 μl SDS lysis buffer [63 mM Tris HCl (pH 6.8), 10 % glycerol, 2 % SDS, 5 % ß-mercaptoethanol, 0.0025 % bromophenol blue (Sigma)]. This sample was ready for SDS-PAGE electrophoresis.

2.7.2 Sodium-dodecyl-sulphate–polyacrylamide gel electrophoresis (SDS-PAGE)

To separate proteins by electrophoresis, protein samples were loaded into precast polyacrylamide gels (Invitrogen) depending on the sizes of the proteins. Normally, the gels contained 10 % polyacrylamide, but for higher molecular weight proteins (usually > 175 kDa) 5 % gels were used. 20 µl of the *T. brucei* whole cell extract (above) was loaded and separated by electrophoresis at 200 V for 60 min. After electrophoresis, the gel was either stained with Coomassie Blue or colloidal coomassie for visualisation of all proteins, or specific proteins were detected by immunoblotting (Section 3.9.5) after transfer to a nitrocellulose membrane.

2.7.3 Coomassie blue staining of SDS gels

After electrophoresis, protein gels were visualised by Coomassie blue staining or in specific instances where low abundant proteins were visualised, colloidal Coomassie staining was used instead.

For Coomassie staining, gels were immersed in Coomassie stain solution [0.25 g Coomassie brilliant blue R [Sigma] in 90 ml of methanol: water [1:1 v/v] and 10 mls glacial acetic acid] and placed on a rocker for 1 – 4 hours. After staining, the gel was destained by immersing gel in a destaining solution [10 % glacial acetic acid, 40 % methanol] for 1 hour.

For colloidal Coomassie staining, protein gels were fixed for 1 hour in fixing solution (40% v/v methanol, 7% v/v acetic acid). Immediately before staining, gel staining solution was prepared by mixing and then vortexing 4 parts of the 1X working solution (1X Brilliant Blue G-Colloidal Coomassie, SIGMA) and 1 part methanol. The gel was then incubated in the staining solution overnight. After staining, the gel was placed in a destaining solution 1 (10% v/v acetic acid, 25%

v/v methanol) for 60 seconds while rocking. The gel was then rinsed with 25% v/v methanol, and destained again in 25% v/v methanol for up to 24 hours. To visualise bands, the gel was scanned at 600 nm wavelength using a trans-UV illuminator (BioRad). Bands were then excised from the gel and sent for mass spectrometry for protein identification.

2.7.4 Identification of proteins by liquid chromatography mass spectrometry [LC-MS/MS]

After colloidal coomassie staining, specific bands were excised directly from SDS gels, and sent to The University of Glasgow, Sir Henry Wellcome Functional Genomics facility for liquid chromatography mass spectrometry (LC-MS/MS). The resulting MS/MS were used to interrogate the TritryDB database using the MASCOT software. The MASCOT output resulting from the queries were used to analyse potential interacting partners. Potential interactors were further confirmed by co-immunoprecipitation experiments.

To identify TbORC1/CDC6 interacting partners, samples were treated as described above but this time the samples were sent to the University of Dundee proteomics facility for mass spectrometry. This is described further in Section 4.7.

2.7.5 Immunoblotting

After separation of proteins by SDS-PAGE, the proteins were transferred to nitrocellulose membrane (GE Healthcare) in transfer buffer [25 mM Tris pH 8.3, 192 mM Glycine and 20% v/v methanol] for acidic proteins or CAPS buffer (SIGMA) pH 11.0 for basic proteins. The transfer was carried out at 100 V for 70 minutes using the Mini Trans-Blot® Electrophoretic Transfer Cell (Bio-Rad). Blots were blocked for 1 hour with PBS containing 5% w/v powdered Milk (Marvel, Chivers Ireland Ltd) and 0.05% v/v Tween 20 (PBST). After blocking, blots were incubated with primary antibody in PBST for 1 hour at RT. Excess antibody was washed off with PBST three times for 5 mins each. Blots were then incubated for 1 hour with horseradish peroxidise (HRP) – conjugated secondary antibody (Invitrogen) in PBST. To detect protein, blots were washed 3 times in PBST and

detected by enhanced chemiluminescence (following the manufacturer's protocols; Piercenet) by exposure to an x-ray film. The x-ray film was developed using a Kodak Ex-omat system (KODAK).

2.7.6 Immunofluorescence

The immunocytochemistry protocol provided with anti-Myc Tag, clone 4A6, Alexa Fluor® 488 Conjugate antibody (Millipore) was used for IFA. Briefly, *T. brucei* cells were grown to a density of ~ 10^6 cells.mL^{-1}, the culture medium was removed after centrifugation at 600 x g for 10 mins and the cells washed twice with PBS. 1 mL of 3.7% (v/v) formaldehyde/PBS was added to the cells and the solution was incubated for 60 minutes at RT. The cells were washed twice with PBS for 5 mins each, resuspended in 1 mL of PBS and permeabilized by addition of Triton X 100 to a final concentration of 0.5 % (v/v) for 5 mins. The cells were washed, again twice, with PBS for 5 mins and incubated with 2 µg.mL^{-1} of anti-Myc Tag, clone 4A6, Alexa Fluor® 488 Conjugate antibody in 5 % BSA in PBS for 1 hr. TREU927 wild type (not TbORC1/CDC6-myc tagged) samples were incubated with the same antibody to serve as a negative control sample. After incubation with antibody, the cells were washed twice with PBS for 5 mins, pelleted, resuspended in 100 µl of PBS and spread evenly onto a glass slide (Menzel-glaser). The slide was then air-dried, counter-stained with DAPI (VectorLabs) and sealed with a cover slip using nail varnish. Cells on the slides were examined using an Axioskop 2 microscope and Openlab 3.00 software was used for analysis of images.

For *T. brucei* cells in which TbORC1/CDC6 had been tagged with Green fluorescent Protein (GFP) at the amino-terminus (see Section 3.5.1), these cells were harvested and fixed with formaldehyde as described above. After fixing with formaldehyde, 10 µl was smeared onto a glass slide, allowed to air dry, counter-stained with DAP and visualised as described above.

2.7.7 Co-immunoprecipitation (co-IP)

For co-IP, whole cell extracts from *T. brucei* were prepared. 10^8 cells were pelleted at 2000 x *g* for 10 minutes and washed twice in ice cold Phosphate buffered Saline (PBS). The cells were then re-centrifuged (2000 x *g* for 10

minutes) and resuspended in 2 mL of WCE buffer below. This was incubated on ice for 1 – 2 hours while magnetic beads were prepared as described below.

2.7.7.1 Preparing magnetic beads

50 µl of M-280 IgG-coated magnetic Dynalbeads (Invitrogen) was pre-blocked by washing the beads with 1 ml of block solution [0.05% Bovine Serum Album (BSA), PBS]. The beads were collected using a magnetic rack and the supernatant removed with a pipette. This step was repeated twice and the beads were resuspended in 250 µl block solution. 5 µg of antibody specific to the targeted epitope tag was added to the beads/block solution and incubated at 4°C overnight on a rotor. After incubation overnight, the beads were washed 3 times in 1 ml block solution and resuspended in 100 µl block solution. The antibody-beads were for ready for IPs.

2.7.7.2 Immunoprecipitation (IP), washes and elution

10^8 PCF cells were pelleted by centrifugation (600 x g), washed once in ice cold PBS pH 7.4 and lysed by incubating in 2 mL of Whole Cell Extract buffer (WCE buffer, see recipe below) at 4 ºC for 1 hr. The lysate was then centrifuged at 15000 x g at 4 ºC for 30 mins. 20 µl from supernatant was aliquoted; protein SDS loading dye was added to it and stored at – 20 °C. This served as the input sample for western blotting. The supernatant was added to the beads prepared in Section 2.7.7.1 and incubated for 2 hrs. After incubation, the samples were washed with 1.0 mL of ice cold wash buffer (see recipe below). The beads were collected and the supernatant was discarded. Wash steps were repeated seven times, after which the samples were washed with 1 mL of TE wash buffer (see recipe below). After the TE wash step, the samples were centrifuged for 3 mins at 1000 x g at 4 ºC and any residual TE wash buffer was carefully removed. 220 µl of Elution buffer (see recipe below) was then added to the samples and incubated in a 65°C water bath for 30 min with intermittent vortexing every 2-5 mins. The beads were then centrifuged for 1 min at 16,100 g at RT. 200 µl of the supernatant was removed after the spin step and transferred to new microcentrifuge tube. This served as the eluate. An aliquot of 20 µl was taken and used for western blot analysis (Section 2.7.5).

Solution	Stock solution concentration	Final Concentration
WCE buffer	1 M HEPES pH 7.55	50 mM
	1 M NaCl	100 mM
	0.5 M EDTA pH 8.0	1.0 mM
	0.5 M EGTA pH 8.0	1.0 mM
	50 % glycerol	10 %
	10 % Triton X100	1 %
	*50 X complete protease inhibitor mix	1 X
Wash buffer	1 M HEPES pH 7.55	50 mM
	1 M LiCl	500 mM
	0.5 M EDTA pH 8.0	1.0 mM
	0.5 M EGTA pH 8.0	1.0 mM
	10 % Na deoxycholate	0.7 %
	10 % NP40	1 %
	*50 X complete protease inhibitor mix	1 X
TE Wash Buffer	50 mM Tris pH 8	10 mM
	10 mM EDTA pH 8.0	1.0 mM
	1M NaCl	50 mM
Elution buffer	1M Tris pH 8	50 mM
	0.5M EDTA pH 8.0	10 mM
	10% SDS	1%

2.8 Chromatin immunoprecipitation (chIP)

ChIP protocol was obtained from the G.A.M. Cross Lab (Rockerfeller Institute, New York). 10^8 PCF cells were used for each ChIP experiment. To prepare cells for cross-linking, the concentration of cells was adjusted thus: 1.0 X 10^8 PCF cells/40 mL in SDM79 medium. These cells were cross-linked by incubating them for 20 min at RT in formaldehyde solution (see recipe below). After cross-linking, Glycine was added to the mix at a final concentration of 125 mM and centrifuged at 4000 x g for 20 min at 4 °C. The cells were then washed with 30 mL ice cold PBS and centrifuged again at 4000 x g for 20 min at 4 °C. After the wash step, cells were resuspended in Lysis buffer 1 (see recipe below), vortexed

thoroughly and rocked on a platform rocker for 10 min. After rocking, the cells were centrifugedat 4000 x g for 20 min at 4 °C and the supernatant was decanted. The previous procedure was repeated with Lysis buffer 2 (see recipe below) but this time the pellet was resuspended in 2 mL of Lysis buffer 3 (see recipe below). Prior to the IP step (below), the whole cell lysates prepared in this way were sonicated for 60 cycles (30 sec on/30 sec off) using a Bioruptor (Diagenode). After sonication, the material was centrifuged at 16,100 x g for 10 min at 4 °C. A 20 µl aliquot of the material was removed before the IP step to serve as the "input sample". IPs were then performed overnight with 10 µg of anti-Myc monoclonal antibody (Millipore) coupled to M-280 IgG magnetic Dynalbeads (Invitrogen). Beads were washed seven times with Wash buffer and eluted as in Section 2.7.7.2. The eluted DNA was prepared for sequencing or hybridisation to a microarray as described below.

After IP, the eluted DNA was reverse cross-linked from proteins. First, 3 volumes of Elution buffer (recipe above; Section 3.9.7.2) was added to the 'input sample' and mixed thoroughly. From here, the 'input sample' was treated like the eluted IP material. Both samples were reverse crosslinked by incubating at 65°C for ~ 9 hours (not more than 15 hours). After incubation, 8 µl of RNaseA (from a stock of 10 $mg.mL^{-1}$) was added to each sample and mixed by inverting the tubes several times and this was incubated at 37°C for 2 hours. Next, 4 µl of Proteinase K (from a stock of 20 $mg.mL^{-1}$) was added to each sample and mixed by inverting tube several times and incubated at 55°C for 2 hours. The DNA was then purified using a Qiagen gel extraction kit and eluted with 60 µl kit elution buffer, as recommended by manufacturer's protocol. The purified DNA was repaired (from nicks introduced during the sonication step) using the Quick Blunting kit (NEB) according to manufacturer's protocol. After the repair step, DNA was purified using the QIAquick PCR Purification Kit (Qiagen) according to manufacturer's instructions and eluted with 41 µl of dH_2O. The purified DNA was then amplified using the Whole Genome Amplification kit (SIGMA) and purified using PCR purification Kit from Qiagen®. 5 µl of the eluted DNA was run on a 1.5 % agarose gel, and the remaining material was quantified using a NanoDrop.

For microarray analysis, ~ 5 µg of input and eluted DNA was sent to Roche Diagnostics (NimbleGen, USA) for differential labelling and hybridisation to a custom-designed microarray. For Solexa sequencing, ~ 100 ng was sent to

GenePool at Edinburgh University, UK for sequencing. The ChIP-chip data was analysed by L. Marcello (University of Glasgow), a collaborator on the project. The ChIP-seq data is still being analysed by N. Dickens, the resident Bioinformatician at the department, and is not considered in this thesis.

Solution	Stock solution concentration	Final Concentration
Formaldehyde Solution	1 M HEPES pH 7.55	50 mM
	1 M NaCl	100 mM
	0.5 M EDTA pH 8.0	1 mM
	0.5 M EGTA pH	0.5 mM
	37% formaldehyde (Fisher)	11%
Lysis Buffer 1	1 M HEPES pH 7.55	50 mM
	1 M NaCl	140 mM
	0.5 M EDTA pH 8.0	1.0 mM
	0.5 M EGTA pH 8.0	1.0 mM
	50 % glycerol	10 %
	10 % Triton X100	0.25 %
	10% NP-40	0.5 %
	*50 X complete protease inhibitor mix	1 X
Lysis buffer 2	1 M Tris-HCl pH 7.8.0	50 mM
	1 M NaCl	200 mM
	0.5 M EDTA pH 8.0	1.0 mM
	0.5 M EGTA pH 8.0	1.0 mM
	*50 X complete protease inhibitor mix	1 X
Lysis buffer 3	1 M Tris-HCl pH 7.8.0	10 mM
	1 M NaCl	100 mM
	0.5 M EDTA pH 8.0	1.0 mM
	1 % Na-Deoxycholate	0.1 %
	10 % N-lauroylsarcosine	0.5 %
	*50 X complete protease inhibitor mix	1 X

Table 2-2 – Table of primers used throughout this thesis

Chapter 3				
Name	**Gene ID/Gene Description**	**Primer Sequence (5 '- 3')**	**Restriction site/Loci description**	**Final vector**
CT_OL1	Tb11.02.5110 /TbORC1/CDC6 RNAi	5' CCC**GGATCC**GAAGCCCACAGCTGTCTTTC	BamH1	pZJM_dual T7
CT_OL2		3' CCC**AAGCTT**TTCTCCGGCAACTTGTAACC	HindIII	
CT_OL3	TbORC1/CDC6 cDNA semi-qPCR	5' GGCTTGCTAACTGTAAAGGA		
CT_OL4		3' ATCCCCGTACCAAAGTCATC		
CT_OL5	Poll primers	5' GTGACGGTTCAGGGAACACT		
CT_OL6		3' GCACCGAAATTTGACTTGGT		
CT_OL7	TbORC1/CDC6_qPCR	5' TTCACCCTGTCATGCAGGTTT		
CT_OL8		3' GGTTCACTGACGCTGTCTTTCC		
CT_OL9	TbORC1/CDC6_ KO 5' UTR	5' CCC**GAGCTC**CATCCGCCTGGTTACAGCCA	SacI	pBlu2KS
CT_OL10		3' CCC**TCTAGA**TGGAGCACGAACACGGTAAA	XbaI	
CT_OL11	TbORC1/CDC6_KO 3' UTR	5' CCC**TTAATTAA**CCGTCTTAGGACGTGTGTGC	PacI	pBlu2KS
CT_OL12		3' CCC**CTCGAG**CCTTCTTTCGGCTTTGGCTT	XhoI	
CT_OL13	TbORC1/CDC6 +/- PUR integration test	5' CCCGTTTGGCGTGAGATATA		
CT_OL14		3' ATGTGGCGGTCCGGGTCGACG		
CT_OL15	TbORC1/CDC6 +/- BSD integration test	5' CCCGTTTGGCGTGAGATATA		
CT_OL16		3' GCCCTCCCACACATAACCAG		
CT_OL17	BSD Primers	5' GGCCAAGCCTTTGTCTCAAG		
CT_OL18		3' GCCCTCCCACACATAACCAG		
CT_OL19	TbORC1/CDC6_eGFP integration test	5' GGCCAAGCCTTTGTCTCAAG		
CT_OL20		3' CTTGAGCTCAGGCAGACTAT		
CT_OL21	MVSG 1.22 (GFP)_qPCR	5' CACAGACCTGCAGATGCACTTTAT		
CT_OL22		3' TGCCTTTATCTTTGCTAAATTTGCT		
CT_OL23	MVSG 1.61 (GFP)_qPCR	5' GGCGGTTTGTCTTTGTTTTTG		
CT_OL24		3' TGACTCCTCTTTGTTGTCGTCTTC		
CT_OL25	MVSG 1.64 (GFP)_qPCR	5' CTGCTGCCCGACAACCA		
CT_OL26		3' TGTGATCGCGCTTCTCGTT		

Chapter 3/4				
Name	**Gene ID/Gene Description**	**Primer Sequence (5 '- 3')**	**Restriction site/Loci description**	**Final vector**
CT_OL27	GPI8_qPCR	5' TCTGAACCCGCGCACTTC	832 bp downstream of start codon	
CT_OL28		3' CCACTCACGGACTGCGTTT		
CT_OL29	Actin-qPCR	5' CGGACGAGGAACAAACTGC	4 bp downstream of start codon	
CT_OL30		3' TTTCCATGTCATCCCAATTGG		
CT_OL31	Tubulin_qPCR	5' TTCAGGCTGGCCAATGCG	18 bp downstream of start codon	
CT_OL32		3' TACGGAGTCCATTGTACCTG		
CT_OL33	VSG221_qPCR	5' AGCAGCCAAGAGGTAACAGC	338 bp downstream of start codon	
CT_OL34		3' CAACTGCAGCTTGCAAGGAA		
CT_OL35	VSG13_qPCR	5' ATAACGCATGGCCATCTTGAC	572 bp downstream of start codon	
CT_OL36		3' GTCGTTGCTGTGGATTGCTC		
CT_OL37	VSGV02_qPCR	5' CAGCGCAAGTACAGGACG	393 bp downstream of start codon	
CT_OL38		3' TGCTTCGTCGTCGCTTAC		
CT_OL39	VSG224_qPCR	5' GACGCAGCAGAATCAACAC	1.1 kb downstream of start codon	
CT_OL40		3' GCTTATTTTGTGTCTGTCGC		
CT_OL41	VSG800_qPCR	5' ACAGACCGCCGACAGTATC	161 bp downstream of start codon	
CT_OL42		3' GTATCTTTGTAGGCCGCTGC		
CT_OL43	TbORC1/CDC6_12Myc	5' CCC**AAGCTT**GCATGGAACTGGCAGCTGAT	HindIII	pNAT-12Myc
CT_OL44		3' CCC**TCTAGA**TATCGAAAACAGTGGGCATATCCC	Xba1	
CT_OL45	TbORC1/CDC6-12Myc integration test	5' TGTCGTGATTCCCATGTTAA		
CT_OL46		3' AATGACGAACGGGAAATGCC		
CT_OL47	TbORC1/CDC6_ORF Probe	5' ATGAAACGAAGGAGGGACAG		
CT_OL48		3' AGACAGCTGTGGGCTTCGAC		
CT_OL49	6HA_Tag	5' CGATCCGGACCAAGT**TCTAGA**TACCCCTACGACGTT	Xba1	
CT_OL50		3' AAATGGGCAGGATCT**GGATCC**TTACCTTGAGGCATA	BamH1	

Chapter 4/5				
Name	**Gene ID/Gene Description**	**Primer Sequence (5 '- 3')**	**Restriction site/Loci description**	**Final vector**
CT_OL51	Tb11.01.3510 / MCM6_HA	5' CCC**AAGCTT**CCTGAGTGAGGATGTTGAAG	HindIII	pNAT-6HA
CT_OL52		3' CCC**TCTAGA**CTGCTTCGTAATATCTGGGT	Xba1	
CT_OL53	Tb11.01.7810 / MCM7_HA	5' CCC**AAGCTT**CGAGTGAGAATGTAAATTTACC	HindIII	pNAT-6HA
CT_OL54		3' CCC**TCTAGA**CTCCAGTGAAAAGTGAACAG	Xba1	
CT_OL55	Tb11.02.5730 / MCM2_HA	5' CCC**AAGCTT**CGAACGCAACAGGAGAGCCG	HindIII	pNAT-6HA
CT_OL56		3' CCC**TCTAGA**CACCAGGGAGTGCTCTATTT	Xba1	
CT_OL57	Tb11.02.3270 / MCM5_HA	5' CCC**GAGCTC**GGGGGTAGTGCTCGTCATTC	SacI	pNAT-6HA
CT_OL58		3' CCC**TCTAGA**ACGTAACCGATGAATTTGGGG	Xba1	
CT_OL59	Tb11.02.4070 / MCM4_HA	5' CCC**AAGCTT**GCCACCAGTCGTGCCCCCAT	HindIII	pNAT-6HA
CT_OL60		3' CCC**TCTAGA**CGAAAGCCTACCGGCGAAAC	Xba1	
CT_OL61	Tb927.2.3930 / MCM3_HA	5' CCC**AAGCTT**GAAGCAATTGTGCGACTAGC	HindIII	pNAT-6HA
CT_OL62		3' CCC**TCTAGA**AATGAACTGAACCCATTCAT	Xba1	
CT_OL63	Tb927.10.13380 / Tb13380_RNAi	5' CCC**GGATCC**CACGTTGTATCCCCTTGCTT	BamH1	pZJM_dual T7
CT_OL64		3' CCC**AAGCTT**TTCAGTTTCGGCGAAGTTCT	HindIII	
CT_OL65	Tb927.10.13380 / Tb13380_HA	5' CCC**AAGCTT**CGTTTCTGCTGTCTTTGGGG	HindIII	pNAT-6HA
CT_OL66		3' CCC**TCTAGA**CACGAGGCTGCGTAATC	Xba1	
CT_OL67	Tb927.10.7980/Tb7980_RNAi	5' CCC**AGATCT**AGTGTGGCTCCGGTTACATC	BglII	pZJM_dual T7
CT_OL68		3' CCC**AAGCTT**TTGTTGCAAAGAGCGTGTTC	HindIII	
CT_OL69	Tb927.10.7980/Tb7980_HA	5' CCCAAGCTTCGAGCAAGTATCGTCACACAGGA	HindIII	pNAT-6HA
CT_OL70		3' CCCTCTAGATCGTGGAATGAGGTCGT	Xba1	
CT_OL71	Tb01_281401_1	5' CGATGAATGCAAAGACATGTGA		
CT_OL72		3' CGATGAATACGCGAAATGGA		
CT_OL73	Tb01_281401_4	5' TGAACTACCCTCCGTGGACAA		
CT_OL74		3' GATCCCATAGTGCCGTTGAAA		
CT_OL75	Tb01_281401_5	5' TTTATTCTTAGCACATTGCGGAGTA		
CT_OL76		3' AGGACGTTCAAAGCAGAATAAGCT		
CT_OL77	Tb01_281421_13	5' GAACTCTCATCTGCCCTTGGTT		
CT_OL78		3' GCAGTCTCAACACGGGTATTTTG		

3 Functional characterisation of TbORC1/CDC6, a putative *Trypanosoma brucei* replication initiation protein

3.1 Introduction

T. brucei is a unicellular eukaryotic parasite responsible for sleeping sickness disease in humans and nagana in livestock. The parasite is characterised by the presence of two genomes: (i) a mitochondrial DNA network that comprises topologically interlocked minicircles (~1 kb in size) and maxicircles (~ 23 kb in length), collectively known as the kinetoplast or mitochondrial genome (Liu *et al.* 2005); and (ii) a nuclear genome which comprises 11 diploid megabase-sized chromosomes (0.9-6 Mb in length), several intermediate-sized chromosomes (200-900 kb in size) and approximately 100 mini-chromosomes (~50-150 kb in size) (Melville *et al.* 1998; Gottesdiener *et al*. 1990). Due to the unique and unconventional composition and organisation of the kinetoplast DNA (kDNA), a distinct repertoire of specialised proteins is dedicated to its replication and segregation to ensure cell propagation and maintain diversity. First isolated and characterised from *Leishmania* by the lab of Larry Simpson in 1971 (Simpson & Dasilva 1971), the kDNA replication machinery has been the subject of immense scientific research for well over 40 years in trypanosomes and *Crithidia* by the labs of P. Englund, M. Klingbeil, D. Ray and J. Shlomai, amongst others [for reviews see: (Liu *et al.* 2005; Shlomai 2004; Klingbeil & Englund 2004; Klingbeil *et al*. 2001; Shapiro & Englund 1995). From initiation to termination, kDNA replication involves the concerted participation of ~30 characterised proteins. For minicircle replication, p38, the universal minicircle binding protein (UMSBP), TbPIF1, at least one primase, DNA polymerases (pols) IB and IC, and Topoisomerase II are just a few factors involved in the initiation events. TbPIF5, structure-specific endonuclease-I (SSE 1), DNA pol ß and DNA ligase kß are involved in later steps (Liu *et al*. 2010; Bruhn *et al*. 2010; Liu *et al*. 2009a; Milman *et al.* 2007; Liu *et al*. 2006; Liu *et al.* 2005). Although much less is known about maxicircle replication, it is estimated that both sets of interlocked DNA classes require ~100 proteins for a complete duplication and segregation (P.T. Englund, personal communication).

Despite the increasing clarity in our understanding of the machinery and mechanism of kDNA replication, until the publication of the sequences of the *T. brucei, T. cruzi* and *L. major* genomes in 2005 (Ivens *et al.* 2005; Berriman *et al.* 2005; El Sayed *et al.* 2005a) the components needed for trypanosomatid nuclear

DNA replication remained relatively uncharacterised. Bioinformatic analysis and gene annotation suggested that the replication initiation machinery may be atypical of eukaryotes (El Sayed *et al.* 2005b).

The DNA replication initiation machinery in higher eukaryotes is covered extensively in Section 1.6.3. The mechanism and machinery of DNA replication initiation is well-conserved among characterised eukaryotes (Dutta & Bell 2006). The six protein origin recognition complex (ORC, Orc1-Orc6), Cdc6, and Cdt1 are recruited sequentially to DNA, and once bound, they load the replicative helicase (MCM, a heterohexamer; subunits Mcm2-7) to form a pre-replicative complex (pre-RC) at potential origins of replication (Dutta & Bell 2006). The largest subunit of ORC, Orc1, is related in sequence to Cdc6, indicative of derivation from a common ancestor (Duncker *et al.* 2009). Such an ancestral molecule appears still to function in archaea (Barry & Bell 2006). These prokaryotes lack Cdc6 and possess a protein named Orc1/Cdc6, which appears to provide all ORC functions, since orthologues of Orcs2-6 are absent. In addition to this, archaeal orthologues of Cdt1 have not been clearly described, though a potentially related factor, named WhiP (winged helix initiator protein), has been found (Robinson & Bell 2007). Comparative genome analysis of *Trypanosoma brucei* and related trypanosomatids (*L. major* and *T. cruzi*) revealed, remarkably, only a single ORC protein that is equally related to eukaryotic Orc1 and Cdc6 (named here TbORC1/CDC6) (El Sayed *et al.* 2005b). In addition, no clear homologue of Cdt1 was found. These observations have been interpreted as suggesting that origin designation in trypanosomatids, although eukaryotic, may be archaeal-like, raising numerous mechanistic and evolutionary questions.

Beyond a core function in origin designation, there is evidence for additional non-replication roles of at least the largest subunit of the ORC complex, Orc1, in higher eukaryotes. First characterised in *S. cerevisiae,* Orc1 has been shown to mediate the formation of transcriptionally silent chromatin domains at silent mating loci, *HMR* and *HML*. Once bound to DNA, *S. cerevisiae* Orc1 interacts with the transcription factors Rap1 and Abf1 via an N-terminal Bromo-adjacent (BAH) domain. This interaction then recruits Silent Information Regulator proteins (Sir 1 – 4), allowing the formation and spreading of heterochromatin. The formation of heterochromatin mediated by ORC proteins appears to be a conserved eukaryotic role reported in yeasts, *Drosophila* and mammals (Sasaki & Gilbert

2007). Bearing in mind the obvious similarity between TbORC1/CDC6 and its eukaryotic Orc1 counterpart (albeit lacking the N-terminal BAH domain), and the high density of localisation of the *T. brucei* protein at subtelomeres (see chapter 5), we hypothesised that TbORC1/CDC6 might play a role in the regulation of *VSG* expression. This hypothesis, if correct, would lend support to the growing data suggesting that epigenetic phenomena are closely associated with antigenic variation and *VSG* transcriptional control in *T. brucei* (Elias & Faria 2009; Siegel *et al*. 2009; Yang *et al*. 2009; Stockdale *et al*. 2008; Figueiredo, Janzen, & Cross 2008).

3.2 Results

3.3 RNA interference of ORC1/CDC6 in procyclic form *T. brucei*

To investigate the role of TbORC1/CDC6 in *T. brucei*, an RNAi approach was taken. The construction of an RNAi vector (Figure 3-1) that uses two opposing tetracycline-inducible T7 promoters to controllably drive the expression of double-stranded RNA to target a specific gene of interest has facilitated the rapid screening of genes essential for viability in *T. brucei* (Motyka & Englund 2004; Wang *et al*. 2000). The construct stably integrates in a ribosomal locus in the *T. brucei* genome after it has been digested with *Not*I and transfected by electroporation (see Materials and Methods) into a transgenic *T. brucei* Lister 427 pLew13 pLew29 cell line constitutively expressing T7 RNA polymerase and the tetracycline repressor (TetR)(Wirtz & Clayton 1995). Upon addition of tetracycline, expression of double stranded RNA (ds-RNA) is induced and this is processed by the RNAi machinery of the parasite, thereby reducing transcript levels of a target gene (Ullu, Tschudi, & Chakraborty 2004; Ullu *et al*. 2002; Ngo *et al*. 1998).

3.3.1 Cloning of TbORC1/CDC6 RNAi constructs

For RNAi of *TbORC1/CDC6*, oligonucleotides were designed for amplification of an 579-bp region (positions 252 – 831 relative to the start codon) of *TbORC1/CDC6* using an automated web-based tool for the selection of RNAi

targets in *T. brucei* (http://trypanofan.path.cam.ac.uk/software/RNAit.html) (Redmond, Vadivelu, & Field 2003). The primers (CT_OL01/CT_OL02, Table 2-2) contained *Bam*HI (forward primer) or *Hin*dIII (reverse primer) recognition sequences. The fragments were amplified by polymerase chain reaction (PCR) from *T. brucei* PCF Lister427 strain genomic DNA (gDNA) using Phusion® Taq DNA polymerase (New England Biolabs ®) – see Materials and Methods. After PCR amplification, the gene fragment was cloned into *Bam*HI/*Hin*dIII site of the pZJM dual T7 vector shown in Figure 3-1. The vector was linearised at the rRNA spacer using the restriction enzyme *Not*I and transfected into the *T. brucei* Lister427 pLew13-pLew29 cell line. Selection of positive clones was carried out using Zeocin at a concentration of 10 $\mu g.mL^{-1}$.

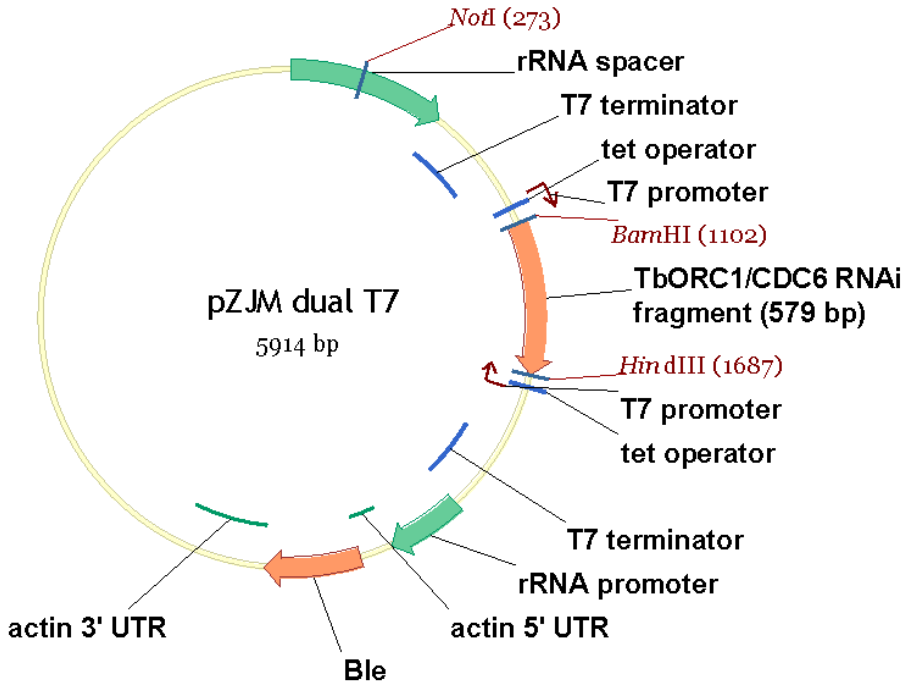


Figure 3-1 - The pZJM dual T7 RNAi vector for *TbORC1/CDC6* RNAi

A fragment of *TbORC1/CDC6* (579 bp) that was cloned into *Bam*HI and *Hin*dIII sites in the plasmid, after PCR amplification, is shown; this is surrounded by T7 promoter sequences and by operator sequences (tet operators) that bind the Tetracycline repressor, together allowing the controlled expression of *ORC1/CDC6* dsRNA. The *Not*I site used for vector linearization to allow integration into the rRNA spacer sequence is also shown. Selection of *T. brucei* transfectant clones was perfomed using the phleomycin resistance gene (*BLE*), whose expression is derived from an rRNA promoter, as indicated. All features shown on the vector map are designed using the VectorNTI resource (Invitrogen). The vector is originally from (Wang *et al*. 2000).

3.3.2 Effect of TbORC1/CDC6 RNAi knockdown on growth

After transfection of the *TbORC1/CDC6* RNAi construct and drug selection, three independent clones were selected and growth curves were generated in the absence or presence of 1.0 μg.ml^{-1} tetracycline, which induces expression of the *TbORC1/CDC6* dsRNA thus initiating RNAi. Growth of the trypanosomes was followed for 144 hours, and induction of *TbORC1/CDC6* RNAi resulted in a growth defect clearly visible from 84 hrs post induction (Figure 3-2), relative to the non-RNAi induced cells. Cell counts were carried out in triplicate for each clone and all growth curves appeared similar for both RNAi-induced and non-induced samples (Figure 3-2). A single clone was selected at random for further analysis. First, we sought to investigate whether the growth defect was specifically due to knockdown of *TbORC1/CDC6* expression.

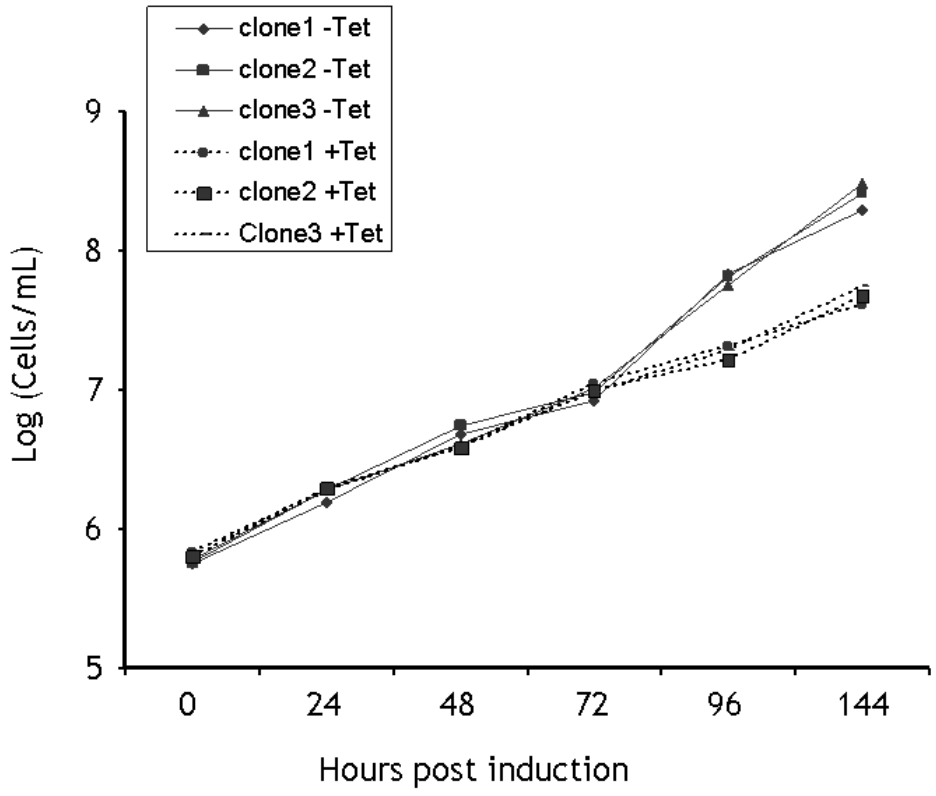


Figure 3-2 - Effect of *TbORC1/CDC6* RNAi on the growth of procyclic form *T. brucei*

Cell counts of *T. brucei* Lister 427 pLew13-pLew29 cells transformed with a *TbORC1/CDC6* RNAi construct are shown over time for 3 independent clones in the absence of *TbORC1/CDC6* RNAi induction (Tet-; solid lines) or following RNAi induction by the addition of 1 μg.ml^{-1} tetracycline (Tet+; broken lines). The cumulative cell numbers reflect correction of the cell densities following dilution after 72 hrs.

3.3.3 mRNA levels after TbORC1/CDC6 RNAi knockdown

To determine if the above growth impairment was due to a decrease in *TbORC1/CDC6* mRNA levels after RNAi induction by addition of tetracycline, semi-quantitative reverse transcription-polymerase chain reaction (semi-qPCR) was first used to examine *TbORC1/CDC6* mRNA levels 96 hrs post induction of RNAi compared with a non-induced sample. Using the single clone selected, total RNA was extracted before and after induction with tetracycline. 1 µg of total RNA was used for cDNA synthesis with 50 µM oligo (dT) primers, as per manufacturer's instructions (SuperScript™ kit, Invitrogen). Identical reactions were set up for each RNA sample with either reverse transcriptase added (RT +) or without reverse transcriptase (RT -) to act as a control for genomic DNA contamination. Using 1 µl of cDNA synthesised per reaction, PCRs were set up for 26 cycles with primers for *TbORC1/CDC6* (CT_OL3/CT_OL4, Table 2-2); a 0.5 kb fragment outside of the fragment cloned into the pZJM vector. As a control, mRNA levels of *RNA Pol I*, an unrelated transcript, were also examined. Using the same cDNA sample from above with primers (CT_OL5/CT_OL6, Table 2-2), a 0.5 kb fragment of *RNA Pol I* was amplified. Both the induced and non-induced samples (RT- and RT+) were resolved on a 1 % agarose gel. The results showed that while less *TbORC1/CDC6* PCR product was generated from the RNAi-induced cDNA relative to the uninduced, the *PolI* RNA control was unaffected (Figure 3-3A).

To quantify levels of mRNA knockdown, quantitative real time PCR (qPCR) was carried out (See Materials and Methods) using a Prism7500 real time PCR machine (Applied Biosystems). Oligonucleotides CT_OL7/CT_OL8 (Table 2-2) for *TbORC1/CDC6* was used to amplify a ~70-bp fragment of the gene distinct from the RNAi vector insert. *Glycosylphosphatidylinositol-8 (GPI8)*, an unrelated transcript, was amplified as an internal control using primers CT_OL23/CT_OL24 (Table 2-2). Distilled H_2O was used as non-template control. All PCRs were set up in triplicate and the data analysed using Applied Biosystems 7500 system software. qPCR of cDNA from the same samples used in the semi-qPCR revealed that, 96 hrs post RNAi induction, *TbORC1/CDC6* mRNA levels were reduced by ~90 % when compared with non-induced samples and with *GPI8*, an unrelated transcript used as endogenous control (Figure 3-3B).

These data demonstrate that knockdown of *TbORC1/CDC6* mRNA by RNAi was efficient by 96 hrs, and that the growth phenotype observed in Section 3.3.2 after this time was in all likelihood associated with specific depletion of *TbORC1/CDC6* mRNA.

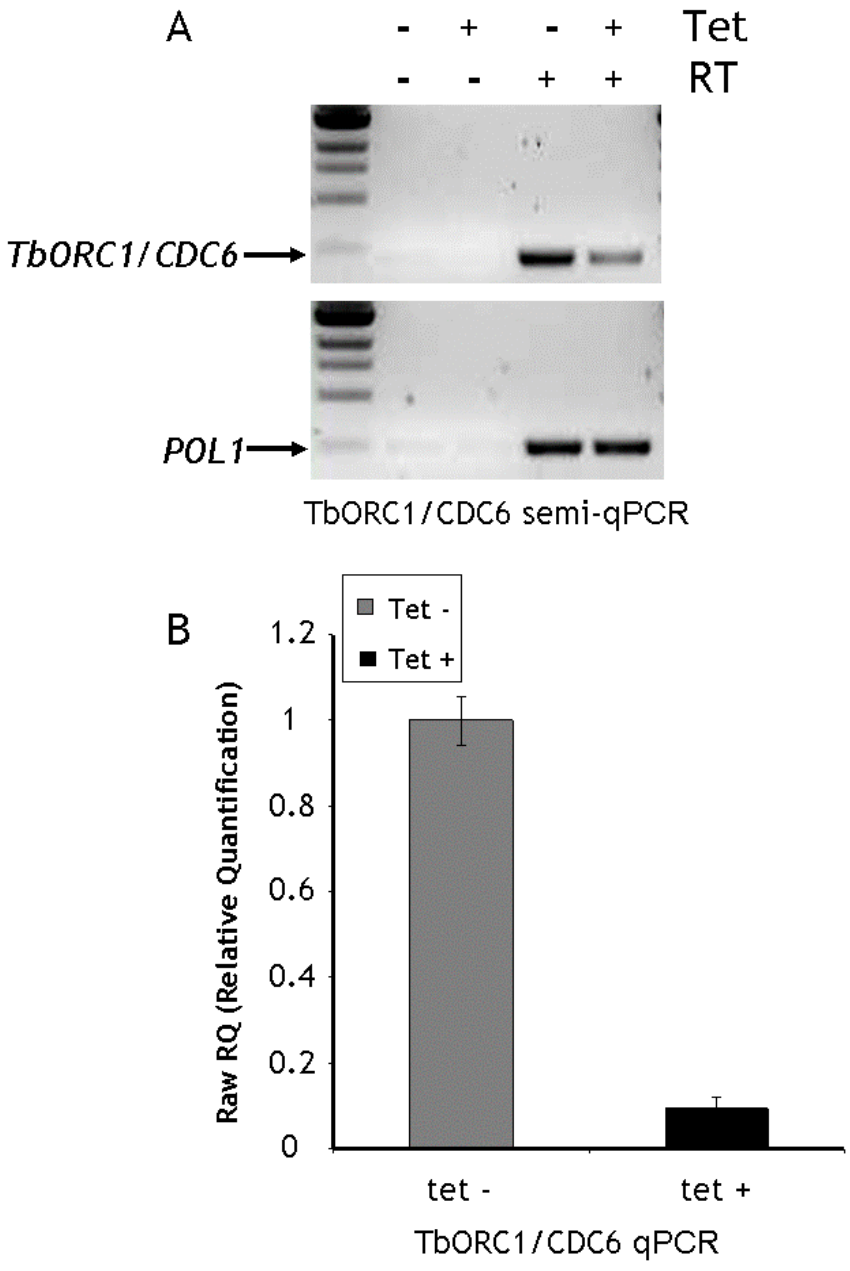


Figure 3-3 - Measurement of *TbORC1/CDC6* transcript levels 96 hrs post induction of RNAi

(A) Semi-quantitative reverse transcriptase PCR (semi-qPCR) was performed by 26 cycles of PCR using oligonucleotide primers for amplification of a region of *TbORC1/CDC6* distinct form that cloned into the pZJM RNAi vector. PCR products are shown from cDNA substrate generated from RNA derived from cells induced for RNAi (+Tet) and uninduced (-Tet); RT+ indicates that reverse transcriptase was used to generate cDNA, while RT- indicates control samples generated in the absence of reverse transcriptase enzyme. The same PCR is shown for an unrelated transcript, RNA Polymerase I (PolI) as loading control. (B) Quantitative PCR (qPCR) to determine *TbORC1/CDC6* mRNAi knockdown levels. The abundance of *TbORC1/CDC6* cDNA from RNAi-induced cells (Tet +, black bar) is shown to relative to control cells without *TbORC1/CDC6* RNAi (tet -, grey bar). The concentration of PCR product in the non-induced sample is normalised to 1.

3.3.4 FACS analysis after TbORC1/CDC6 RNAi knockdown

To measure the relative DNA content of cells, and to analyse the cell distribution during the various phases of the cell cycle before and after RNAi of *TbORC1/CDC6* in procyclic form *T. brucei* Fluorescent-Activated Cell Sorting (FACS) was performed on propidium-iodide (PI) - stained RNAi induced and non-induced control cells. In this analysis, cells were examined 72 hrs and 120 hrs after RNAi induction (Figure 3-4). A total of 10^4 cells were sorted for DNA content with Becton Dickinson FACSCalibur using detector FL2-A and an AmpGain value of 1.75. In such FACS plots, cells in the G_1 phase of the cell cycle appear at position 200 (labelled 2C, where C represents a haploid DNA content); cells with double DNA content (labelled 4C), in $G_{2,}$ appear at position 400; while S-phase cells are in-between G_1 and G_2 (Figure 3-4). Aberrant cells can also be detected: those with DNA content < 2C (normally anucleate cells, or zoids) appear on the far left of the plot at approximately position zero on the PI-axis; cells with >4C content are found to the right of position 400.

After knockdown of *TbORC1/CDC6* for 72 hrs, there was an accumulation of putative zoids with a corresponding decrease in the population of 2C cells (Figure 3-4A). This became more pronounced after 120 hours, with an increase in zoids of ~ 15 fold relative to a non-induced sample (Figure 3-4B). After 120 hrs of inducing *TbORC1/CDC6* RNAi there was also a marked reduction in 4C cells, and potentially of 2C cells, with greater staining of S-phase cells (between 2C and 4C). This suggests that cells appear unable to progress beyond S-phase after RNAi, perhaps due to impaired G1/S and/or G2/M transition. This observation correlates with an inability of *TbORC1/CDC6* RNAi-induced cells to synthesise new nuclear DNA (see below).

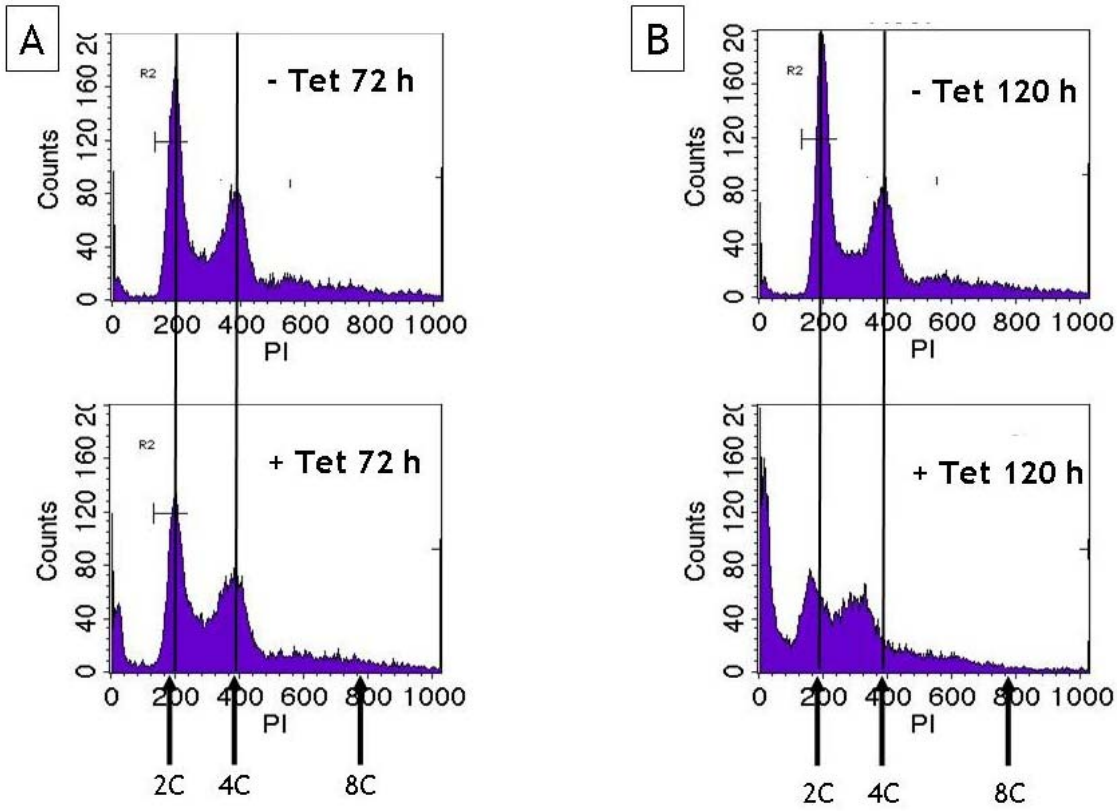


Figure 3-4 - FACS profiles of PI-stained cells after *TbORC1/CDC6* RNAi

Histograms show propidium iodide-stained *T. brucei* procyclic-form cells after FACS sorting, sampled 72 hrs (A) or 120 hrs (B) pre- and post - induction of RNAi for *TbORC1/CDC6* (- Tet and + Tet ,respectively). The peaks corresponding with cells containing 2C and 4C DNA content are indicated, as is the peak position for cells with 8C content; where C represents a haploid DNA content.

3.3.5 Cell morphology after TbORC1/CDC6 RNAi knockdown

In order to get a clearer picture of the cell cycle distribution of cells, and to distinguish between cell debris (which co-localise with zoids in FACS analyses) and genuine anucleate cells, a quantification of the various cellular morphologies throughout the period of RNAi induction was carried out by sampling the cells and staining them with 4,6-Diamidino-2-phenylindole (DAPI) – see Materials and Methods. DAPI-staining of DNA in trypanosomes provides a tool to monitor cell cycle progression since the number of kinetoplasts and nuclei in an individual cell enables cell cycle stage-specific classification (see below). Approximately 200 cells were classified into various subgroups, depending on the number of nuclei (N) or kinetoplast (K) DNA that could be detected microscopically. Without *TbORC1/CDC6* RNAi induction, cells appeared normal with the expected distribution of N-K DNA configuration (Woodward & Gull 1990)

(Figure 3-5A): cells with one kinetoplast and one nucleus (1K1N) made up ~ 80 % of the total population; cells with two kinetoplasts and a single nucleus (2K1N, having divided the kDNA but not nDNA) comprised ~10 %; and, after nuclear division but before completion of mitosis, cells with two nuclei and two kinetoplasts (2K2N) made up ~5 % of the total population. Aberrant cells, not corresponding to these N-K configurations, represented only ~3 % of cells in the absence of *TbORC1/CDC6* RNAi. In a typical scenario, 2K2N cells will undergo cytokinesis to produce two daughter progeny each with a 1K1N configuration and the cycle continues. However, aberrant cells can also be present in the population and these usually constitute ~3.5 % (Woodward & Gull 1990), similar to that observed here. Thus, without induction of RNAi up to the 144 hrs sampled, the N-K distribution was as expected. In contrast, from 48 -144 hrs after *TbORC1/CDC6* RNAi there was a steady decrease in the numbers of 1N1K cells, falling from ~70% - to ~35 % of the population. The drop in 1N1K cells was accompanied by an increase in the population of anucleate cells (0N1K; Figure 3-5C), rising from ~5% of the entire population at 48 hrs to ~50% at 144 hrs (Figure 3-5B). There was no apparent increase in other forms of aberrant cells (grouped here as 'unclassified'). These DAPI data appear to correspond with the FACS data and, taken together, suggest that knockdown of *TbORC1/CDC6* in procyclic-form *T. brucei* cells prevents nuclear DNA synthesis but has no observable effect on kinetoplast DNA synthesis or on cytokinesis.

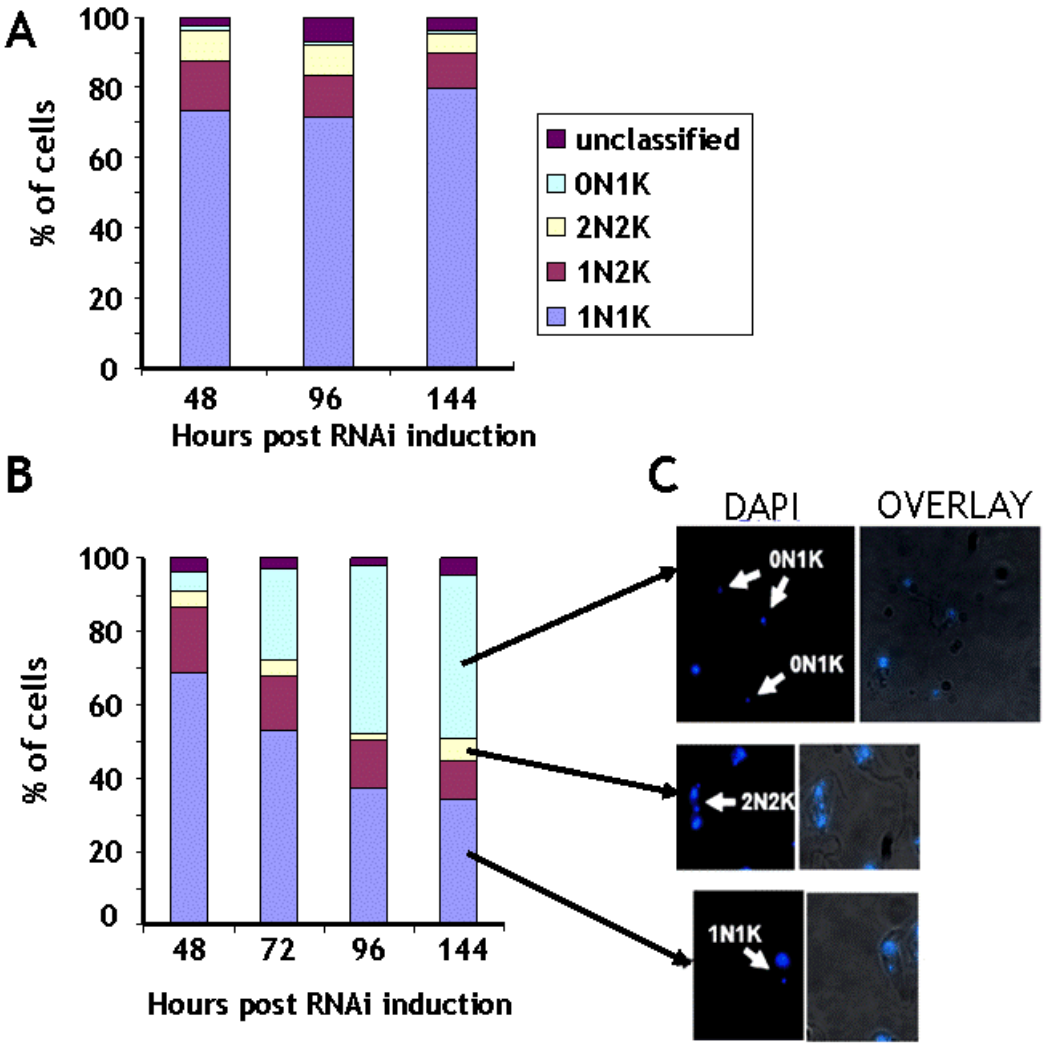


Figure 3-5 – Microscopic quantitation of nuclear (N) and kinetoplast (K) DNA content after DAPI staining

Procyclic form *T. brucei* cells were counted (N = 200) at each time point for non RNAi-induced control cells (A), or cells depleted in TbORC1/CDC6 by RNAi (B). Examples of cells of differing DNA content are shown stained with DAPI, and as an overlay with a phase contrast image (C).

3.3.6 Quantitative assessment of DNA Synthesis during TbORC1/CDC6 RNAi knockdown

So far, we have proposed, based on FACS and DAPI analyses, that RNAi knockdown of *TbORC1/CDC6* mRNA results in the generation of cells which can no longer synthesise nuclear DNA. To examine this directly, we used a quantitative dot blot assay to assess *in vivo* DNA synthesis (Ueda *et al*. 2005). After induction of RNAi for 96 hrs, induced and uninduced cells were incubated with 5-Bromo-2'-Deoxyuridine (BrdU) for 60 minutes, total genomic DNA was

purified from the cells, denatured, 50 ng spotted onto a nitrocellulose membrane, and the incorporated 5-bromo-2-deoxyuridine triphosphate (BrdUTP) was detected with a monoclonal antibody against BrdUTP by western blotting (see Materials and Methods). Using this approach, there was very little visibly detectable BrdUTP signal after *TbORC1/CDC6* RNAi, compared with un-induced samples where signal was readily detected (Figure 3-6A). In contrast, there was no visible difference in the signal intensity of dot-blotted DNA after silencing *TbWhiP/CDT1*, an unrelated gene that also induces zoid formation upon RNAi (Figure 3-6C). Quantifying the signal intensities by densitometry showed a > 4 fold difference in intensity between samples where *TbORC1/CDC6* was RNAi depleted compared with samples where RNAi was not induced (Figure 3-6B), while quantifying *TbWhiP/CDT1* showed no significant difference in intensities between induced and un-induced samples. These data strongly support our previous observations that cells which lack ~ 90% of *TbORC1/CDC6* transcript fail to synthesis new nuclear DNA.

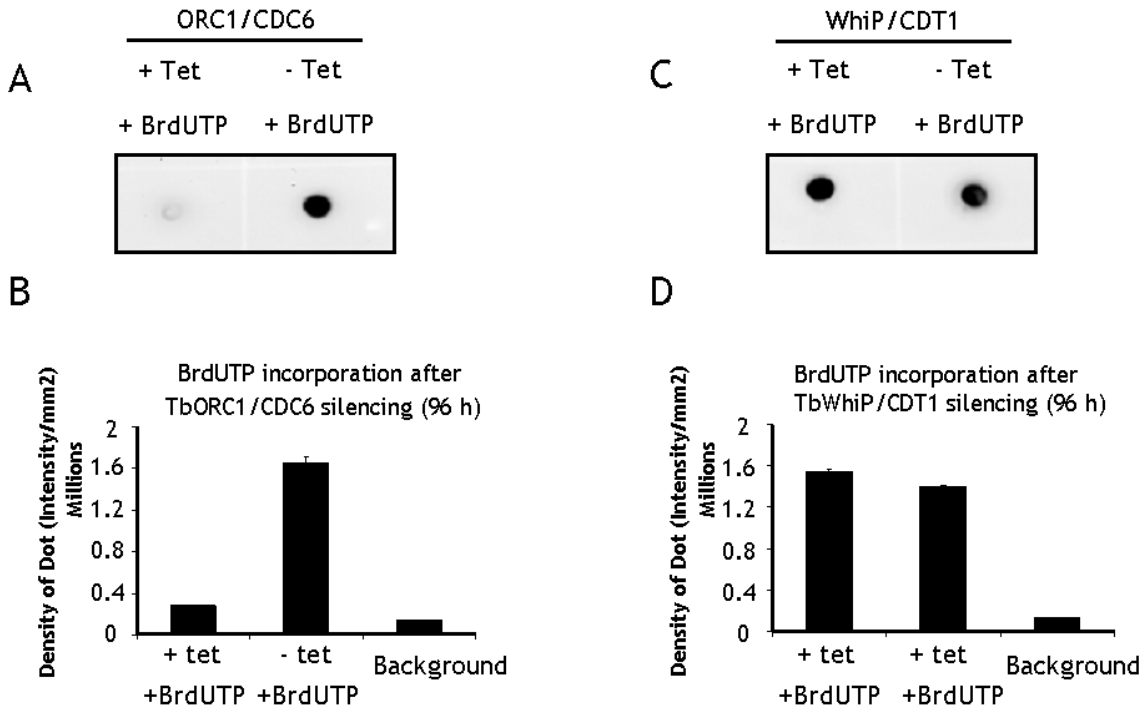


Figure 3-6 - Assaying DNA synthesis by BrdUTP incorporation after RNAi of *TbORC1/CDC6*

Dot-blots are shown of *T. brucei* DNA probed with anti-BrdUTP antibody. In (A) the cells were incubated with BrdU after *TbORC1/CDC6* RNAi was induced by tetracycline (+Tet) for 96 h, and are compared with control cells in which RNAi was not induced (-Tet). In (C) dot blots are shown for a similar analysis performed for *TbWhiP/CDT1*. (B); (D) show quantitative densitometric analysis of the dot-blot samples, calculated relative to a background spot on the membranes.

3.4 Is TbORC1/CDC6 essential for viability of *T. brucei?*

3.4.1 Design of TbORC1/CDC6 knockout constructs and generation of null mutants

The above RNAi data are consistent with *TbORC1/CDC6* being an essential gene, as would be expected for a core factor of replication initiation, but are inconclusive as the extent of protein knockdown and longevity of the knockdown have not been examined. To try and determine more clearly if the gene is essential, attempts were made to make *TbORC1/CDC6* homozygous (null) mutants. To generate a *TbORC1/CDC6* null mutant cell line, constructs were made in which the 5' and 3' flanks (Untranslated Regions; UTRs) of the ORF were cloned into a modified pBluescript KS vector (vector was a Gift from M. Swiderski, WTCMP University of Glasgow). This pBluescript KS vector had antibiotic resistance cassettes encoding resistance to either puromycin (*PUR*) or blasticidin (*BSD*) cloned into it. The *PUR* or *BSD* cassette present in this vector was flanked upstream by a sequence derived from the processing signal in-between ß and α tubulin (ßα), and downstream by a sequence derived from the actin intergenic region (Act) (Figure 3-7A). The 5' UTR region used for targeting *TbORC1/CDC6* was PCR-amplified using primers CT_OL9/CT_OL10, while the 3' UTR region of *TbORC1/CDC6* used was PCR-amplified using primers CT_OL11/CT_OL12, (Table 2-2). The 5' UTR and 3' UTR PCR fragments, of 348 bp and 237 bp respectively, were PCR-amplified from genomic DNA extracted from strain TREU927 using Phusion® Taq (NEB) DNA polymerase enzyme, and were digested with *Sac*I/*Xba*I, and *Pac*I/*Xho*I, respectively. The two fragments were cloned sequentially into *Sac*I/*Xba*I (5' UTR) and *Pac*I/*Xho*I (3'UTR) restriction sites of the pBluescript KS vector (Figure 3-7B). Restriction mapping and DNA sequencing of the final constructs confirmed correct insertion of both UTRs and both drug resistance cassettes. Once both constructs were confirmed, they were called *ΔORC1/CDC6::BSD* or *ΔORC1/CDC6::PUR* (Figure 3-7B). Transfection of *ΔORC1/CDC6::BSD* or *ΔORC1/CDC6::PUR* in *T. brucei* should allow integration of the drug resistance cassettes by homologous recombination, targeting the 5' and 3' UTRs of *TbORC1/CDC6*, and deleting the ORF (Figure 3-7 A, B).

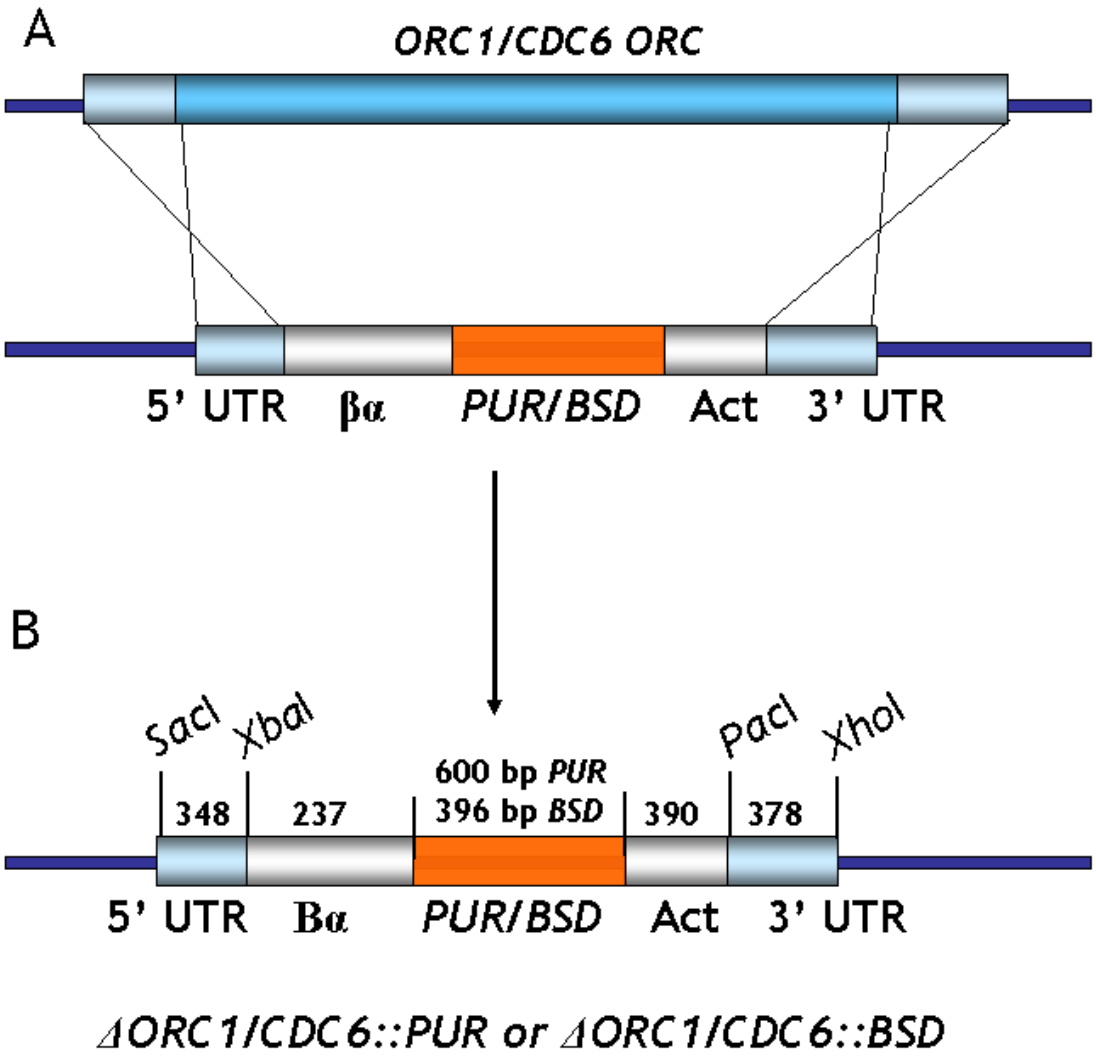


Figure 3-7 – *TbORC1/CDC6* knockout construct and knockout strategy

(A) *TbORC1/CDC6* knockout strategy: the 5' and 3' Untranslated Regions (UTR; sky blue boxes) of *TbORC1/CDC6* that were used for targeted integration of a puromycin resistance gene (*PUR*) or *Blasticidin S Deaminase* (*BSD*) gene are shown flanked 5' by a *Tubulin* UTR (βα) and 3' by an *Actin* UTR (Act). Integration of the linearised construct is meant to occur by homologous recombination (indicated by black crosses) using the 5' and 3' UTRs of *TbORC1/CDC6* ORF.

(B) The resulting locus after integration is represented schematically (*ΔORC1/CDC6::PUR or ΔORC1/CDC6::BSD)*. Numbers shown represent length of DNA fragments (in bp) Location of restriction sites used for digestion and cloning the 5' and 3' UTRs, and linearization are shown by vertical lines.

3.4.2 Confirmation of TbORC1/CDC6 gene disruption by PCR

Once the *ΔORC1/CDC6::PUR* construct was generated and linearised with *Sac*I/*Xho*I, it was transfected into wild type TREU927 procyclic form *T. brucei* and puromycin-resistant clones recovered. Total genomic DNA was prepared from five drug resistant clones and tested by conventional PCR for accurate

integration at the *TbORC1/CDC6* locus. All five clones selected indicated correct integration of the *ΔORC1/CDC6::PUR* construct using the primer pair (CT_OL13/CT_OL14, Table 2-2) indicated by red arrows (Figure 3-8 B, C). As expected, an ~800 bp PCR product was seen when primer CT_OL14 (reverse primer) hybridised 153 bp downstream of the 5' end of the *PUR* gene and primer CT_OL13 (forward primer) hybridised at a sequence 38 bp upstream of the 5' end of the 5' UTR region cloned (Figure 3-8 B, C; red arrows). Figure 3-8C shows that we were able to delete one allele of *TbORC1/CDC6* to generate a heterozygote null mutant; since only in the presence of the *ΔORC1/CDC6::PUR* cassette can we get a PCR product.

To attempt to mutate both alleles of *TbORC1/CDC6*, one of the puromycin-resistant heterozygote mutant clones generated was transfected with the *ΔORC1/CDC6::BSD* construct. Three independent transfection experiments each failed to delete the second allele of *TbORC1/CDC6*. Figure 3-8 B and D show that no PCR products were generated using primers CT_OL15/CT_OL16 (Table 2-2), indicated by blue arrows. The forward primer hybridised upstream of the 5' UTR cloned and the reverse primer hybridised within the blasticidin cassette. However, all five clones selected were blasticidin resistant, suggesting that the *ΔORC1/CDC6::BSD* construct had indeed integrated in the genome. To check if the second *TbORC1/CDC6* allele had been affected by the *ΔORC1/CDC6::BSD* construct integration, we used primers CT_OL1/CT_OL2 indicated by purple arrows (Figure 3-8A) to PCR-amplify the *TbORC1/CDC6* ORF. As seen on Figure 3-8E, the second allele of *TbORC1/CDC6* was unaffected by the *ΔORC1/CDC6::BSD* construct integration. Using primers CT_OL17/CT_OL18, indicated by green arrows (Figure 3-8A), to the *BSD* ORF, PCR confirmed that indeed the cells were resistant to blasticidin due to the presence of the *Blasticidin S Deaminase* gene present in the genome (Figure 3-8F).

The results observed from all experiments described in this section are consistent with *TbORC1/CDC6* being an essential gene as we could not generate a homozygous null mutant. Using blasticidin at the same concentration, we were able to tag the endogenous *TbORC1/CDC6* locus with a myc epitope (see chapter 4) in the puromycin-resistant heterozygous clone, meaning that second allele is accessible for genetic manipulation. Though we have not formally shown that the *BSD* null mutant construct is capable of integrating into the *TbORC1/CDC6*

locus in these conditions, it is equivalent to the *PUR* construct in all regards except the antibiotic resistance ORF.

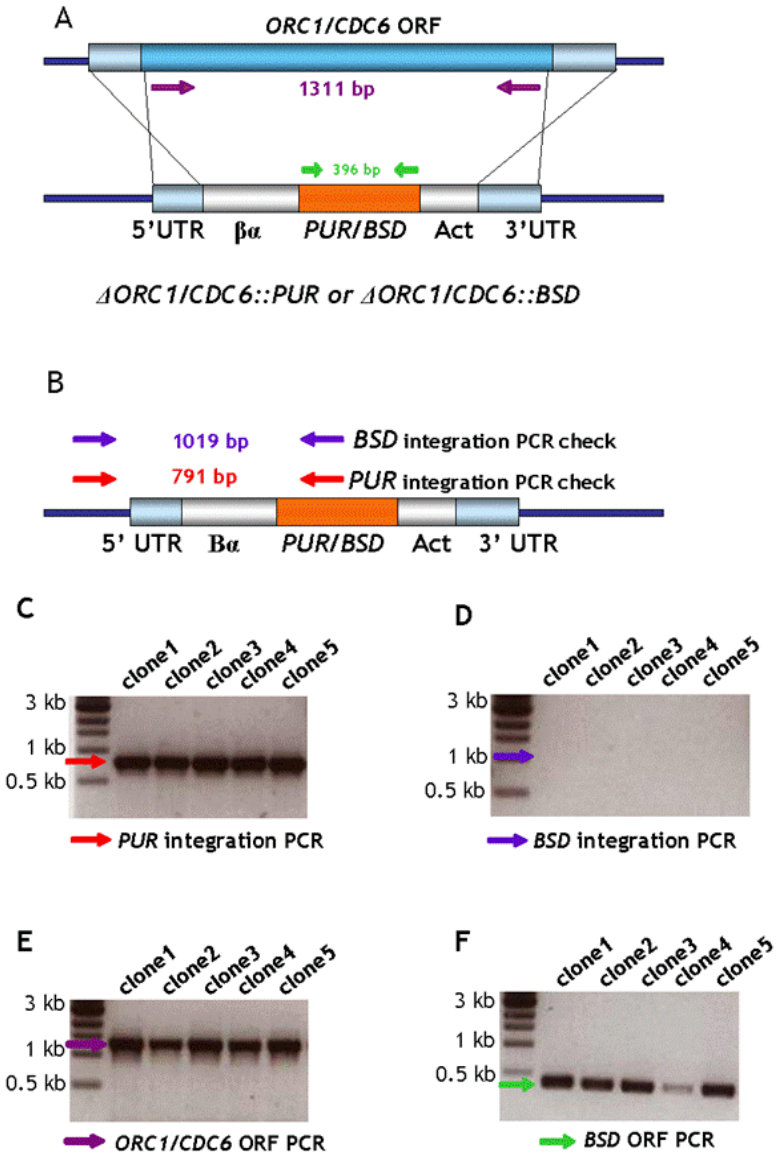


Figure 3-8 - Attempt to generate null mutants of *TbORC1/CDC6* in procyclic form *T. brucei* by targeted gene replacement

(A) Described in Figure 3-8; purple and green arrows indicate position of primers to PCR amplify either *TbORC1/CDC6* ORF (purple arrows) or the *BSD* ORF (green arrows). (B) Linearised schematic map after integration at the *TbORC1/CDC6* locus. Arrows indicate position of primers to check for correct integration by PCR either for the *ΔORC1/CDC6::PUR* construct (red arrows) or for the *ΔORC1/CDC6::BSD* construct (blue arrows). (C) Agarose gel showing PCR products generated by PCR amplification using primers indicated by red arrows in (B) on genomic DNA isolated from 5 puromycin-resistant clones. A band is visible at ~800 bp (red arrow) corresponding to red arrow positions in (B). (D) Agarose gel shows that no PCR product was generated when PCR amplification using primers indicated by blue arrows in (B) is carried out. (E) Agarose gel shows a ~1.3 kb product (purple arrows) which corresponds to the size of a full length wild type *TbORC1/CDC6* allele, derived from primers indicated by purple arrows in (A). (F) Agarose gel shows the presence of a 0.4 kb PCR product which corresponds to the size of full length *BSD* ORF when primers indicated by green arrows (A) are used.

3.5 Localisation of TbORC1/CDC6

3.5.1 Generation of eGFP-tagged TbORC1/CDC6 construct

Procyclic form *T. brucei* cells expressing TbORC1/CDC6-eGFP were generated using a PCR-based approach as described in (Shen *et al*. 2001). Briefly, constructs were designed with oligonucleotides: forward primer [5'-UTR(5'primer)] 5' - ATACGCTCAGCGCTGCCACCCCAGCCCTAAGTGCCGCAATGGACTTACCGTTACTGTCCCTCCTTCGTTTCTTGTACAGCTCGTCCATGC -3' and reverse primer [5'-ORF (3' primer)] 5' - ATACGCTCAGCGCTGCCACCCCAGCCCTAAGTGCCGCAATGGACTTACCGTTACTGTCCCTCCTTCGTTTCTTGTACAGCTCGTCCATGC – 3'. The sequences underlined are the 5' UTR immediately upstream of the start codon of *TbORC1/CDC6* ORF ("ATG") and the 5' *TbORC1/CDC6* ORF immediately downstream of its "ATG", for the forward and reverse primers respectively (see Figure 3-9 for strategy). The primers, which consist of 70 nucleotides of the 5' UTR plus 20 nucleotides of the drug resistance marker (*BSD*) for the forward primer, and 70 nucleotides of the *TbORC1/CDC6* ORF plus 20 nucleotides of the Green fluorescent protein (GFP) ORF sequence for the reverse primer, were used for PCR-amplification as shown in Figure 3-9. The template for the PCR was a plasmid, PGL1464 (a gift from Tansy Hammarton, University of Glasgow). PGL1464 has the 3' end of the *TY1* epitope tag and *eGFP* ORF subcloned into it to generate a *TY1:eGFP* fusion construct. The TY1 epitope tag corresponds to 10 amino acids from the immunologically well-characterised major structural protein of *S. cerevisiae* TY1 virus-like particle (Shen *et al*. 2001). Just upstream (5' end) of the *TY1:eGFP* fusion is an intergenic region to provide RNA processing signals for trans-splicing and polyadenylation, and a *BSD* gene upstream (5' end of intergenic region) to serve as a selectable marker for stable transfectants. After PCR, as shown in Figure 3-9, the product was subcloned into pCR2.1 Topo (Invitrogen). The pCR2.1 construct was sequenced to check for mutations in the oligonucleotide primers incorporated during synthesis or mutations incorporated by the PCR. The pCR2.1 construct was digested with *Eco*R1, ethanol precipitated and *T. brucei* strain TREU927 wild type parasites were transfected as described in Materials and Methods. Selection of positive transfection was achieved at 10

µg.ml^{-1} Blasticidin. Two clones were selected for *TbORC1/CDC6* tagged to *eGFP* and checked for correct integration of the construct by PCR.

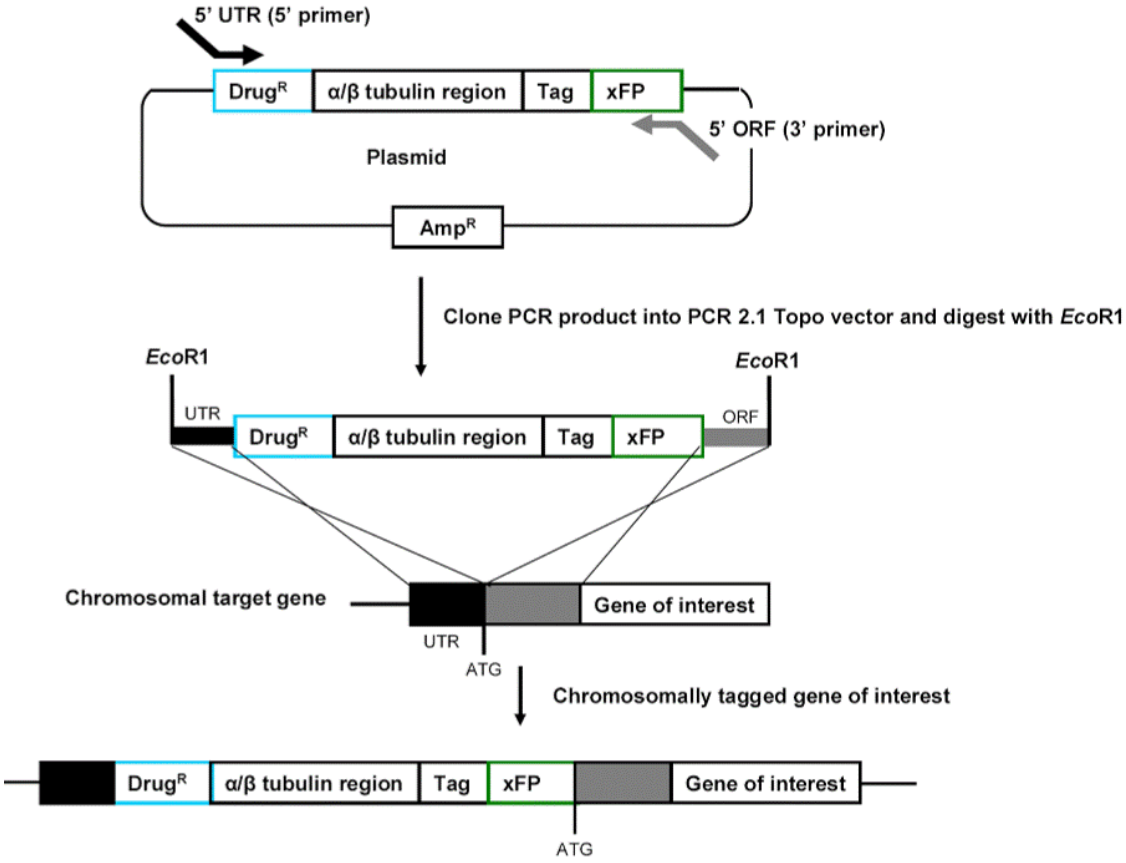


Figure 3-9 - Strategy for *in vivo* eGFP tagging of *TbORC1/CDC6* ORF in *T. brucei*

Key: [5' UTR (5' primer)] = 70 nucleotides (ntds) of 5' UTR immediately upstream of the start codon ("ATG") of *TbORC1/CDC6* gene + 20 ntds of *Drug^R* gene (*BSD* gene); [5' ORF (3' primer)] = 70 ntds of ORF immediately downstream of *TbORC1/CDC6* "ATG" + 20 ntds of *Green fluorescent protein* (xFP) ORF, xFP = *Green fluorescent protein*, Tag = *TY*. Adapted from (Shen *et al*. 2001)

3.5.2 Confirmation of in vivo eGFP-TbORC1/CDC6 tagging by PCR and western blot analysis

The construct generated in Section 3.5.1 was transfected into wild type TREU927 line and positive clones were recovered after drug selection with 10 µg.mL^{-1} Blasticidin. To check for correct integration two independent clones were selected and PCR was carried out using oligonucleotides CT_OL19/CT_OL20 (Table 2-2), indicated by red arrows (Figure 3-10A and Figure 3-10B). Western blot analyses of whole parasite extracts using anti-GFP antibody (SIGMA) and BB2 monoclonal antibody (anti-TY antibody from T. Hammarton) was used to check expression of TY-GFP to determine if they have been translationally fused to

TbORC1/CDC6. In all experiments, untransfected *T. brucei* strain 927 parasite whole extracts was used as a negative control to exclude non-specific bands in the western blots and non-specific PCR products.

The results showed a PCR product size of ~ 2.4 kb as expected (Figure 3-10A, B) if the forward primer (CT_OL19) hybridised at the 5' end of the *BSD* ORF and the reverse primer (CT_OL20) hybridised 629 bp downstream of the *TbORC1/CDC6* ORF. This PCR indicated that *TY-GFP* was fused to full length *TbORC1/CDC6* at its genomic locus. Western blot with anti-GFP and anti-TY also confirmed that full length protein was expressed with a band visible at ~ 65 kDa (Figure 3-10 C, D).

Since the TREU927 line used was not a heterozygote null for *TbORC1/CDC6*, we could not claim that the tagged TbORC1/CDC6 protein was fully functional, as the other allele remains untagged. However, confirmation of expression by western blot implied that the protein was not targeted for degradation. Therefore we used this line for immunofluorescence to localise TbORC1/CDC6 in *T. brucei*.

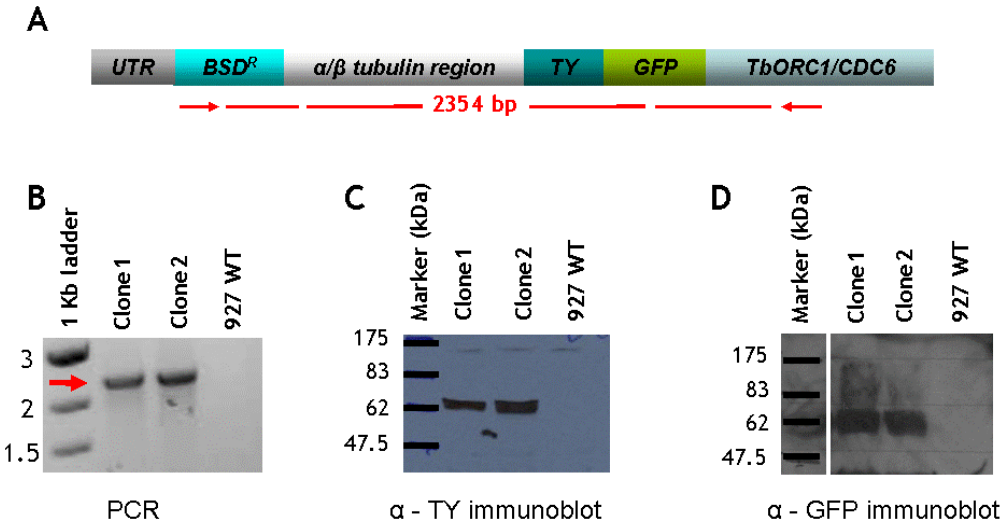


Figure 3-10 – *TbORC1/CDC6-eGFP* at the endogenous locus

(A) Linearised schematic map after integration at the *TbORC1/CDC6* locus. Red arrows indicate position of primers to check for correct integration by PCR; forward primer binds at the 5' end of *BSD* ORF while reverse primer binds at middle of *TbORC1/CDC6* ORF. (B) Agarose gel shows PCR product at ~ 2.4 kb for two clones corresponding to position of red arrows in (A), with no band present in an untransfected line (927 WT). (C) Western blot of whole cell parasite extract of two independent clones using anti-TY antisera show a band at ~ 65 KDa with no band present in whole parasite extract from an untransfected line (927 WT). (D) Western blot of whole cell parasite extract of two independent clones using anti-GFP antisera show a band at ~ 65 KDa with no band present in whole parasite extract from an untransfected line (927 WT).

3.5.3 Confirmation of TbORC1/CDC6 localisation by Immuno-fluorescence

Procyclic form *T. brucei* cells generated expressing TbORC1/CDC6-eGFP or TbORC1/CDC6-12Myc was used for localisation of TbORC1/CDC6 by fluorescent microscopy or immuno-fluorescence, respectively. The generation and confirmation of TbORC1/CDC6-12Myc construct and cells is described in detail in Section 4.5. DNA in the cell was stained with DAPI. For cells in which TbORC1/CDC6 had been tagged with eGFP at the endogenous locus, microscopic visualisation was carried out using either a DAPI filter to see the DNA staining, a FITC filter for the eGFP signal and phase contrast microscopy for DIC imaging of the whole cell. AlexaFluor488 conjugated anti-myc antibody (Millipore®) at 1:1000 dilution was used to localise TbORC1/CDC6 in a fixed cell line expressing 12-Myc tagged to TbORC1/CDC6 at the endogenous locus (see Materials and Methods, Section 2.7.6). In both cases, TbORC1/CDC6 was localised to the nucleus and nowhere else in the cell (Figure 3-11).

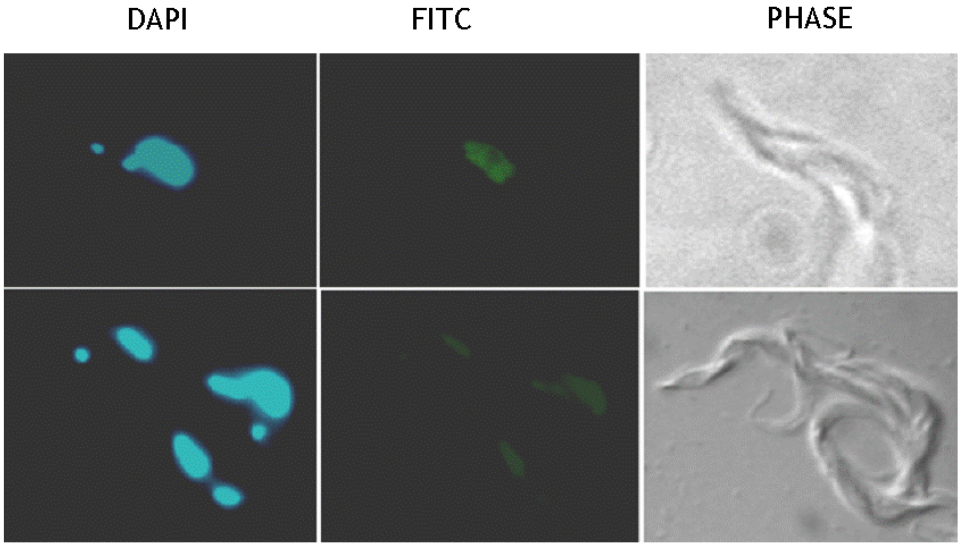


Figure 3-11 - Subcellular localization of TbORC1/CDC6

Top Panel: An example of a *T. brucei* cell expressing endogenous TbORC1/CDC6 C-terminally tagged with 12Myc. Localisation of the protein is shown by immuno-fluorescence using monoclonal anti-myc antibody and alexafluor-conjugated anti-mouse secondary antiserum (FITC); DNA is shown by DAPI stain (DAPI); and a phase contrast image (PHASE) shows the intact cell.

Bottom Panel: Examples of *T. brucei* cells expressing endogenous TbORC1/CDC6-eGFP; localisation of the protein is shown by fluorescent microscopy (FITC); DNA is shown by DAPI stain (DAPI); and a phase contrast image (PHASE) shows the intact cells.

3.6 RNA interference of TbORC1/CDC6 in bloodstream form *T. brucei*

3.6.1 Generation of TbORC1/CDC6 RNAi construct for BSF analysis

To investigate the role of TbORC1/CDC6 *in vivo* in *T. brucei* BSF cells, the same approach described in Section 3.3 was adopted, except that the *TbORC1/CDC6* pZJM dual T7 vector with the same RNAi gene fragment was transfected into the bloodstream form 427 pLew13 pLew90 cell line. This cell line has the same tetracycline inducible properties already described in Section 3.3. After stable transfection and drug recovery with 2.5 $\mu g.mL^{-1}$ of Phleomycin, two independent clones were selected and growth curves were generated in the absence or presence of 1 $\mu g.ml^{-1}$ tetracycline, which induces expression of *TbORC1/CDC6* RNAi.

3.6.2 Effect of TbORC1/CDC6 RNAi knockdown on growth

In the bloodstream form (BSF), cell numbers were counted for two clones for 30 hrs with or without tetracycline induction of RNAi. The results showed growth retardation from 12 hrs and an almost complete disappearance of cells after 30 hrs, as seen in the decrease in cell numbers (Figure 3-12A). *TbORC1/CDC6* RNAi effect on growth in BSF cells was significantly more rapid than was observed in PCF cells, where growth arrest was seen only after ~84 hrs (Section 3.3.2). For a single clone, we asked if the growth defect was accompanied by a reduction in *TbORC1/CDC6*- specific transcript levels. Surprisingly, this greater growth arrest was accompanied by a smaller, though still significant, RNAi-induced knockdown in *TbORC1/CDC6* transcript than seen in PCF cells. Where RNAi in the former caused an ~90 % reduction in *TbORC1/CDC6* transcript, this reduction was only ~ 40 % in BSF cells, as judged by quantitative RT-PCR of RNAi-induced cDNA compared with non-RNAi sample, and using *GPI8* as a control (Figure 3-12B).

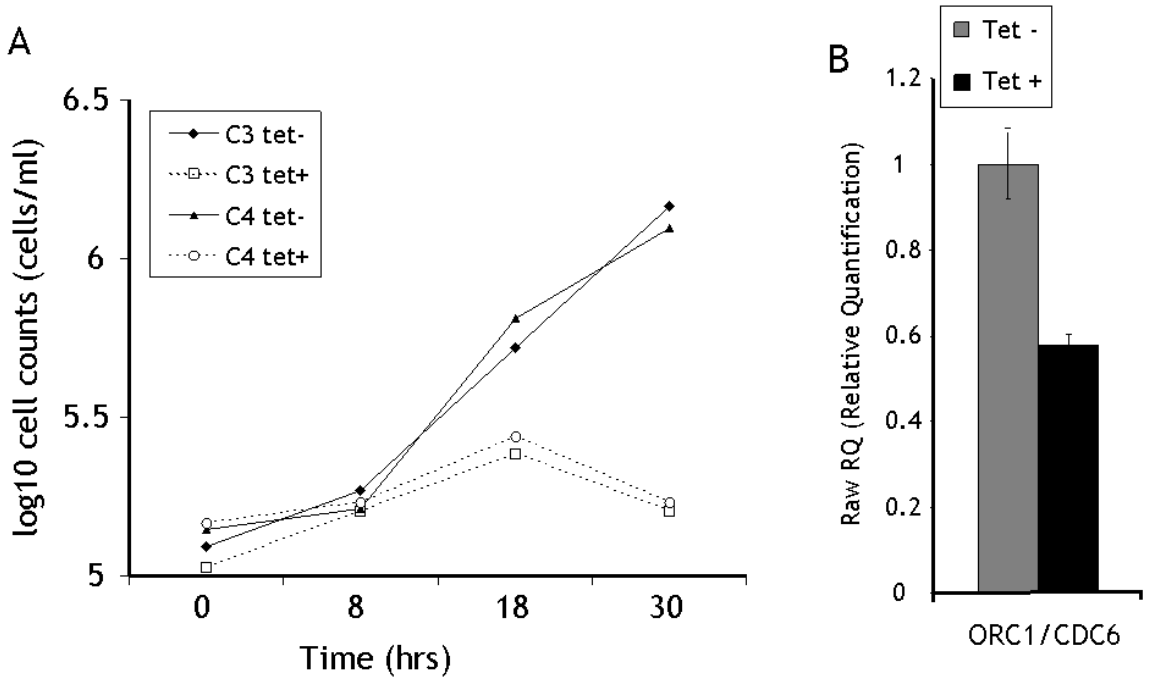


Figure 3-12 - Effect of *TbORC1/CDC6* RNAi in bloodstream cells

(A) Growth was monitored in bloodstream cell lines in which *TbORC1/CDC6* RNAi was induced (+ Tet, solid line) or without induction (-Tet, broken line) for 2 clones; C3 and C4. (B) mRNA levels was measured by quantitative PCR before (-Tet, grey bar) and after (+ Tet, black bar) tetracycline addition for *TbORC1/CDC6* relative to *GPI8* as an endogenous control.

3.6.3 Effect of TbORC1/CDC6 RNAi knockdown on cell morphology

To examine the phenotype of *TbORC1/CDC6* RNAi in BSF cells, nucleus (N) and kinetoplast (K) configurations were monitored for 24 hours in the absence and presence of tetracycline. Analysis of the N and K DNA configuration of the BSF cells after *TbORC1/CDC6* RNAi revealed a drastic contrast to PCF observations described above (Section 3.3.5).

As the two clones responded, by growth, nearly identically, one clone was analysed in this way. ~100 cells were counted after DAPI staining to examine their DNA content. Given the severity of the growth impairment after RNAi, these analyses were conducted only up to 24 hrs after RNAi induction. Unlike in PCF cells, where the most abundant abnormal N-K configuration after *TbORC1/CDC6* RNAI was zoids (0N1K), in the BSF cells most abnormal cells had greater than 2N and 2K DNA configurations by DAPI staining and, in most cases, were unclassifiable. These data are shown in Figure 3-13. Non RNAi-induced cells

displayed the expected ratios of 'normal' N-K configurations (1N1K, 1N2K and 2N2K) for the length of the experiment, and only a minor proportion of aberrant cells (~3-4 %) were seen (Figure 3-13A). In contrast, multi-kinetoplast and multi-nucleate cells accumulated rapidly after RNAi, with ~70 % of cells being aberrant and most (~ 60 %) having >2N>2K DNA content by 18 hrs, which increased to >80 % by 24 hrs, at which time the increased nuclear DNA material was mainly (~80 % of cells) a single condensed mass (Figure 2-8).

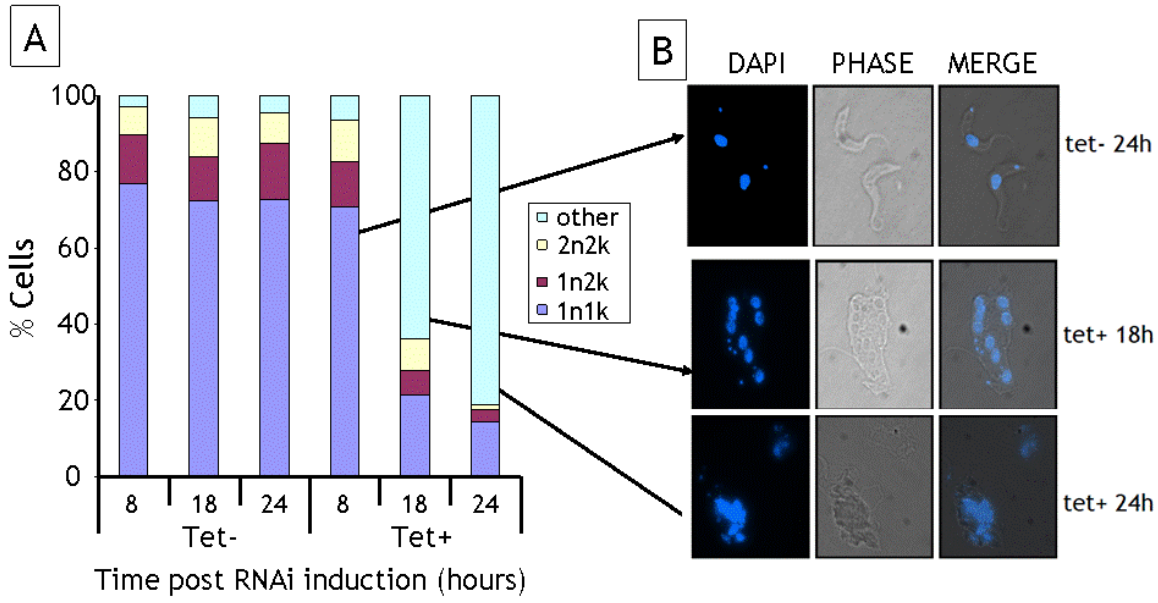


Figure 3-13 – Analysis of nuclear and kinetoplast DNA configuration of *TbORC1/CDC6* RNAi cells for a single clone

(A) DNA DAPI staining and microscopic counting of nuclear (N) and kinetoplast (K) DNA content after DAPI staining. ~ 100 bloodstream form *T. brucei* cells were counted at each time point for non RNAi-induced control cells (A: 8, 18, 24 hrs; Tet -), or cells depleted in TbORC1/CDC6 by RNAi (A: 8, 18, 24 hrs; Tet +). (B) Images of normal cells (Tet –, 24 hrs) and abnormal cells after RNAi induction (Tet +; 18 and 24 hrs)

3.6.4 Effect of TbORC1/CDC6 RNAi knockdown on cell cycle

To further characterise the effect of depleting TbORC1/CDC6 on the *T. brucei* DNA content, and to understand its possible consequences for cell cycle regulation, flow cytometry was performed on PI-stained RNAi-induced versus non-induced cells for up to 18 hrs (Figure 3-14). FACS profiles of these cells showed that *TbORC1/CDC6* RNAi in the BSF resulted in a decrease in the proportion of cells with 2C and 4C DNA content (the former being depleted more rapidly), with a concomitant increase in the proportion of cells with 8C DNA

content. The increase in the 8C peaks implied that these cells re-replicated their nuclei following induction of *TbORC1/CDC6* RNAi. This observation seems consistent with the accumulation of multinucleate cells observed from DNA DAPI staining above. Again consistent with the lack of observable 0N1K zoids observed in the DAPI staining, and contrasting with their pronounced accumulation in PCF cells, no significant accumulation of cells of <2C content was seen in the BSF RNAi FACS.

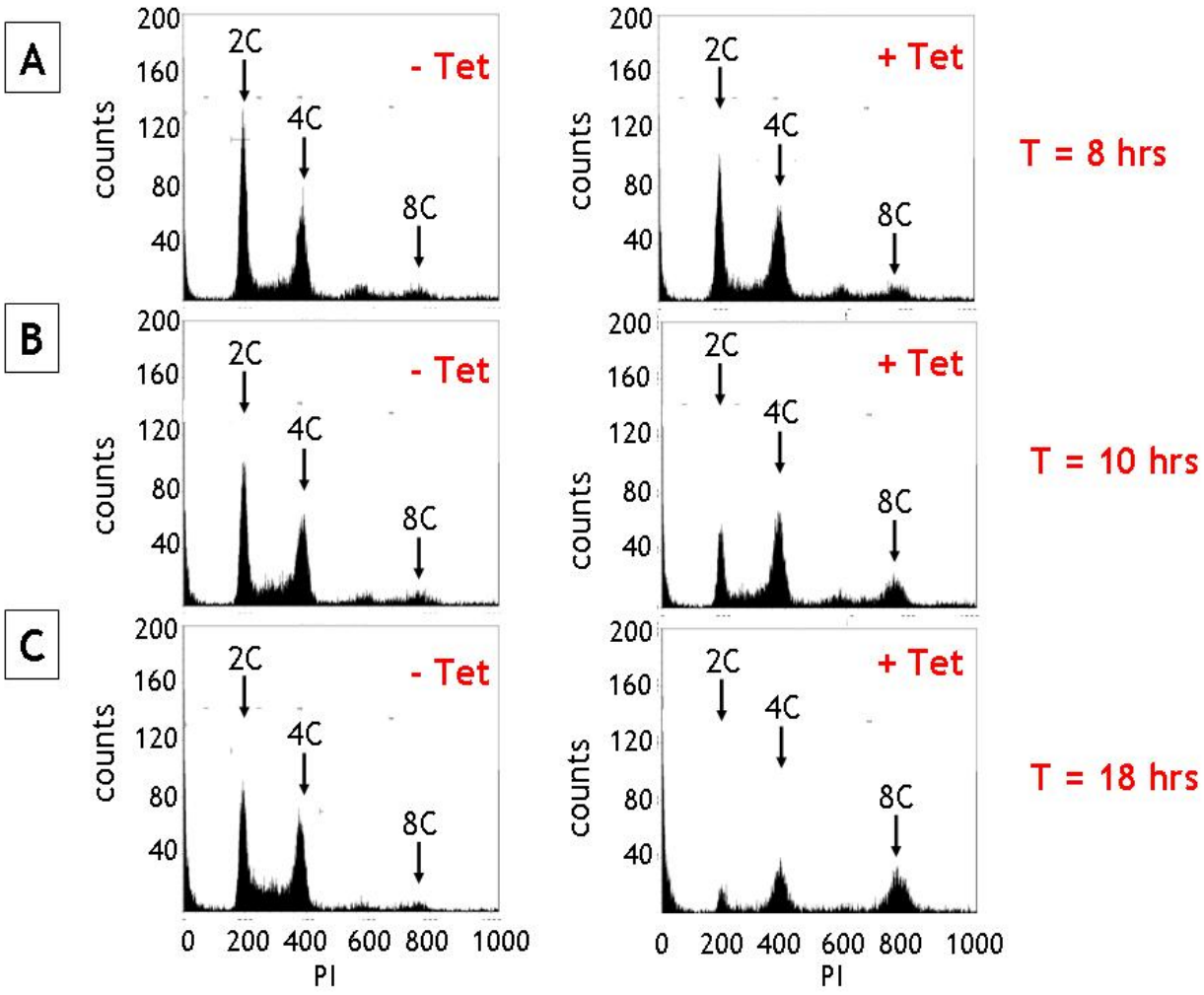


Figure 3-14 - FACS analysis of BSF *TbORC1/CDC6* RNAi cells

FACS analysis was performed on a single clone at 8 hrs, 10 hrs, and 18 hrs post induction. *Left panel* = non-induced cells and *Right panel* = tetracycline- induced cells.

3.7 Does TbORC1/CDC6 have broader functions beyond DNA replication?

Because Orc1 has been reported to initiate and maintain the formation of heterochromatin and hence transcriptionally silent loci in budding yeasts (Foss *et al*. 1993; Bell *et al*. 1993), we next used the RNAi cell lines generated and examined the possibility that TbORC1/CDC6 might play a similar role in regulation of *VSG* gene expression in procyclic and bloodstream form trypanosomes.

3.7.1 Effect of TbORC1/CDC6 RNAi on BVSG expression in Lister 427 PCF T. brucei

According to Hertz-Fowler *et al*, there are 14 different *VSG* genes present in 15 unique telomeric bloodstream expression sites (BES) in the Lister427 strain of *T. brucei* (Hertz-Fowler *et al*. 2008). Out of the 15 BES, only 1 is transcriptionally active in the BSF cells, and it is not fully understood how the cell selects only one BES out of the many. In the Lister 427 cell line used for RNAi analysis, the active BES (BES1) contains *VSG221* (*VSG221/427-2*) (Hertz-Fowler *et al*. 2008). In PCF cells, all the BES are transcriptionally silenced to an equivalent level and the VSG coat is replaced with procyclin (Chaves *et al*. 1999)

We first investigated whether depletion of *TbORC1/CDC6* by RNAi results in activation of transcription from the BES in PCF cells. To do this, we used quantitative real time RT-PCR to measure mRNA levels for 5 silent BES *VSGs*: *VSG13 (VSG427-13), VSG224 (VSG224/427-3), VSG800 (VSG800/427-18), VSGV02 (VSGVO2/427-9)* and *VSG221 (VSG221/427-2)*. This was performed 96 hours after RNAi induction using oligonucleotides CT_OL35/CT_OL36, CT_OL39/CT_OL40, CT_OL41/CT_OL42, CT_OL37/CT_OL38, and CT_OL33/CT_OL34, respectively. Oligonucleotides CT_OL07/CT_OL08 and CT_OL31/CT_OL32 were used for *TbORC1/CDC6* and *Tubulin* qPCRs. In two biologically independent experiments, with RNAi knockdown resulting in a reduction in *TbORC1/CDC6* mRNA levels by ~70 %, we detected a statistically significant ($P < 0.05$) increase between induced and non-induced samples in mRNA levels for *VSG221* (~2 fold increase),

VSG800 and *VSGv02* (each ~1.5 fold increase), but no significant differences for *VSG13, VSG224* and the chromosome-internal gene *tubulin* (Figure 3-15).

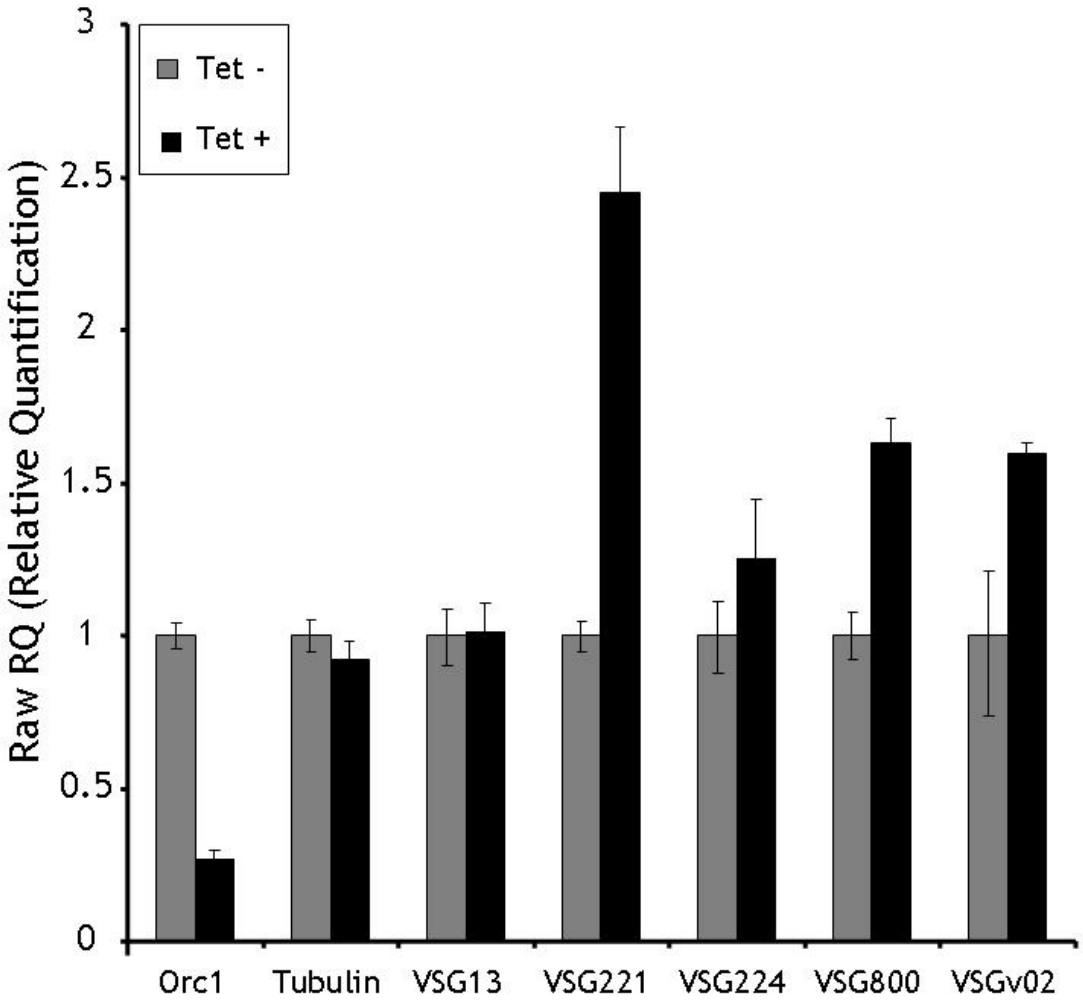


Figure 3-15 - Bloodstream *VSG* expression in procyclic form *T. brucei* after *TbORC1/CDC6* RNAi

Expression levels of genes were measured by quantitative real time RT-PCR, comparing mRNA levels for each relative to *GPI8* as endogenous control, before (Tet -, grey bar) and after induction (Tet +, black bar) of *TbORC1/CDC6* RNAi. In each case the level of mRNA of the uninduced sample is shown as 1 and the mRNA level in the RNA-induced sample shown relative to that. *TbORC1/CDC6* is indicated by Orc1 and *VSG*s are described in the results. Vertical lines indicate standard deviation from triplicates within the same biological sample.

3.7.2 Effect of TbORC1/CDC6 RNAi on BVSG gene expression in Lister 427 BSF T. brucei

During a tsetse blood meal metacyclic trypomastigotes are transmitted from the salivary glands of the fly to the bloodstream of the mammalian host, where the parasites differentiate to replicative long slender bloodstream forms.

Metacyclic form *T. brucei* express MVSG surface coat molecules until a few days post-entry into the mammalian bloodstream at which time, by a process that is not well understood, MVSG expression becomes attenuated with a concomitant activation of a single B*VSG* (Tetley *et al.* 1987). With the observation that some silent *VSG* genes are modestly derepressed in PCF cells upon RNAi depletion of *TbORC1/CDC6* (see above), we investigated whether the inactive BES also become activated after RNAi in established BSF cells expressing a single *BVSG*. To do this, we used to Lister 427 (90-13) bloodstream form *TbORC1/CDC6* RNAi cells described above (Section 3.6), and used quantitative RT-PCR to measure mRNA levels of the *BVSG* genes previously examined in Lister 427 PCF cells (Section 3.6). The same oligonucleotides used in Section 3.7.1 were used for this experiment. Of these *BVSG*s, *VSG221* is found in the single actively transcribed BES, and the others are found in silent BES. mRNA levels were compared in *TbORC1/CDC6* RNAi-induced versus non-induced samples relative to *Actin* as endogenous control, and 12 hours after RNAi induction (Figure 3-16), a shorter time relative to PCF cells given the severity of the RNAi growth phenotype in the bloodstream (see Section 3.6.3). Perhaps surprisingly (at $P < 0.05$), we detected a significant increase in transcript levels, of only 1.5 fold, for only one of the silent VSGs out of the five tested (expression of the housekeeping genes *GPI8* and *Tubulin* were also not significantly affected). The single gene with significantly greater expression was *VSG221*, in the actively transcribed BES. As before, RNAi depletion of *TbORC1/CDC6* transcript levels was lower than in procyclic form cells, at ~40 % of uninduced levels.

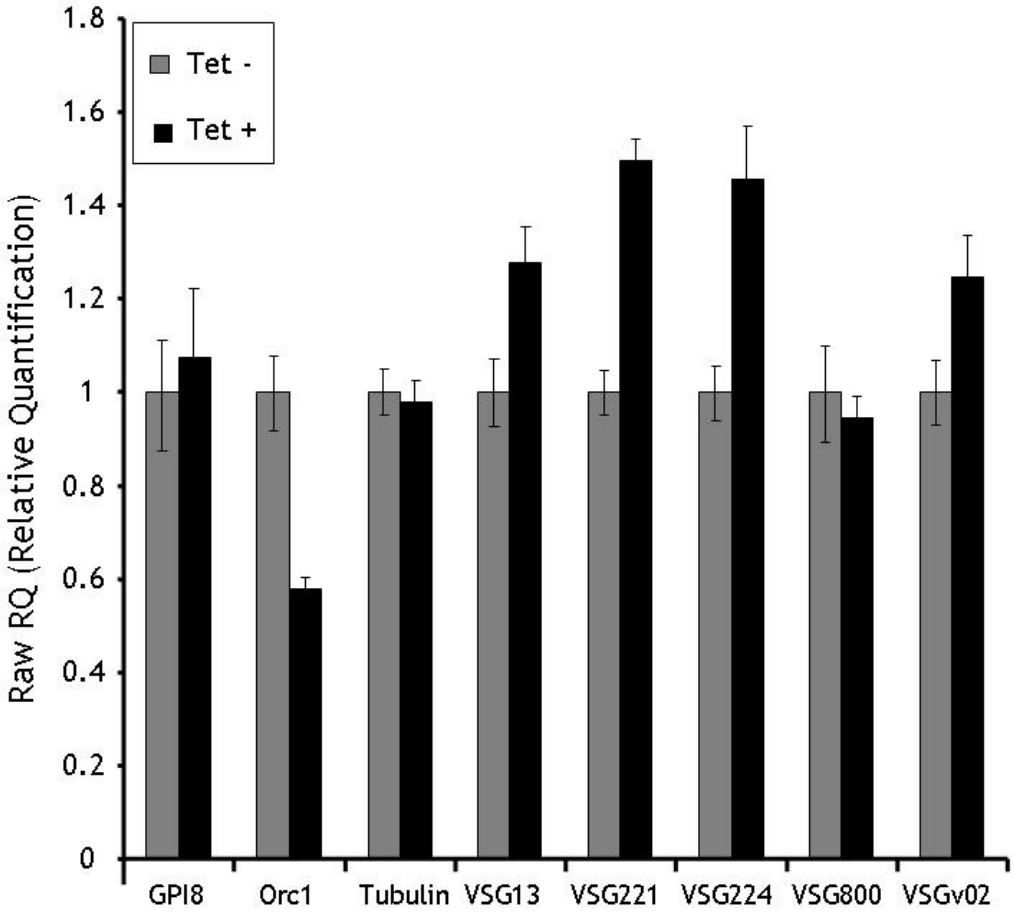


Figure 3-16 - Bloodstream *VSG* expression in bloodstream form *T. brucei* after *TbORC1/CDC6* RNAi

Expression levels of genes were measured by quantitative real time RT-PCR, comparing mRNA levels for each relative to *Actin* as endogenous control, before (Tet -, grey bar) and after induction (Tet +, black bar) of *TbORC1/CDC6* RNAi. In each case the level of mRNA of the uninduced sample is shown as 1 and the mRNA level in the RNA-induced sample shown relative to that. *TbORC1/CDC6* is indicated by Orc1 and *VSG*s are described in the results. Vertical lines indicate standard deviation from triplicates within the same biological sample

3.7.3 ***Effect of TbORC1/CDC6 RNAi on MVSG expression in EATRO 795*** *PCF T. brucei*

VSGs are expressed on the surface of *T. brucei* in metacyclic stage cells in the salivary gland of the tsetse fly prior to a mammalian infection: these are known as metacyclic VSGs (MVSGs). Like all surface coat molecules in trypanosomes MVSGs are expressed in a life-cycle stage-specific manner. Like *BVSG* genes, *MVSG* transcription is also controlled by RNA Pol I, but differs in that it is monocistronic, since the *MVSG* ESs do not contain *ESAG*s (Alarcon *et al*. 1994). The complete repertoire of *MVSG* ESs is not fully known as they require targeted cloning and sequencing, which thus far has been limited to the *BVSG* ESs (Hertz-

Fowler *et al*. 2008; Taylor & Rudenko 2006). So far, the promoter elements of *MVSG* ESs containing four *VSG*s have been characterised in the EATRO795 *T. brucei* strain: *MVSG1.22*, *MVSG1.61*, *MVSG1.63* and *MVSG1.64* (Ginger *et al*. 2002). Having observed statistically significant derepression of some *BVSG* transcripts in PCF cells upon RNAi knockdown of *TbORC1/CDC6*, we therefore extended this analysis to these *MVSGs* in the same life cycle stage. To do this, we utilised an EATRO795 pLew29pLEW13 RNAi line developed by M. Swiderski (WTCMP, University of Glasgow). This was transfected with the *TbORC1/CDC6* RNAi construct (Section 3.3) and two clones isolated. To determine if the RNAi effect was equivalent to that seen in Lister427 procyclic form cells, we analysed one clone by measuring growth for 144 hours. Although the *TbORC1/CDC6* RNAi EATRO795 cell line had almost equal growth with the uninduced samples (Figure 3-17A), qPCR showed that RNAi in this line resulted in a ~70 % knockdown of *TbORC1/CDC6* mRNA after 96 hours (Figure 3-17B), compared with a 90% knockdown observed in the Lister 427 cell line (Section 3.3.3). To then ask about the effect of *TbORC1/CDC6* RNAi knockdown on *MVSG* expression, we performed quantitative RT-PCR on three of the four *MVSGs*, 96 hours after induction (Figure 3-17C). This showed a statistically significant increase in mRNA levels of 6 – 12 fold for M*VSG 1.22*, 5 – 10 fold for M*VSG 1.61* and 6 – 13 fold for M*VSG 1.64*, when comparing RNAi-induced and non-induced samples relative to *GPI8* as an endogenous control. Thus, *TbORC1/CDC6* RNAi appears to have a greater derepression effect on *MVSG* genes than on *BVSG* genes in this life cycle stage. It is worth mentioning that more clones need to be tested with *Tubulin* run as an unrelated control relative to *GPI8* as an endogenous control.

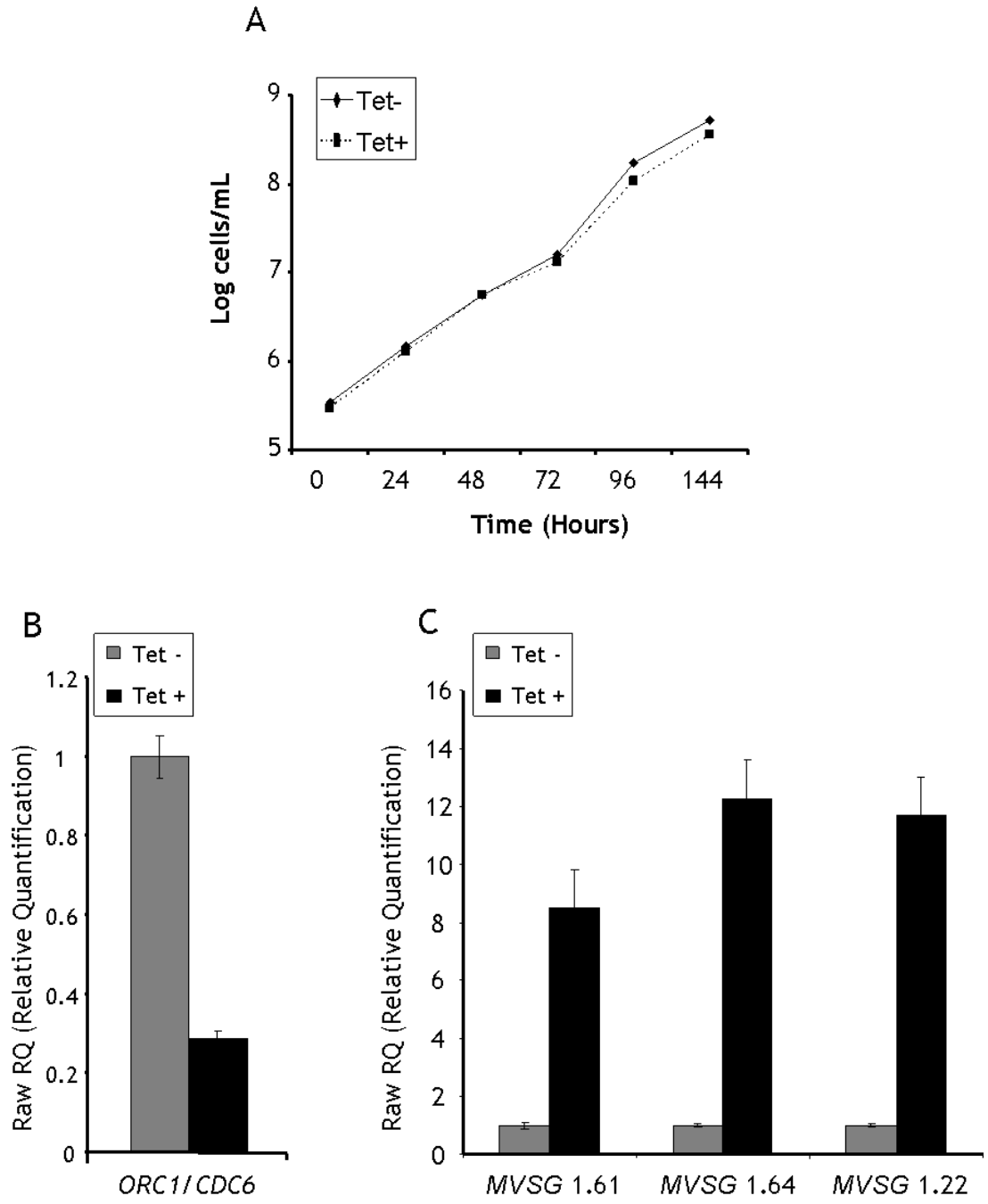


Figure 3-17 - Metacyclic *VSG* expression in procyclic form *T. brucei* after TbORC1/CDC6 RNAi

(A) Cell counts of *T. brucei* Lister 427 pLew13-pLew29 cells transformed with a *TbORC1/CDC6* RNAi construct are shown over time for 3 independent clones in the absence of *TbORC1/CDC6* RNAi induction (Tet-; solid lines) or following RNAi induction by the addition of 1 µg.ml^{-1} tetracycline (Tet+; broken lines). Expression levels of *MVSG* genes were measured by quantitative real time RT-PCR, comparing mRNA levels for each relative to *GPI8* as endogenous control, before (Tet -, grey bar) and after induction (Tet +, black bar) of *TbORC1/CDC6* RNAi. In each case the level of mRNA of the uninduced sample is shown as 1 and the mRNA level in the RNA-induced sample shown relative to that. *TbORC1/CDC6* mRNA knockdown levels is shown in (B) and *MVSG*s in (C). Vertical lines indicate standard deviation from triplicates within the same biological sample

3.8 Discussion

This chapter presents the results of functional characterisation of *T. brucei* ORC1/CDC6 by RNAi, showing a role in nuclear DNA replication, and preliminary data that suggest a further role beyond DNA replication.

3.8.1 RNAi of TbORC1/CDC6 defines an essential role in PCF

First reported in *Caenorhabditis elegans* as a mechanism to specifically interfere with endogenous gene function by injection of RNA into cells, RNAi has evolved as a powerful tool that has been widely used in a variety higher eukaryotes as well as parasitic pathogenic protozoa to rapidly assess gene function (Ullu *et al.* 2004; Fire *et al*. 1998). Since its inception in *T. brucei* RNAi has become the method of choice, in the first instance, to functionally characterise the effect of partial depletion of target genes. In this chapter, we have exploited this tool to investigate the function of a newly identified protein thought to control DNA replication in *T. brucei*. Before work in this chapter was begun, Michele Klingbeil and colleagues at the University of Massachusetts in Amherst had already performed RNAi of *TbORC1/CDC6* and presented her data at the Kinetoplastid Molecular and Cellular Biology meeting II (2007) in Woods Hole, USA. Being a collaborator of Richard McCulloch and Dave Barry (my supervisors), Michele Kinglbeil kindly shared her unpublished data with us, and it was agreed that corroboration of this data, by performing the same initial PCF RNAi analysis in Glasgow, was valuable. Unknown to us, a group in Brazil headed by Caroline Elias was working on the same project, and have since published their findings in 2009 (Godoy *et al.* 2009), Given this set of circumstances, I discuss below the overlapping data from all three labs, as well as unique aspects of my work, which suggest TbORC1/CDC6 is a key nuclear DNA replication factor. For the purposes of coherence, I provide a brief summary of the findings from Michele Klingbeil and colleagues and from Godoy et al (2009), and then focus the rest of my discussion on my data.

Klingbeil and colleagues showed that: (i) a TbORC1/CDC6-GFP C-terminal fusion was a nuclear protein; (ii) *TbORC1/CDC6* RNAi resulted in growth inhibition from 4 days after induction in PCF *T. brucei*; (iii) the pre-dominant population of cells

post *TbORC1/CDC6* RNAi induction was zoids (0N1K cells); (iv) BrdU incorporation, assayed microscopically, after *TbORC1/CDC6* RNAi led to a decrease in nuclear DNA synthesis with a subset of zoids still capable of synthesising kDNA. Based on these findings, Klingbeil *et al* concluded that TbORC1/CDC6's main role is in nuclear DNA synthesis. Godoy *et al* (2009) characterised ORC1/CDC6 from *T. cruzi* and *T. brucei* and showed that: (i) both proteins possess ATPase activity [a typical property of eukaryotic ORC proteins (Speck *et al.* 2005)] that increased in the presence of non-specific DNA; (ii) TbORC1/CDC6 and TcORC1/CDC6 were able to complement *S. cerevisiae* Cdc6 mutants, but not *S. cerevisiae* Orc1 mutants; (iii) *TbORC1/CDC6* depletion by RNAi resulted in zoid formation in *T. brucei* PCF cells; (iv) using antiserum against TcORC1/CDC6 and TbORC1/CDC6, both proteins localised to the nucleus and remained attached to chromatin throughout the cell cycle. Based on these data Godoy *et al* concluded that, unlike higher eukaryotes (yeasts and humans, for example), trypanosomes utilise a simplified pre-replication machinery as previously described for archaea (Godoy *et al.* 2009).

Based on database mining and phylogenetic analyses, we (Godoy *et al*, Klingbeil et al, and us) all hypothesise that *T. brucei*, *T. cruzi* and *L. major* possess a single protein that is homologous to eukaryotic Orc1 and Cdc6 proteins, and defines nuclear DNA replication origins by providing the activities of each. It is important to stress, however, that although BLAST searches using all ORC subunit homologues from yeast, humans and *Arabidopsis* have been unable to find trypanosomatid proteins with domains that match other members of the eukaryotic ORC family of proteins, such proteins might be present, but severely evolutionarily divergent and hence only identifiable through direct experiments, such as immunoprecipitation (see chapter 4). However, what is clear is that the *ORC1/CDC6* homologues in each trypanosomatid show synteny across the fully sequenced genomes (Aslett *et al*. 2010), consistent with a critical function within the group. Given the homology between ORC1 and CDC6 in all eukaryotes (Kawakami & Katayama 2010; Duncker *et al.* 2009), and the ability to detect only trypanosomatid ORC1/CDC6 by BLAST searches with either gene, it seems likely that, like in archaea (De Felice *et al*. 2006; Grabowski & Kelman 2003), this singular trypanosomatid gene encodes a protein with ORC1 and CDC6 functions. Indeed, this may not be unique to trypanosomatids as a single *ORC1/CDC6* gene

has also been described in *Giardia lamblia* (Morrison *et al*. 2007), and we detect the same in *Entamoeba histolytica* (see chapter 4). Whether this reflects an ancestral eukaryotic state remains unknown, but the ability of TcORC1/CDC6 to complement the absence of CDC6 when expressed in yeast CDC6 mutants (Godoy *et al*. 2009) provides evidence of Cdc6 function in the trypanosomatid protein.

If the hypothesised nuclear DNA replication function for ORC1/CDC6 is true, the gene should be essential in trypanosomatids and RNAi knockdown should result in a complete block of nuclear DNA synthesis and rapid cell death. This has been tested now in *T. brucei*, and our work and others (Klingbeil et al, unpublished; and Godoy *et al,* 2009) show, surprisingly, that depletion of *TbORC1/CDC6* mRNA in procyclic form cells does not immediately lead to cell death, but instead normal growth continues for up to 3-4 days post induction of RNAi. Why could this be the case? For any cell to be successfully propagated, its DNA content must be duplicated in S-phase, segregated in M-phase and then during cytokinesis passed onto its progeny. In budding and fission yeasts to ensure that DNA is replicated only once per cell cycle and to avoid re-replication, sites of initiation of DNA replication are licensed in G_1 and only at these sites does DNA synthesis occur in S-phase (Ogawa, Takahashi, & Masukata 1999; Aparicio, Weinstein, & Bell 1997; Santocanale & Diffley 1997). We know from experiments by Godoy *et al* that, like yeast Orc1, TbORC1/CDC6 remains bound to chromatin throughout the cell cycle (Godoy *et al*. 2009). It also known that ORC in yeasts is not targeted for degradation after S phase, but rather its activity is regulated by post-translational modification events that release it from chromatin in mitosis and, upon mitotic exit, it is able to re-bind origins (Tsakraklides & Bell 2010; DePamphilis 2005). Any such modifications have not been explored in *T. brucei*. Nevertheless, based on these observations, we can speculate that the equal growth observed before 72 hrs (approximately 6-8 cell divisions *in vitro*) for both *TbORC1/CDC6* RNAi-induced and non-induced cells results from incomplete depletion of TbORC1/CDC6 protein in the cell, meaning that there is sufficient protein in these cell divisions to bind all origins, or sufficient numbers of origins defined to continue effective replication. Only once all, or most, TbORC1/CDC6 is depleted does a growth defect become apparent, because the cell can no longer initiate sufficient DNA synthesis. Consistent with this, Godoy et al (2009)

show that TbORC1/CDC6 protein is still detectable up to 7-8 days post-RNAi induction (we have not examined this in our cell lines).

Our FACS data, and that of Godoy *et al* (2009) and Klingbeil *et al* (unpublished), showed that *TbORC1/CDC6* RNAi depleted cells have an altered nuclear DNA content (the contribution of kinetoplast DNA to the PI-stained population in FACS is considered negligible), consistent with a perturbation in cell cycle progression. *T. brucei* follows the conventional G1-S-G2-M cell cycle transition, with the replication of kinetoplast DNA occurring in temporal and partial synchrony with the replication of the nuclear genome (Woodward & Gull 1990), although all the events co-ordinating this cell division are not completely understood (Hammarton 2007). From Figure 3-4, the population of cells labelled 2C are in G_1, 4C cells with double nuclear DNA content are in G_2, and S-phase cells are between the G_1 and G_2 peaks. Like wild type cells, non RNAi-induced G_1 cells transit nuclear DNA synthesis in S-phase to enter G_2 and, following mitosis and cytokinesis, revert to G_1. In the case of *TbORC1/CDC6* RNAi cells, we infer that this routine is interrupted, since an almost complete disappearance of 4C cells is observed 120 hrs post-RNAi induction. In addition to this, there is an obvious reduction in 2C cells, and ultimately a majority of the cells have less than 2C nuclear DNA content 5 days post-RNAi. These data suggest that TbORC1/CDC6 depletion leads to inhibition of G_1 to S transition, consistent with loss of TbORC1/CDC6 leaving origin DNA sites free from pre-licensing and hence no subsequent replication in S-phase. The additional loss of 2C cells and accumulation of <2C cells suggest further cell cycle abnormalities, discussed below.

To complement the FACS data and further understand the potential role of *TbORC1/CDC6* in the regulation of nuclear DNA replication in procyclic form *T. brucei*, microscopic counting of DAPI-stained DNA was carried out and quantified by us and by Godoy *et al* (2009) and Klingbeil *et al* (unpublished). Progression of the cell cycle in *T. brucei* involves the ordered and coordinated duplication of two genomes; the kinetoplast and the nuclear genomes. However, in cases of aberrant DNA synthesis and/or segregation of either genome, progeny could inherit unusual N-K configurations, such as 0N1K (anucleate cells, or zoids), 1N0K, 2N1K, or >2N1K cells. Such aberrant cells usually maximally constitute ~3.5 % of a wild type population (Woodward & Gull 1990). RNAi of *TbORC1/CDC6*

had a deleterious effect on nuclear synthesis, but no effect on kDNA synthesis, since cells lacking kDNA DAPI staining were never seen. In contrast, *TbORC1/CDC6* RNAi virtually exclusively resulted in the generation of 0N1K zoids. In agreement with the growth phenotype and FACS analysis, zoids did not appear in the *TbORC1/CDC6* RNAi-induced cell population until 72 hrs. Zoid appearance was primarily concomitant with a reduction of 1N1K cells, though 2N2K cells appeared less prevalent relative to non-RNAi induced parasites, whereas 1N2K cells numbers appeared unchanged. Put together, these data support previous reports that there is apparently little coordination between nuclear DNA synthesis and cell division in PCF *T. brucei* (discussed below). Hence, we suggest that *TbORC1/CDC6* RNAi inhibits nuclear but not kinetoplast DNA synthesis, but the cells can still proceed through to cytokinesis, resulting in the accumulation of zoids from division of 1N2K cells.

The dissociation of nuclear DNA synthesis, nuclear mitosis and cytokinesis in *T. brucei* has been reported for a number of cell cycle proteins silenced by RNAi or following drug treatment of PCF cells. RNAi of cyclins CycE1/CYC2 and Cyc B2, a polo-like kinase, PLK, and cdc2-related kinase-1s, CRK1 and CRK3, all result in the production of anucleate cells in PCF *T. brucei* (Kumar & Wang 2006; Hammarton, Engstler, & Mottram 2004; Hammarton *et al.* 2003; Li & Wang 2003). Studies by Ploubidou *et al* (1999) used 60 μM aphidicolin to inhibit DNA synthesis by DNA Pol α (hence nuclear DNA synthesis) and showed that cells become blocked in S-phase, leading to the formation of zoids. This treatment also resulted in the complete absence of 2N2K cells in the population and an increase in 1N2K cells, suggesting that the zoids were the products of division of the IN2K population. In the same study, an anti-microtubule polymerisation inhibitor, rhizoxin, also resulted in the formation of zoids. From both experiments Ploubidou *et al* concluded that neither inhibition of nuclear DNA synthesis nor inhibition of mitosis has any effect on *T. brucei* cytokinesis (Ploubidou *et al*. 1999). Given that zoid formation also resulted from RNAi of all the above cell cycle regulators, the question arises if a general cell death pathway initiates in PCF cells once nuclear DNA synthesis or nuclear mitosis is inhibited, or does the data indicate a role of the cell cycle proteins in nuclear DNA synthesis? It seems possible that once procyclic form cells sense a perturbation with cell cycle progression, nuclear DNA synthesis and consequently mitosis are shut down,

although progression to cytokinesis is not: hence the common production of anucleate cells. Therefore, the suggestion by Godoy *et al* (2009) that zoid formation upon RNAi of *TbORC1/CDC6* implies a singular involvement of the protein in nuclear DNA synthesis may be flawed, and other roles cannot be ruled out. Godoy *et al* showed that the protein has ATPase activity, is localised to the nucleus and binds chromatin. Does that mean every protein with ATPase activity (at least ~ 135 in *T. brucei* identified in the TritrypDB database), that binds chromatin (at least ~ 110 in *T. brucei* by Tritryp DB), or that produces zoids after RNAi is involved in nuclear DNA replication? Like in Klingbeil *et al*'s analysis of TbORC1/CDC6 work and Ploubidou *et al* (1999) an assay, such as BrdUTP incorporation, is necessary to measure DNA synthesis and distinguish between cells which die by zoid formation because of cell cycle inhibition and death by the same route because nuclear DNA replication is inhibited.

3.8.2 Assaying BrdU incorporation shows that TbORC1/CDC6 RNAi inhibits nuclear DNA synthesis

Having established by FACS and microscopy that RNAi knockdown of TbORC1/CDC6 led to inhibition of nuclear DNA replication and led to zoid formation in PCF cells, we wondered if the persistence of some normal cells (for instance, 40 % of cells were 1N1K) even 120 hrs post RNAi induction was due to incomplete TbORC1/CDC6 depletion, and so tested this by BrdUTP incorporation. Incorporation of the thymidine synthetic analogue 5-bromo-2′-deoxyuridine (BrdU) has become a common route to study S-phase events in living cells (Nowakowski, Lewin, & Miller 1989; Gratzner 1982). Cells that are actively replicating their DNA, when incubated with BrdU and deoxycytidine, can be visualised by immunofluorescence using fluorescently-labelled monoclonal antibodies specific for BrdU. Variations in fluorescent intensities due to different immunofluorescent detection methods, the requirement of high magnification microscopes, and the time-consuming process of fixing and labelling cells are limitations in assessing and discriminating BrdU positive and negative cells by such approaches. We therefore used a novel dot blot technique to assess and quantify DNA synthesis by BrdU incorporation in whole genomic DNA isolated from parasite populations before and after induction of *TbORC1/CDC6* RNAi (Ueda *et al*. 2005). This approach demonstrated that 96 hrs post induction of

TbORC1/CDC6 RNAi, all living cells in the population could not make new DNA, since the amount of BrdU incorporated was essentially indistinguishable from background. This is consistent with our FACS and microscopic analyses, and suggests that TbORC1/CDC6 depletion was sufficient for abrogation of *T. brucei* DNA replication.

3.8.3 TbORC1/CDC6 seems to be an essential gene

Although RNAi has become the primary forward genetic tool to analyse gene function in trypanosomes, classical reverse genetic approaches such as targeted replacement of genes still remain the ultimate proof to test essentiality of genes for viability. Being a diploid organism, for single copy *T. brucei* genes both alleles must be deleted using conventional homologous recombination of regions flanking the gene of interest, which are replaced by selectable drug resistance markers. From a wild type cell line, a heterozygote null is first generated with one drug resistance marker, and using this line, a second marker is then used to attempt to delete the second allele. If the second allele cannot be deleted, this may suggest that the gene has an essential function, since both copies cannot be lost. A *TbORC1/CDC6* heterozygous PCF cell line, with one allele replaced by a *PUR* cassette, was successfully generated. However, three separate attempts aimed at removing the second allele of *TbORC1/CDC6* failed. In all three attempts, using a PCR-based approach, no PCR products could be generated for correct integration of the *BSD* cassette at the *TbORC1/CDC6* locus when the *PUR* cassette was also present. Collectively, these results establish that there are, as expected, two alleles of *TbORC1/CDC6* and that one can be disrupted without effecting cell viability. Here, we have only shown a PCR-based approach to verify that the *TbORC1/CDC6* locus was targeted, but Southern blot analyses have also been carried out to confirm the results from the PCR (see chapter 4).

A definitive technique to confirm essentiality would be to carry out a conditional gene knock-out strategy, first established in *T. brucei* by the lab of G.A.M Cross (Wirtz *et al.* 1999). Here, after one allele of the target gene has been replaced by a drug resistant cassette, the wild type copy of the gene is tetracycline-inducibly expressed ectopically from the rRNA spacer. Once the heterozygote null recombinant parasites of the gene is obtained that can conditionally restore

expression of the wild type copy, then the second allele of the gene can be deleted with a second drug resistant marker. Removal of tetracycline to shut down ectopic expression of the gene in a homozygous null background should lead to cell death, if the gene is essential for viability. This approach has been used to test the essentiality of a number of genes in *T. brucei* (Martin & Smith 2006; Martin & Smith 2005; Helfert *et al*. 2001; Wirtz *et al*. 1999; Ochatt *et al*. 1999). With hindsight, this would have been the approach adapted to ultimately prove that *TbORC1/CDC6* is an essential gene.

3.8.4 TbORC1/CDC6 is a nuclear protein

Immunolocalisation and fluorescent imaging of C-terminally GFP– or Myc– tagged TbORC1/CDC6 localises the signal to the nucleus, consistent with data from Klingbeil *et al* (unpublished) and Godoy *et al* (2009). In later work, furthermore, we show that the Myc-tagged variant is functional (see chapter 4). In conjunction with FACS and DAPI-staining microscopy data, and attempted knockout experiments, these data are consistent with an essential role for TbORC1/CDC6 in procyclic form nuclear DNA synthesis.

3.8.5 TbORC1/CDC6 is important for BSF T. brucei life cycle progression

In the above analyses, and in the work of Klingbeil *et al* and Godoy *et al*, the effect of *TbORC1/CDC6* RNAi was limited to insect stage *T. brucei* procyclic form cells and demonstrated clearly that it is an essential gene involved in nuclear DNA synthesis. What about the mammalian-infective stages, the bloodstream form cells, which are clinically relevant in terms of drug discovery? Does TbORC1/CDC6 perform the same function in both life cycle stages? As shown in the Section 3.6.2, induction of *TbORC1/CDC6* RNAi inhibits parasite growth very rapidly in the BSF, since after 8 hrs (~1 cell division) post-induction no increase in cell density was observed. This was followed by an accumulation of post-S cells with no apparent manifestation of a cleavage furrow to indicate the onset of cytokinesis. The impairment of cytokinesis in these cells appeared to be associated with continued synthesis and replication of kinetoplast and nuclear DNA, since multi-nucleate, multi-kinetoplast cells accumulated from 8 hrs

onwards, with the nuclei eventually becoming condensed and non-discrete 24 hrs post RNAi induction. However, despite the rapid effect of RNAi on DNA content, knockdown levels measured by quantitative RT-PCR only revealed a 40 % depletion of *TbORC1/CDC6* transcripts 12 hrs post-induction. Even at this level of RNAi, >80 % of the cells in the RNAi-induced population after 24 hrs had aberrant DNA configurations. The rapidity of the onset of growth arrest, the extent of aberrant cells observed and the lack of anucleate cells in the BSF all appear to contrast with the *TbORC1/CDC6* RNAi phenotypes seen in PCF cells. Looking at the literature there is substantial evidence to support the concept that there are differences in the regulation of cell cycle events between BSF and PCF *T. brucei*. A few examples of proteins whose RNAi knockdown show differences between bloodstream and procyclic forms are described below.

In all studied eukaryotes, Mob proteins are known to bind and up-regulate the activity of members of the NDR (nuclear Dbf2-related) subfamily of kinases (Hergovich *et al*. 2006). In budding yeasts, they have been shown to control mitotic exit and entry into cytokinesis via their interaction with Dbf2/Dbf20 (Mah, Jang, & Deshaies 2001). In *T. brucei* there are two Mob1 proteins: MOB1-A and MOB1-B. When depleted by RNAi both proteins exhibit an increase in abnormal cells (Hammarton *et al*. 2005), usually seen as >2N>2K configurations in the BSF 24 hrs post RNAi induction. The MOB1 RNAi-induced cells also showed prematurely terminated cytokinesis, as observed by a block in the later stages of cytokinesis which occurred in a subset of post-mitotic cells (Hammarton *et al*. 2005). It is worthy of note that MOB1 in BSF *T. brucei* cells does not localise to the nucleus at any stage in the cell cycle. Although the same localisation experiments were not done in PCF cells, RNAi of MOB1-A and MOB1-B in this life cycle stage resulted in the production of zoids. The authors concluded that there is differential regulation of the cell cycle in the two life cycle stages. How do these findings compare with TbORC1/CDC6 RNAi? Clearly, there are broad similarities in terms of RNAi phenotypes of both TbORC1/CDC6 and TbMOB1; indeed, in both studies very similar time points were used for examination of samples, implying that the results are very comparable. However, localisation of MOB1 proteins was analysed in BSF cells and not in PCF cells, while localisation of TbORC1/CDC6, in all 3 studies, was done in PCF cells and not in the BSF. Due to time constraints in this project we have not looked at the localisation of

TbORC1/CDC6 in the BSF. Although it is possible that TbORC1/CDC6 might have other roles in BSF cells that may not restrict it to the nucleus, without experimental proof this is speculation.

Structural Maintenance of Chromosome (SMC) proteins are known in eukaryotes to play essential roles in DNA repair and recombination, as well as in chromosome segregation (for review see (Schubert 2009)). An example of the SMC class of proteins is the cohesin complex, whose function is to hold duplicated sister chromatids together until segregation. The cohesin complex proteins are conserved in *T. brucei* and characterisation of the cohesin subunit I (SCC1) by RNAi has been carried out by Gluenz and colleagues (Gluenz *et al*. 2008). In BSF cells, TbSCC1 localises to the nucleus during S-phase and G_2 and is absent in G_1 and M-phases of the cell cycle. Although phenotypes arise much slower than observed for TbORC1/CDC6, RNAi of *TbSCC1* also results in zoid formation in PCF cells and in multikinetoplast, multinuclei and multi-flagellated cells in the BSF. Here, therefore, is a protein that is localised to the nucleus at least for part of the cell cycle and whose RNAi phenotypes are analogous to that observed for *TbORC1/CDC6* RNAi in both BSF and PCF cells.

In higher eukaryotes the function of cyclin-dependent kinases (CRK) as key regulators of cell cycle progression is well established [for review see (Pines 1995) and references therein]. Recently, in *T. brucei,* RNAi studies on a number of cyclins and their corresponding kinases has begun to shine a light on our understanding of their roles as key cell cycle regulators in both PCF and BSF *T. brucei* cells (Hammarton *et al*. 2004; Tu & Wang 2004; Li & Wang 2003). In two independent studies the cyclin CYC6 and its corresponding kinase CRK3 have been shown to show stage-specific differences in cell cycle controls when analysed by RNAi (Tu & Wang 2004; Hammarton *et al*. 2003). Although in both studies the localisation of the proteins are not established in any of the parasite life cycle stages, it is very likely that at least for part of the cell cycle these proteins will be in the nucleus if they are *bona fide* cell cycle regulators. With this assumption in mind, depletion of *TbCYC6* by RNAi generates zoids in PCF cells, whereas in BSF cells multikinetoplast and multinucleat cells are formed. Not surprisingly, the same phenotypes are observed upon RNAi knockdown of *TbCRK3*, except that the multiple nuclei in BSF cells remain condensed to each other. Both papers conclude that there are fundamental differences in cell cycle

regulation in PCF and BSF cells. Again, these phenotypes are broadly similar to the observed differences for *TbORC1/CDC6* RNAi in the two life cycle stage forms, though not directly comparable due to lack of localisation experiments.

Taking the above data together, we can propose that in *T. brucei* cells, BSF-specific cell cycle checkpoints become activated in a number of contexts that halt cell division, or at least cytokinesis. For TbORC1/CDC6 such a response occurs quickly after RNAi, even before TbORC1/CDC6 levels are reduced sufficiently to halt DNA replication. Three alternative hypotheses are possible. One is that *T. brucei* expresses a BSF-specific factor that is accidently targeted by the TbORC1/CDC6 RNAi, initiating an off-target effect that exceeds the small depletion of TbORC1/CDC6 (40 % decrease in mRNA levels). Second, there may be a BSF-specific factor that interacts with TbORC1/CDC6, or a direct BSF-specific function of TbORC1/CDC6, whose importance to *T. brucei* growth in this life cycle stage supersedes the role of TbORC1/CDC6 as the orchestrator of nuclear DNA replication. Thirdly, might depletion of essential proteins by RNAi in BSF cells always produce a stereotyped phenotype of multinucleate and mutikinetoplast cells, implying a general cell death phenomenon in BSF? Examples of protein depletion by RNAi that fail to induce these phenotypes appear to rule out this last possibility (see below).

In a quest to understand the mechanisms that lead to cell death in BSF *T. brucei*, Worthen *et al* (2010) used RNAi to knockdown the large and small subunits of topoisomerase IB (TOP1BS and TOPIBL respectively), mitochondrial topoisomerase II (TOPIImt) and a nucleolus localised protein, NOPP44/46. Knockdown of TOP1BS and TOPIBL resulted in zoid formation, knockdown of TOPIImt resulted in accumulation of 1N0K cells, while knockdown of nuclear NOPP44/46 resulted in loss of 2N cells and accumulation of >4N cells. Coupled with a variety of drug treatments, the authors suggest that there is a limited number of pathways that lead to cell type in BSF cells, and whether this involves the disruption of biochemical pathways and/or depends on underlying genetic variation still remains to be investigated (Worthen, Jensen, & Parsons 2010). Nevertheless, it is clear that the formation of multinucleate and multikinetoplast cells, as seen following RNAi of TbORC1/CDC6 and a number of cell cycle regulators, is not a default RNAi phenotype.

The differences in *TbORC1/CDC6* RNAi phenotypes between PCF and BSF cells may, of course, have simpler explanations. It may simply be down to differences in metabolic rates or in generation times (~ 6 hrs for BSF cells, and ~ 11 hrs for PCF). Differences in protein abundance is also possible, though no significant difference in mRNA expression levels of TbORC1/CDC6 between BSF and PCF *T. brucei* have been reported, at least based on RNA-Seq data available on TriTryDB (Aslett *et al*. 2010; Siegel *et al*. 2010). Further work will therefore be needed to explain this, but one non-DNA replication function of TbORC1/CDC6 that may be key to this is discussed below.

3.8.6 Regulation of BSVG and MSVG gene expression by TbORC1/CDC6

ORC proteins mediate the formation of heterochromatin, for instance at the silent mating loci *HML and HMR* in yeasts, and at telomeres and subtelomeres via interactions with Telomere repeat binding factor 2 (TRF2), heterochromatin protein 1 (HP1), and histone H3 trimethyl K9 (H3 K9me3) in *Drosophila* and mammalian cells (Deng *et al*. 2009; Sasaki & Gilbert 2007; Lidonnici *et al*. 2004). Also, in the malaria parasite *Plasmodium falciparum*, where only 1 out of ~ 50 *var* genes is expressed at one time, Orc1 in conjunction with Sir2 has been shown to be involved in telomeric silencing (Mancio-Silva *et al*. 2008). Except in *P. falciparum,* eukaryotic Orc1's role in silencing is thought be orchestrated by the presence of an N-terminal BAH domain (reviewed in [Sasaki & Gilbert 2007)]. Although *T. brucei* ORC1/CDC6 lacks the BAH domain, sub-telomeric localisation revealed by chIP-chip data (see Chapter 5) prompted us to investigate if TbORC1/CDC6 played additional roles in PCF and BSF cells beyond DNA replication.

Mapping TbORC1/CDC6 binding along the *T. brucei* megabase chromosomes showed high density clusters of binding sites at telomeric and subtelomeric loci, overlapping with sites which are known to contain *BVSG* and *MVSG* expression sites (see Chapter 5). To investigate the potential function of TbORC1/CDC6 binding at these loci we asked if depletion of the protein by RNAi had an effect on *BVSG* or *MVSG* expression. In Lister 427 PCF cells, RNAi had limited effect on *BVSG* mRNA levels (3/5 genes tested showed derepression, of 1.5-2.7 fold), and

no discernible effect on *BVSGs* in Lister 427 BSF cells (no silent genes showed derepression). In contrast, in EATRO795 PCF *T. brucei*, RNAi of TbORC1/CDC6 led to substantial (5-13 fold) derepression of all *MVSG*s tested. On first look this is an interesting observation, but upon closer scrutiny several questions arise; why is the effect of *TbORC1/CDC6* RNAi different at BES and MES, and perhaps in different life cycle stages? Are these due to differences in the structure of the two ES classes, or due to differential regulatory mechanisms acting on them or in the two life cycle stages?

Recently, Hertz-Fowler and colleagues used transformation–associated recombination cloning (TAR clone) to analyse the complete collection of BES from the Lister 427 strain of *T. brucei* (Hertz-Fowler *et al.* 2008). Their results confirm a number of differences from the MES (Ginger *et al.* 2002). Figure 3-18 highlights the major differences in the architecture of the MES and BES. Briefly, Hertz-Fowler showed that *BVSGs* are located 40 – 60 kb downstream of their promoters, between which are located a number of expression site-associated genes (*ESAG*s) and pseudogenes that are co-transcribed with the *VSG*. *ESAG 7, ESAG6,* and *ESAG5* are located closest to the promoter and are thus furthest from the *VSG* ORF. Between the most telomere-proximal *ESAG* (*ESAG1*) and the *VSG* ORF is an array of 70 bp repeats whose length is variable depending on the BES (Hertz-Fowler *et al.* 2008). On the other hand, the MES show a much simpler design than the BES (Ginger *et al.* 2002; Kim & Donelson 1997; Son *et al*. 1989). Eight MES characterised showed gross architectural similarities amongst each other with a number of differences compared to BES (Figure 3-18 A and B). The differences include: (a) much reduced numbers, and even absence, of 70-bp repeats between the promoter and the *MVSG* ORF; and (b) the complete absence of *ESAGs* between the promoter and the *MVSG* ORF (though some ESAGs are located upstream of the MVSG promoter, alongside a number of ESAG pseudogenes and *ingi* retrotransposons) (Ginger *et al.* 2002).

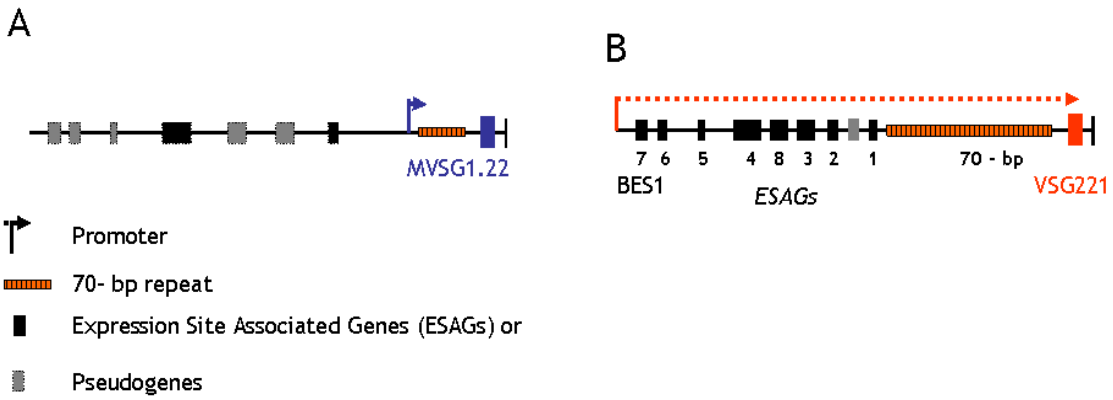


Figure 3-18 – Architecture of *T. brucei* expression sites

(A) Metacyclic *VSG* expression site (MES) and (B) bloodstream *VSG* expression site (BES) show structural differences in: (i) location of the promoter relative to the VSG; (ii) number of 70 bp repeats between promoter and VSG; and (iii) position and number of expression site associated genes (ESAGs). Functional *ESAGs* are shown by black boxes, pseudigenes are shown as grey boxes, arrow heads indicate the position of the promoter, and coloured boxes show the ORF of either *BVSG* or *MVSG* shown below it. Figure made by R. McCulloch – copied with permission. BES structure originally in (Hertz-Fowler *et al*. 2008)

Like *BVSGs* in the mammal bloodstream, *MVSGs* are expressed in a monoallelic fashion in the tsetse salivary glands (Alarcon *et al*. 1994), and in neither case is it fully understood how this is achieved. Unlike protein coding genes, which are transcribed by RNA Pol II, both types of *VSG* gene are transcribed by RNA Pol I (Gunzl *et al*. 2003), though the BES and MES promoters are not significantly homologous to each other (Ginger *et al*. 2002). Navarro and Gull showed that RNA Pol I localises to a single, extranucleolar site in the nucleus of BSF cells, termed the expression site body (ESB), and thus they postulate that this is limited to containing one BES and therefore drives monoallelic expression (Navarro & Gull 2001). It is unknown if an equivalent structure is found in metacyclic cells and associated with MES expression, but it is not observed in PCF cells, when no VSG are normally expressed (Navarro & Gull 2001). If the ESB hypothesis is true, is the single ESB primarily necessary and sufficient to render the other 14 sites transcriptionally silent? There is accumulating evidence in the literature to suggest that epigenetic phenomena are also involved in maintaining transcriptionally silent chromatin at inactive BES. Deletion of *TbDOT1B*, and RNAi of *TbISWI*, *TbSCC1* and *TbRAP1*, have all been shown to cause derepression of silent BES to varying extents in PCF and BSF (Yang *et al*. 2009; Landeira *et al*. 2009; Figueiredo *et al*. 2008; Hughes *et al*. 2007).

The first experiments to show that derepression of BES could involve changes in local chromatin structure came from the lab of G. Rudenko, by studying the chromatin remodelling protein TbISWI through RNAi. Using fluorescent gene reporter assays integrated downstream of BES promoters they showed that depletion of TbISWI in PCF cells led to 10 -17 fold depression of the reporter, while in BSF depletion resulted in 30 – 60 fold derepression. Silent BES activation was measured either by the reporter gene or by measuring transcript levels for *ESAGs* located downstream of the reporter. Interestingly, in both PCF and BSF cells RNAi caused derepression of the reporter gene and *ESAGs* proximal to the promoter (*ESAG 5, 6, and 7*), while those genes distant from the promoters, including the *BVSGs*, were not derepressed. They concluded that BES activation/silencing is a multi-step process that involves regulation of the activity of the promoter occurring separately from regulation at the level of transcription, which involves regulation of termination/progression of transcription through the BES (Hughes *et al*. 2007). These data may be consistent with the lack of *BVSG* derepression after *TbORC1/CDC6* RNAi, particularly in BSF cells. If some BES control is indeed exerted at the level of the promoter, with transcription gradually attenuated in silent sites as you move further from it, then the level of derepression following *TbORC1/CDC6* RNAi for *BVSG* transcription in PCF might be stronger than reported here; i.e. there may be significant derepression of promoter proximal genes (which was not examined), but little increase in *VSG* mRNA levels (which corresponds with what was observed). The slightly greater evidence for *BVSG* derepression in PCF cells may reflect some life cycle stage differences in the balance between promoter and transcription elongation controls. To test this, transcript levels for *ESAG5, ESAG6* and *ESAG7* would need to be assessed. In contrast to the work of Hughes *et al* (2007), Yang et al (2009) showed that RNAi depletion of TbRAP1 causes derepression of *VSG*s from silent BESs in BSF *T. brucei*, but the extent of this derepression is reduced in genes more proximal to the telomere. TbRAP1, like its homologues in other eukaryotes, associates with telomeres (Yang *et al*. 2009) and may be part of a trypanosome shelterin complex that protects chromosome ends from repair and can mediate heterochromatin-silencing (Dreesen, Li, & Cross 2007). It is known that in mammals ORC can also associate with telomeres, where it is recruited via two other telomere binding components of the shelterin complex, TRF1 and TRF2, and can also interact with HP1, which can bind

methylated histone H3 at telomeres (Deng *et al*. 2009). Given that TbRAP1 RNAi causes telomere-proximal BES derepression in BSF cells, and TbORC1/CDC6 does not, this might be consistent with TbORC1/CDC6 acting at promoters, and not at telomeres. This could be addressed by chIP, which has not been done. However, this is complicated by that fact that BSF TbTRF2 RNAi appears not to result in VSG de-repression (Li, Espinal, & Cross 2005).

Despite the lack of a clear role for TbORC1/CDC6 in determining BES monoallelic expression, we found compelling evidence that *TbORC1/CDC6* RNAi results in derepression of *MVSG*s in PCF cells. Using an EATRO 795 RNAi strain developed by M. Swiderski, we observed a significant derepression of all *MVSGs* tested. RNAi depletion of PLK, Clathrin and an ATPase factor as control experiments clearly demonstrate that the observed derepression is not a stress response caused by RNAi of essential genes, since none of these experiments led to MVSG expression (M. Swiderski, personal communication). To date, no genes involved in *MVSG* silencing in PCF cells have been reported, but M. Swiderski has revealed others in addition to TbORC1/CDC6 (M.Swiderski and J.D.Barry, unpublished). The level of derepression observed following *TbORC1/CDC6* RNAi is comparable to that observed for a nuclear envelope protein, TbNUP-1 (6 – 16 fold increase in MVSG mRNA levels; M. Swiderski, personal communication). Interestingly, *MVSG*s are derepressed to an even greater extent by RNAi of *TbRAP1* and *TbTRF2* (16 – 80 fold) (M. Swiderski, personal communication), which was not reported to affect BES silencing. Why, then, does TbORC1/CDC6 appear to have a role in *MVSG* silencing, but not *BVSG*? This may be consistent with TbORC1/CDC6 exerting its influence at the promoter, since in MES the promoter is much closer to the *VSG* than in the BES and any alleviation of promoter-proximal silencing could encompass the *MVSG*. This is complicated by the lack of homology between MES and BES promoters, however. An alternative explanation is that the monoallelic regulatory factors and mechanisms that operate on the MES and BES are quite different. Indeed, this may be consistent with the observed role for TRF2 in MES silencing in PCF cells, but not in BES silencing in BSF cells (Li *et al*. 2005). Answering this would require the demonstration that *TbORC1/CDC6* or *TRF2* RNAi in BSF cells causes MES derepression, and this is not currently technically possible as MESs in Lister427 *T. brucei* have not been identified and EATRO795 *T. brucei* do not grow in culture as BSF cells. Nevertheless, the

localisation of TbORC1/CDC6 at *MVSG* loci (see chapter 5), coupled with known functions of Orc1 as mediators of heterochromatin (Deng *et al*. 2009; Sasaki & Gilbert 2007), could provide a potential broad explanation of how TbORC1/CDC6 functions at telomeric loci in this parasite, which deserve to be followed up.

A final complication in analysing the above data may provide yet another explanation for the lack of observable effect of *TbORC1/CDC6* RNAi on *BVSG* expression in BSF cells. It could simply be that the *TbORC1/CDC6* RNAi-induced cells die so quickly that any changes in transcript levels are masked by switching off the transcription and translation machinery. The rapidity of *TbORC1/CDC6* BSF RNAi-induced cell death is faster than that of either *TbISWI* (Hughes *et al*. 2007) or *TbRAP1* (Yang *et al*. 2009), which caused BES up-regulation, but may be closer to that following RNAi of *TbTRF2* (Li *et al*. 2005). However, previous observations do suggest there are fundamental differences in the way BSF and PCF cells regulate ES promoter activity, perhaps arguing against this technical explanation. In BSF cells, an rDNA promoter integrated in the BES behaves exactly like the endogenous promoter, whereas in PCF cells integration of an rDNA promoter prevents repression of BES activity (Rudenko *et al*. 1995; Horn & Cross 1995; Rudenko *et al*. 1994). Thus, changes in sequence could mediate differences in TbORC1/CDC6 occupancy at these loci, leading to differential behaviour in terms of regulating ES activation.

The last, and perhaps most critical question, especially given the genome-wide binding of TbORC1/CDC6 (see chapter 5), is how widespread is TbORC1/CDC6 RNAi effect on gene expression? Is it limited only to telomeric *VSG*s? Using our TbORC1/CDC6 427 PCF RNAi strain, RNA-seq was carried out to ask if there are any genome-wide or *VSG*-specific changes in transcript levels between RNAi-induced and non-induced samples. Analysis of RNA-Seq data and quantification of genome-wide RNA levels in two samples (tet – and tet +) is still been performed. The results may also answer the questions on changes in *ESAG* transcript levels upon *TbORC1/CDC6* RNAi induction discussed above.

3.9 Highlight of major findings

To summarise this chapter, the major findings are: (i) *TbORC1/CDC6* is essential for nuclear DNA synthesis in *T.brucei* as revealed by BrdUTP incorporation in PCF cells; (ii) *TbORC1/CDC6* is essential in BSF cells, where its knockdown results in accumulation of multinucleate and multikinetoplast cells with an absence of cytokinesis, a distinct phenotype to the accumulation of zoids in PCF cells; (iii) TbORC1/CDC6 is a nuclear protein; (iv) several failed attempts to generate *TbORC1/CDC6* null mutants, despite being able to generate a heterozygote null and tag the second allele at the endogenous locus, are consistent with it being an essential gene; and (v) TbORC1/CDC6 is involved in the silencing of telomeric *MVSG* loci in PCF cells; whether this role extends more widely to antigenic variation in the BSF remains to be shown.

4 Identification of putative novel components of the *Trypanosoma brucei* Pre-Replication machinery

4.1 Introduction

4.1.1 Is DNA replication initiation in T. brucei atypical of eukaryotes?

Trypanosoma brucei and related species (such as *L .major and T. cruzi*) are single cell eukaryotic parasitic protozoa that display several unique biological features. These organisms have been studied for over a century not only because they cause clinically important diseases in humans and a variety of animals, but also because aspects of their fundamental biology are remarkably distinct from other eukaryotes and therefore shed new light on basic eukaryotic processes. Obviously, if basic biological processes can be shown to be significantly divergent from the eukaryotic hosts that trypanosomatids parasitize they could be exploited for rational drug design to treat the diseases caused by these organisms. A number of divergent biological aspects in trypanosomatids have already been characterised. A few examples include: (a) the sequestration of glycolytic enzymes into a peroxisome-like organelle known as the glycosome; (b) the apparently non-regulated polycistronic transcription initiation of mRNA genes by RNA Pol II; (c) the lack of restriction of RNA Pol I transcription to *rRNA* genes, since the enzyme also drives expression of several mRNA genes, e.g. *MVSG* and *BVSG* genes; (d) *trans*-splicing of nearly all mRNA transcripts; (e) the presence of a hypermodified base, beta-d-glucopyranosyloxymethyluracil (base J), in repetitive DNA sequences; and (f) the presence of an unusual mitochondrial (kinetoplast) DNA, which consists of a catenated network of minicircles and maxicircles with extensive editing of its RNA products (Borst & Sabatini 2008; Lukes, Hashimi, & Zikova 2005; Lukes *et al*. 2002; Clayton 2002; Simpson *et al*. 2000). Broadly, therefore, aspects of energy metabolism, gene expression and regulation and, of particular note, mitochondrial DNA replication seem to be atypical of eukaryotes. So, is trypanosomatid nuclear DNA replication, or at least its initiation, also unusual?

Annotation of putative DNA replication factors using protein homology search algorithms suggests that trypanosomatid parasites, though eukaryotic, possess aspects of their replication initiation machinery that may be archaeal-like in organisation. This is supported by some experimental work, which reinforces the

hypothesis that only a single protein constitutes the pre-RC machinery in *T. brucei* and *T. cruzi* (Godoy *et al.* 2009).This chapter seeks to test this hypothesis further. Here, we focus on the events of DNA replication occurring from the point of recognition of *T. brucei* origins of replication by the ORC complex (or, putatively, simply by TbORC1/CDC6) up to the point at which the origin becomes licensed and is rendered competent for replication firing. The chapter describes a bioinformatic comparison of the pre-replication initiator proteins across a few eukaryotic lineages, and transgenic approaches to examine functional interactions amongst DNA replication proteins in *T. brucei*. The generation of a functional TbORC1/CDC6 epitope – tagged cell line, and lines co-expressing TbMCMs and TbORC1/CDC6 fusion proteins, will be described. Experiments to test for direct interactions between TbMCMs and TbORC1/CDC6, and subsequently using TbORC1/CDC6 as 'bait' to probe for potentially novel interactors, are discussed.

4.2 Results

4.3 A Bioinformatic screen for ORC and Cdc6 components in protozoa relative to higher eukaryotes

Chromosomal DNA replication initiation in "standard" eukaryotes, such as yeasts and metazoans, requires the Origin Recognition Complex (composed of six proteins, Orcs1-6), and Cdc6 to recognise and bind a number of DNA sequences per chromosome, known as origins of replication (Speck *et al.* 2005; Bell & Dutta 2002). This event is a fundamental early requirement for the establishment of replication initiation. The experiments until now have considered the potentially novel circumstances found in kinetoplastid parasites where only a single protein, indistuguishable between Orc1 and Cdc6, can de identified from the above seven proteins. To ask if this is unique to these organisms amongst eukaryotes, we decided to examine the distribution of ORC components, plus CDC6, in a number of protozoans whose genomes have been sequenced. In doing so, we can ask if there is a clear evolutionary picture of how conserved these factors are amongst eukaryotes beyond simply yeasts, *Drosophila*, *Xenopus*, and humans where most work has been done [reviewed in (Gerbi, Strezoska, & Waggener 2002; Bell & Dutta 2002)]. To do this, we performed BLAST searches of all the genomes

detailed in Figure 4-1 with Orc1, Orc2, Orc3, Orc4, Orc5, Orc6 and Cdc6 from *H. sapiens*, *S. cerevisiae* and from *Arabidopsis thaliana*.

From Figure 4-1, it is clear that bioinformatic examination of a wide range of eukaryotic genomes reveals that the absence of clear homologues of each of the six ORC components is not uncommon, and that a single protein related to both Orc1 and Cdc6 is not limited to kinetoplastids. Within the Opisthokonts and Amoebozoa, yeasts, metazoans and the social amoeba *Dictyostileum discoidieum* have six ORC proteins, and Cdc6 as a separate protein, the 'standard' for all higher eukaryotes studied so far (Bell & Dutta 2002). This complement of factors is also found in plants belonging to the Archaeplastida supergroup. However, two members of the Opisthokont supergoup, *Encephalitozoon cuniculi* and *Entamoeba histolytica,* appear not to have this composition: in the former, Orc1 and Cdc6 are separate proteins, but Orcs 3, 4 and 6 are not found; in the latter, Orc2 alone is identified and a single Orc1/Cdc6 protein. The case of *E. cuniculi* is perhaps illuminating. This is an intracellular microsporidian parasite, related to fungi such as yeast, and might therefore be expected to display greater sequence homology, including in the ORC subunits, with Opisthokont eukaryotes. Considerable evidence suggests *E. cuniculi* is undergoing a process of genome reduction (Gill & Fast 2007; Katinka *et al*. 2001), and the absence of ORC components may then be a relatively recent evolutionary loss, illustrating that eukaryotic ORC function can be adapted to a streamlined version lacking some components. In the Excavata supergroup, all genomes examined appeared to encode Orc1 and Cdc6 as a single protein, and none had all six ORC components: *G. Intestinalis* and *Trichomonas vaginalis* encoded recognisable orthologues of the Orc4 subunit, and the Orc2 and Orc4 subunits, respectively. On the other hand, the Chromalveolates, which include members of the Apicomplexan taxon, appear to encode Orc1 and Cdc6 as separate proteins, and a variable number of other Orc subunits were present. Trypanosomatids belong to the taxon *Euglenozoa*, in the supergroup Excavata. From this genome-wide screen trypanosomatids appear to be the only organisms examined that possess Orc1 and Cdc6 as a single protein, with no other subunits of ORC identified. As such, they may have the most simplified eukaryotic machinery involved in the initiation of DNA replication. To summarise, this genome-wide screen could not reveal a clear evolutionary trend for the conservation or absence of ORC

proteins or Cdc6 across the selected examples of eukaryotes for which genome sequences were available.

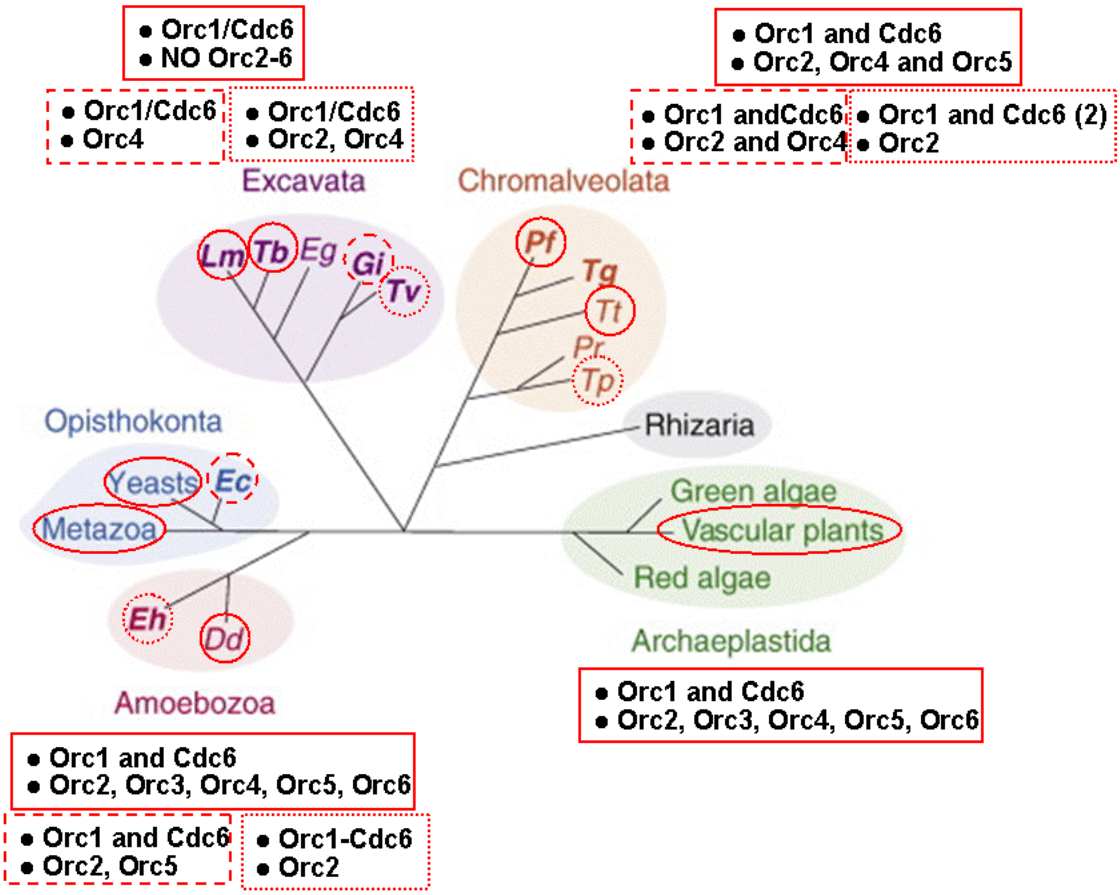


Figure 4-1 – Comparison of ORC complex components and Cdc6 across eukaryotes with full genome sequences published or available

The tree generated according the new level of classification by Adl *et al* (2005), shows the presence of the ORC complex components (Orcs 1-6) and Cdc6 revealed by database mining of genome sequences. Each species and the presence or absence of Orcs1-6 or Cdc6 within each supergroup is indicated by a corresponding broken or solid line; Orc1/Cdc6 indicates a single protein related to both Orc1 and Cdc6, while Orc1 and Cdc6 indicates the presence of both proteins; two copies of a protein is indicated by (2). Lm = *Leishmania major*, Tb = *Trypanosoma brucei*, Gi = *Giardia intestinalis*, Tv = *Trichomonas vaginalis*, Eh = *Entamoeba histolytica*, and Dd = *Dictyostelium discoideum*, Ec = *Encephalitozoon cuniculi*, Pf = *Plasmodium falciparum*, Tt = *Tetrahymena thermophila,* Tp = *Thalassiosira pseudonana*.

4.4 Are typical eukaryotic Pre-RC components conserved in Trypanosomatids?

In budding yeasts (*S. cerevisiae)*, the pre-RC is established at DNA replication origins by the regulated binding of four different proteins: ORC, Cdc6, Cdt1, and the MCM complex (Diffley *et al*. 1994). With no specific pattern observed for

retention or loss of ORC and Cdc6 initiator proteins in the selected eukaryotes described above, a further BLAST was carried out to examine the extent of conservation of the downstream pre-RC factors, Cdt1 and the MCM complex. The molecular events that regulate and follow pre-RC formation are beyond the scope of this thesis and are not explored. Using CLC Genomics Workbench v4 (CLC Bio, Denmark), with a local database of the 9068 predicted ORFs of *T. brucei* from TriTrypDB (http://tritrypdb.org/tritrypdb), pre-RC proteins from other organisms available in the NCBI protein database were used to carry out a local BLAST search for homologues in *T. brucei*. The BLAST search results summarised in Table 4-1 revealed no clear homologue of the eukaryotic helicase recruiter Cdt1 (Randell *et al.* 2006; Tanaka & Diffley 2002). Bearing in mind the previously reported potential analogy of TbORC1/CDC6 to archaea (Godoy *et al.* 2009)(chapter 3), the proteins were then compared to archaeal components. Sequence BLAST using seven archaeal WhiP proteins [*Sulfolobus acidocaldarius* (GenBank accession no. YP_256028); *Metallosphaera sedula* (GenBank accession no. ZP_01600793); *Solfolubus solfataricus* (GenBank accession no. NP_342366); *Aeropyrum pernix* (GenBank accession no. NP_148313); *Hyperthermus butylicus* (GenBank accession no. YP_001013395)] were used to query local database using CLC Genomics workbench. Again, the BLAST results revealed 13 hits all with E-values >1. Therefore, we concluded that there was no obvious homologue of a helicase recruiter protein, since *T. brucei* orthologues of WhiP (Robinson & Bell 2007) were not identified (Table 4-1). Using the same approach but this time using eukaryotic MCM proteins as query (20 versions of each protein from eukaryotes available in the NCBI database – see Appendix 1), all six subunits of the MCM complex were found in *T. brucei*. The hits identified corresponded to annotations present in the *T. brucei* database, TriTrypDB (http://tritrypdb.org/tritrypdb). The gene IDs and corresponding *T. brucei* MCM subunits are shown in Table 4-2.

Table 4-1 – Comparison of Pre-RC components in archaea, "standard" eukaryotes, and *T. brucei*

Function	Archaea	Eukarya	*T. brucei*
Origin recognition	Orc1/Cdc6-fusion protein	ORC 1-6 (heterohexamer)	ORC1/CDC6-fusion protein
Helicase recruiter	WhiP	Cdc6/Cdt1	----------
Helicase	MCM6 (homohexamer)	MCM2-7 (heterohexamer)	MCM2-7 protein Heterohexamer?

The *in silico* database mining experiments above and in Section 4.3 suggest that a number of factors of *T. brucei* DNA replication initiation are not recognisable: Orc2-6 and Cdt1. This means that either (a) *T. brucei* lacks eukaryotic homologues of these proteins, or (b) that these proteins are present but significantly diverged in sequence such that they cannot be identified by conventional bioinformatic approaches. A putative absence that can be tested experimentally is the lack of a helicase recruiter, Cdt1. If this is truly lacking, *T. brucei* replication initiation may be unusual, at least for a eukaryote, in that TbORC1/CDC6 functions alone as the initiator and at the same time as the recruiter of the helicase, which should be detectable as direct interaction between TbORC1/CDC6 and one or more subunits of the MCM heterohexamer. This hypothesis is tested below (Section 4.6).

4.5 TbORC1/CDC6-Myc expressed at the endogenous locus is functional

4.5.1 Cloning of TbORC1/CDC6-Myc construct

For Myc tagging of *TbORC1/CDC6*, oligonucleotides were designed for amplification of a 752-bp region excluding the STOP codon (positions 557 – 1308 relative to the start codon) of *TbORC1/CDC6* using VectorNTI software (Invitrogen®). The primers (CT_OL43/CT_OL44, Table 2-2) contained restriction

sites *Hin*dIII (forward primer) or *Xba*I (reverse primer). The fragments were amplified by polymerase chain reaction (PCR) from *T. brucei* PCF TREU927 strain genomic DNA (gDNA) using Phusion® Taq DNA polymerase (NEB ®). After PCR amplification, the gene fragment was cloned into *Hin*dIII/*Xba*I site of the pNAT^{12MYC} vector shown in Figure 4-2. The vector was linearised within the *TbORC1/CDC6* C-terminal fragment cloned using the restriction enzyme *Xho*I and transfected into the *T. brucei* TREU927 procyclic heterozygote cell line (*TbOrc1/Cdc6* +/-). The strategy for integration of the linearised vector is described below. Selection of positive clones was carried out using blasticidin at a concentration of 15 µg.ml^{-1}.

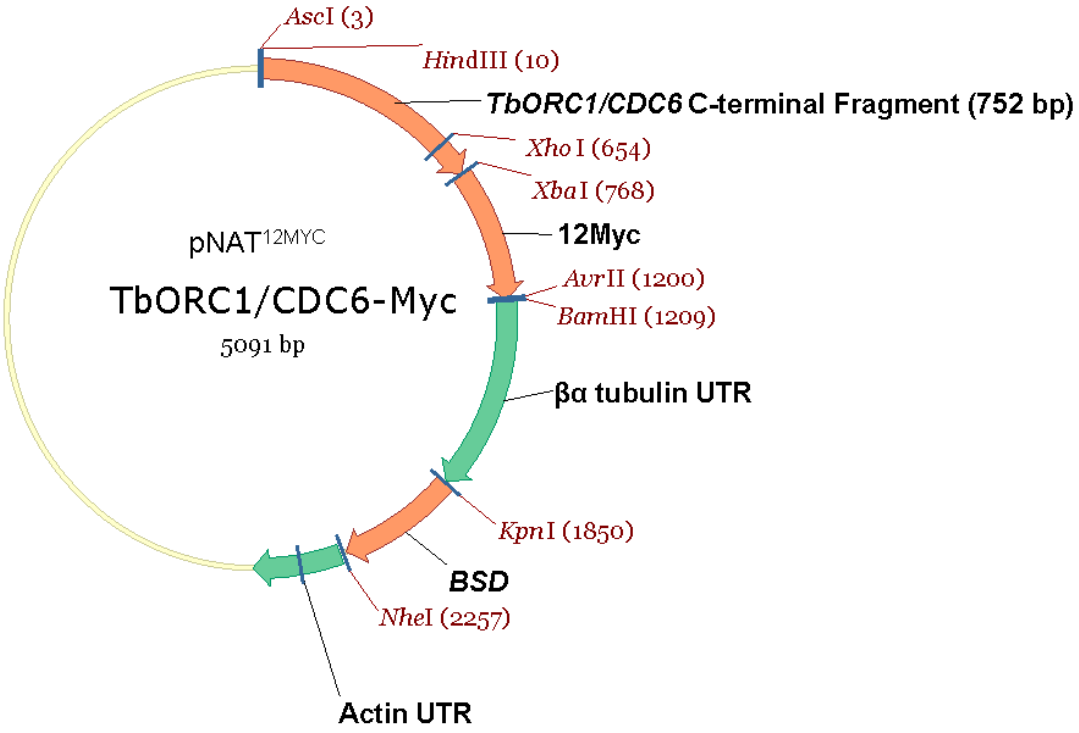


Figure 4-2 – The pNAT^{12MYC} vector for *TbORC1/CDC6-Myc* epitope tagging

A C-terminal coding sequence of *TbORC1/CDC6* (752 bp) that was cloned into *Hin*dIII and *Xba*I sites in the plasmid after PCR amplification is shown; the C-terminal *TbORC1/CDC6* fragment was cloned fused to 12 tandem repeats of Myc (12Myc) to allow the expression of *ORC1/CDC6* inframe with the 12Myc epitope tag. The *Xho*I site used for vector linearization to allow integration into the *TbORC1/CDC6* locus is also shown. Selection of *T. brucei* transformant clones was performed using the blasticidin resistance gene (*BSD*) flanked by Actin and tubulin 3' and 5' mRNA processing regions (Actin UTR) and (βα tubulin UTR). All features shown on the vector map are designed using the VectorNTI resource (Invitrogen). The vector is originally from the lab of David Horn (Alsford & Horn 2008)

4.5.2 Strategy for C-terminal tagging of TbORC1/CDC6

The strategy for C-terminal tagging using the pNAT^{12MYC} employs the same strategy previously described for PTP-tagging (Alsford & Horn 2008; Schimanski, Nguyen, & Gunzl 2005). The *TbORC1/CDC6* ORF is targeted using the C-terminal coding region of the gene (grey bar) which has a unique restriction site (*Xho*I) that is indicated by arrowheads (Figure 4-3 A and B). As illustrated in Figure 4-3 A, B and C, after linearization of the construct (B) with *Xho*I, there is homology-directed integration with the concomitant displacement of a fragment the 3' end of the ORF disrupted (C). The resulting target gene (TbORC1/CDC6; C) becomes modified, fused at the C-terminus inframe with the epitope tag (12Myc), as demonstrated in Figure 4-3.

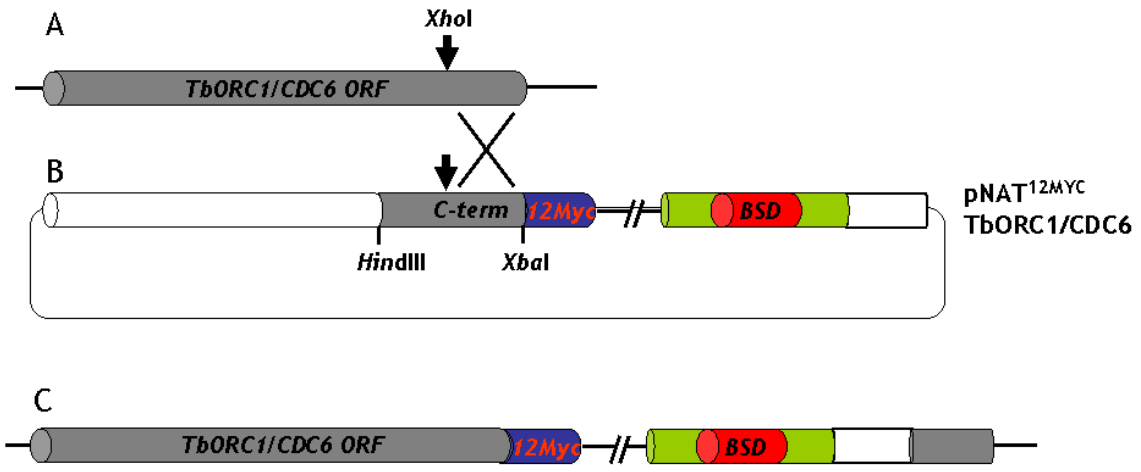


Figure 4-3 – Strategy for C-terminal Myc-tagging of *TbORC1/CDC6*

(A) Wild type open reading frame of *TbORC1/CDC6* (TbORC1/CDC6 ORF) showing the presence of a unique restriction site (*Xho*I) depicted by an arrow head; (B) schematic representation of pNAT^{12MYC} TbORC1/CDC6 C-term vector showing: unique linearization site (*Xho*I) depicted by an arrow head; *Hin*dIII and *Xba*I restriction sites for cloning of the C-terminal fragment of *TbORC1/CDC6* are also shown; the *Blasticidin S Deaminase* gene cassette for selection of positive transformants is shown (BSD); integration occurs by recombination of the c-terminal homologous sequences post linearization (indicated by cross) and (C) illustration of the resulting 12Myc C-terminally tagged *TbORC1/CDC6* locus post integration. Illustration adapted from (Schimanski *et al*. 2005)

4.5.3 Confirmation of TbORC1/CDC6-Myc tagging by PCR

We have previously shown that both alleles of TbORC1/CDC6 cannot be deleted to yield a *TbOrc1/Cdc6* null (-/-) mutant. We also know that one allele can be successfully deleted (Chapter 3). Using the *TbOrc1/Cdc6* heterozygote line in TREU927 strain procyclic form cells (*TbOrc1/Cdc6* +/-, see Chapter 3) we generated transgenic lines that express TbORC1/CDC6 tagged with 12 tandem repeats of Myc at the C-terminus. Six transformant clones resistant to both puromycin (*PUR*; selecting for TbOrc1/Cdc6 +/-) and blasticidin (*BSD*; selecting for TbORC1/CDC6-Myc) were selected and analysed for correct integration of both constructs (Figure 4-4 A and B).

To confirm that the *TbORC1/CDC6::PUR* knockout allele was still intact in these transformants, primers indicated in blue (Figure 4-4 A) were used to carry out a diagnostic PCR. The forward primer was located upstream of the 5' UTR used in the knockout construct, while the reverse primer was located 362 bp from the 5' end of the *PUR* gene ORF. As expected a PCR product size of ~800 bp was generated from each transformant, and no band in the wildtype (WT) TREU927 cells that lacked the *PUR* gene (Figure 4-4 C). This confirmed that the construct had integrated in the TbORC1/CDC6 locus. PCR to test for integration of the *TbORC1/CDC6-myc*-tag construct was also carried out. Using a forward primer that binds to the *TbORC1/CDC6* ORF upstream of the gene fragment cloned into the tagging construct and a reverse primer downstream (after the stop codon) of 12 Myc ORF (CT_OL45/CT_OL46; Table 2-2), five clones out of the six gave a PCR product of the expected size, while the WT cells showed no band (Figure 4-4 B and D). Both PCR reactions validated the correct integration of both constructs at the *TbORC1/CDC6* loci, generating cell lines expected to encode TbORC1/CDC6-Myc at the endogenous locus in a *TbORC1/CDC6*+/- background. These data cannot, however, exclude that a WT *TbORC1/CDC6* allele has been retained, perhaps because the myc-tagged protein is non-functional.

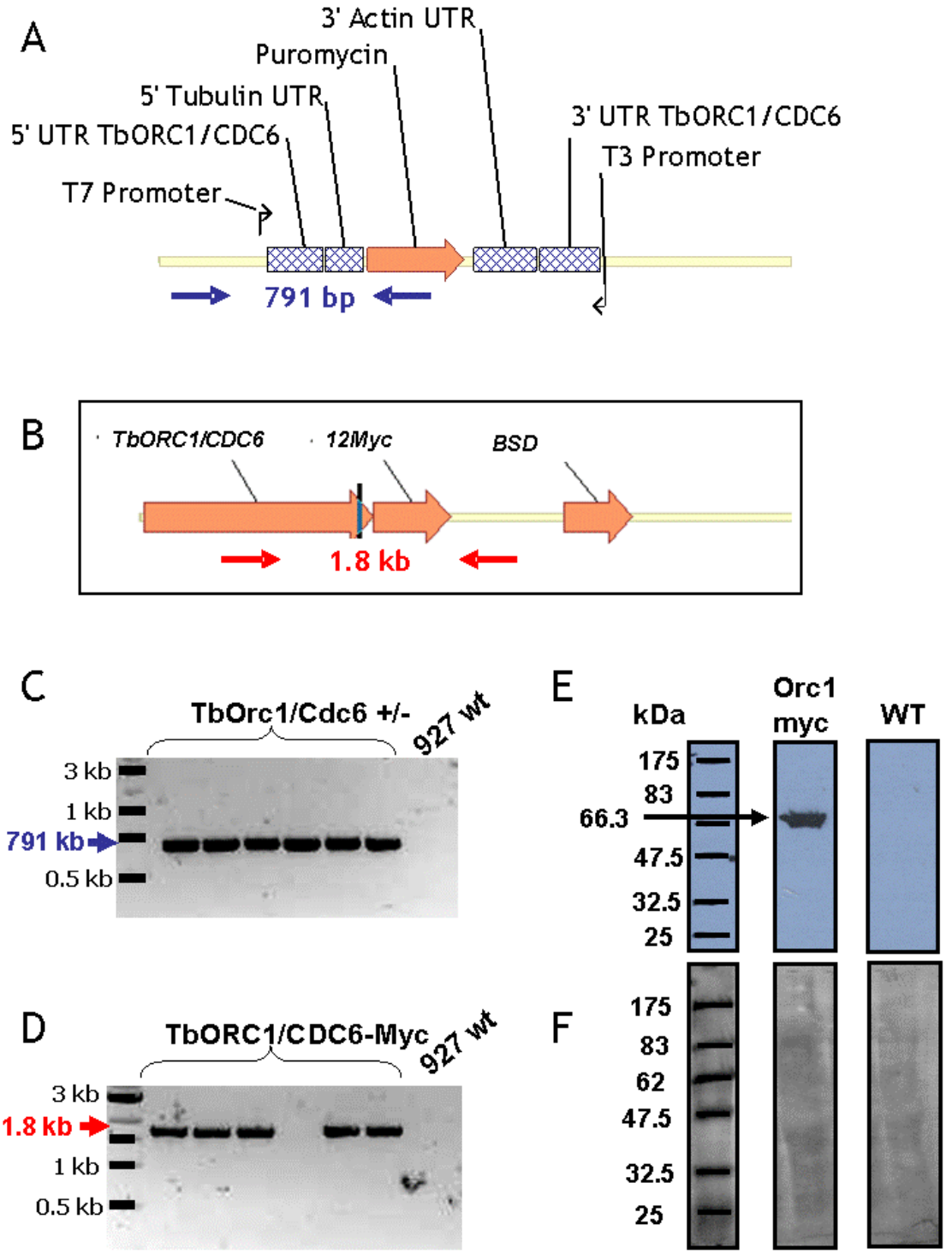


Figure 4-4 –Analysis of TbORC1/CDC6-Myc expressing procyclic form cells

(A) Cartoon representation of *TbORC1/CDC6* loci after integration of a puromycin knock out cassette. Blue arrows indicate position of primers used to check correct integration of construct at the *TbORC1/CDC6* locus. (B) Cartoon representation of *TbORC1/CDC6-12Myc* loci after integration of myc-tagging construct. Red arrows indicate the position of primers for checking correct integration into the *TbORC1/CDC6* locus. (C) Image of a 1% agarose gel to show PCR products from 6 clones checked for integration of construct in (A) with an expected size of 791 bp; a control is shown of PCR from untransformed TREU927 wild type (927 wt) cells. (D) Image of a 1% agarose gel to show PCR products of 6 clones checked for integration of construct in (B) with an expected size of 1.8 kb (E) Western blot of cell extracts from procyclic form cells expressing TbORC1/CDC6-Myc (Orc1-Myc) probed with anti-myc monoclonal antibody; with control lane (WT) showing no cross-reacting band. (F) Ponceau stained membrane used for western blot to demonstrate equal loading of whole cell extracts prior to western blot.

4.5.4 Confirmation of TbORC1/CDC6-Myc tagging by Western blot

Western blot analysis was carried using an anti-Myc monoclonal antibody (Millipore®) to check if TbORC1/CDC6-Myc was expressed in these transformant cell lines. To do this, one clone was examined (Figure 4-4 E). On western blotting, anti-myc antibody recognised a band of ~65 kDa, corresponding to the expected size of a TbORC1/CDC6-Myc fusion (66.3 kDa; 48.6 kDa for TbORC1/CDC6 and 17.7 kDa for 12Myc), absent from WT cells. No further cross-reacting bands were detected in either the TbORC1/CDC6-Myc or WT extracts (Figure 4-4 E). Equal loading of extracts from the both TbORC1/CDC6-Myc and WT cells was checked by Ponceau staining of the membrane (Figure 4-4 F).

4.5.5 Confirmation of TbORC1/CDC6-Myc tagging by Southern blot

Having shown that TbORC1/CDC6-Myc is expressed in this cell line, and that the anti-Myc antibody does not cross-react with other antigens in the parasite whole cell extract, Southern blot analysis was carried out to test whether or not only the *TbORC1/CDC6-Myc* and *TbORC1/CDC6::PUR* alleles were present, as shown in the schematic illustration in Figure 4-5 A. One allele of *TbORC1/CDC6* was replaced by homologous recombination with the *puromycin* resistance cassette, while the second allele was tagged at the C-terminus with 12 repeats of Myc. Using *Eco*RV to digest genomic DNA prepared from the transformant and WT cells, Southern blotting was used with a probe hybridising to the 5' end of the ORF of *TbORC1/CDC6* (Figure 4-5 A, B). This showed that both alleles of this loci are identical in size, and that the gene is single copy, as indicated by a single band at 4.1 kb in the WT lane (Figure 4-5 B). The same size band was seen in the *TbORC1/Cdc6* +/- cells, indicating that at least one allele is intact. In the puromycin and blasticidin-resistant transformant, with the *TbORC/CDC6-Myc* allele, a single band of 3.1 Kb was seen, due to modification of the locus caused by the introduction of an *Eco*RV site during integration of the tagging construct. The absence of the 4.1 kb band seen in the WT and +/- lanes indicates that the *TbORC*/CDC6 ORF is only present as a *TbORC/CDC6-Myc* allele.

To confirm that the antibiotic resistance cassettes were integrated as expected in the TbORC1/CDC6-Myc line, we used the same Southern blot and a probe recognising the 5' UTR of *TbORC1/CDC6*. Since there is an *Eco*RV site within the *PUR* resistance cassette, using the 5' UTR probe we can distinguish the WT *TbORC1/CDC6* and *TbORC/CDC6::PUR* alleles, and these in turn can be distinguished from the *TbORC/CDC6-Myc* allele. Digestion of genomic DNA from all three lines with *Eco*RV and probing with a 5' UTR probe showed a 1.9 kb band corresponding to the *TbORC/CDC6::PUR* allele only in the *TbOrc1/Cdc6* +/- and TbORC1/CDC6-Myc cells, and no band in WT (Figure 4-5 C). In addition, a 3.1 kb band was seen in the TbORC1/CDC6-Myc cells, where there was no evidence of a 4.1 kb band corresponding to WT *TbORC1/CDC6* seen in the WT and +/- cells.

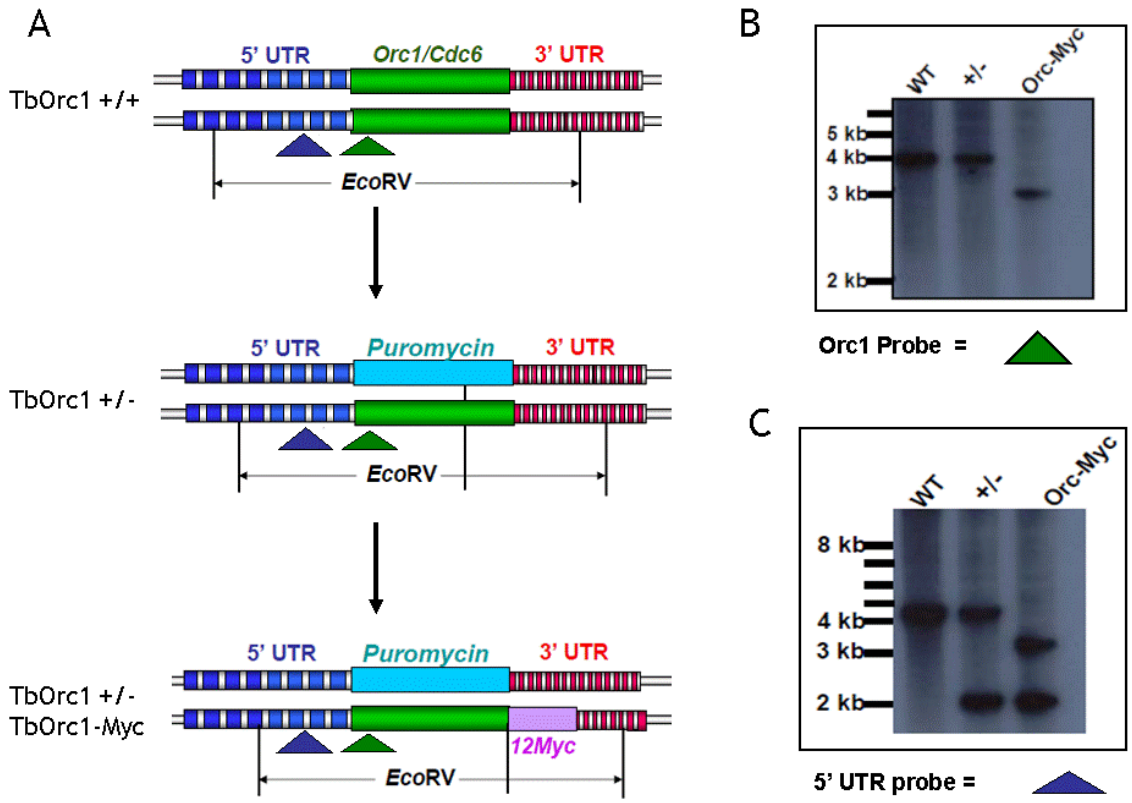


Figure 4-5 – Southern blot confirmation of TbORC1/CDC6-myc expressing lines

(A) Cartoon representation of the gene disruption strategy from a wild type line (TbOrc1/Cdc6 +/+) where both alleles are intact: first, to heterozygote null line where one allele has been replaced with a transfected puromycin cassette (TbOrc1/Cdc6 +/-) is generated; and then, after a second round of transfection, a TbOrc1/cdc6 +/- and TbORC1/CDC6-Myc line is generated where one allele has been replaced with the *PUR* cassette and the other allele is tagged to 12 tandem repeats of *Myc*. The Blue triangle indicates the position of a probe to the 5' UTR of *TbORC1/CDC6* and the green triangle indicates the position of a probe to the 5' end of the ORF of *TbORC1/CDC6*. *Eco*RV was used to digest genomic DNA prepared from each of the three lines WT = TbOrc1/Cdc6 +/+, +/- = TbOrc1/Cdc6 +/-, and Orc-Myc = TbOrc1/Cdc6 +/- and TbORC1/CDC6-Myc. (B) Autoradiogram using a probe to the 5' end of *TbORC1/CDC6* ORF (green triangle) to confirm loss of the wild type allele, and retention of a modified allele (C) Autoradiogram using a probe to the 5' UTR of *TbORC1/CDC6* (blue triangle) to confirm loss of the wild type *TbORC1/CDC6* allele with a concomitant retention of the heterozygote allele at the endogenous locus.

Southern blot analysis was also carried out with a *Hin*dIII digest of genomic DNA from these cell lines and the results revealed a similar pattern after probing with the same probes (data not shown). Given that both alleles of *TbORC1/CDC6* could not be deleted (Chapter 3), the loss of WT *TbORC1/CDC6* when the gene was Myc-tagged at the C-terminus in a heterozygote mutant clearly demonstrates that the locus is accessible and amenable to genetic manipulation and that TbORC1/CDC6-Myc is functional.

4.6 Does TbORC1/CDC6 interact with TbMCM?

In eukaryotes both Cdc6 and Cdt1 function in conjunction with the ORC complex to provide a platform at the origin of replication to recruit the Mcm2-7 helicase complex for local DNA unwinding (Bell & Dutta 2002). As explained above, *T. brucei* bioinformatically lacks Cdc6 as a distinct replication initiation component, lacks a clear eukaryotic Cdt1 homologue, and lacks an obvious archaeal WhiP homologue, which may provide Cdt1 functions (Robinson & Bell 2007). On the other hand all six MCM subunits, named Mcm2-7 (named here as TbMCM2-7) are unambiguously present in the genome of *T. brucei*. To ask if TbORC1/CDC6 directly interacts with the MCM helicase, we generated cell lines that co-express TbORC1/CDC6-Myc with C-terminally tagged variants of each of the MCM subunits.

4.6.1 Cloning of TbMCM-HA constructs

To HA-tag the endogenous genes of each TbMCM subunit, six constructs were generated, each as described below in Figure 4-6, possessing a C-terminal fragment of a *TbMCM* ORF translationally fused with 6 tandem repeats encoding *HA* (Figure 4-6 A – F). The same principle used for construction of the pNAT^{12MYC}-TbORC1/CDC6-Myc construct (Section 4.5.2) was used for the TbMCM-HA constructs. A derivative of the pNAT^{12MYC}-TbORC1/CDC6-Myc vector in which the *BSD* gene had been replaced with a bleomycin gene (*BLE)* was obtained from M. Swiderski (WTCMP, University of Glasgow). Using the latter, the *12Myc* fragment in this vector was replaced with a *6HA* fragment using primers CT_OL49/ CT_OL50 (Table 2-2). A 201 bp *6HA* fragment was amplified by PCR from plasmid PGL1728 (Gift from T. Hammarton, University of Glasgow) using Phusion® Taq

DNA polymerase (NEB ®). After PCR-amplification, the PCR product was cloned into *Xba*I/*Bam*HI site of the *BLE* resistant pNAT^{12MYC}-TbORC1/CDC6-Myc vector to obtain the vector pNAT^{6HA}-TbORC1/CDC6-HA. The latter was used as a template to clone all six *TbMCMs* C-terminal regions using the same strategy described in Section 4.5.2. The primers used (for sequences see Table 2-2), C-terminal fragment sizes used for each *TbMCM*, the restriction sites, and the sites for linearization sites are shown in Table 4-2.

Table 4-2 – Construction of TbMCM-HA tagging vectors

Gene ID	Primers	Restriction sites	Fragment size	Linearization site
Tb11.02.5730 / *TbMCM2*	CT_OL55 / CT_OL56	*Hin*dIII/*Xba*1	774 bp	*Cla*I
Tb927.2.3930 / *TbMCM3*	CT_OL61 / CT_OL62	*Hin*dIII/*Xba*1	540 bp	*Hpa*I
Tb11.02.4070 / *TbMCM4*	CT_OL59 / CT_OL60	*Hin*dIII/*Xba*1	648 bp	*Xho*I
Tb11.02.3270 / *TbMCM5*	CT_OL57 / CT_OL58	*Sac*I/Xba1	600 bp	*Pvu*II
Tb11.01.3510 / *TbMCM6*	CT_OL51 / CT_OL52	*Hin*dIII/*Xba*1	604 bp	*Pvu*II
Tb11.01.7810 / *TbMCM7*	CT_OL53 / CT_OL54	*Hin*dIII/*Xba*1	683 bp	*Age*I

Each plasmid has the *Bleomycin* resistance gene (*BLE*) that confers resistance to phleomycin and can be used to select for *TbMCM-HA* transformants after transfection into the *TbORC1/CDC6-Myc* line, described above (which is blasticidin- and puromycin-resistant). Each *TbMCM* construct was transfected and phleomycin-resistant clones selected with 10 μg.ml^{-1} zeocin.

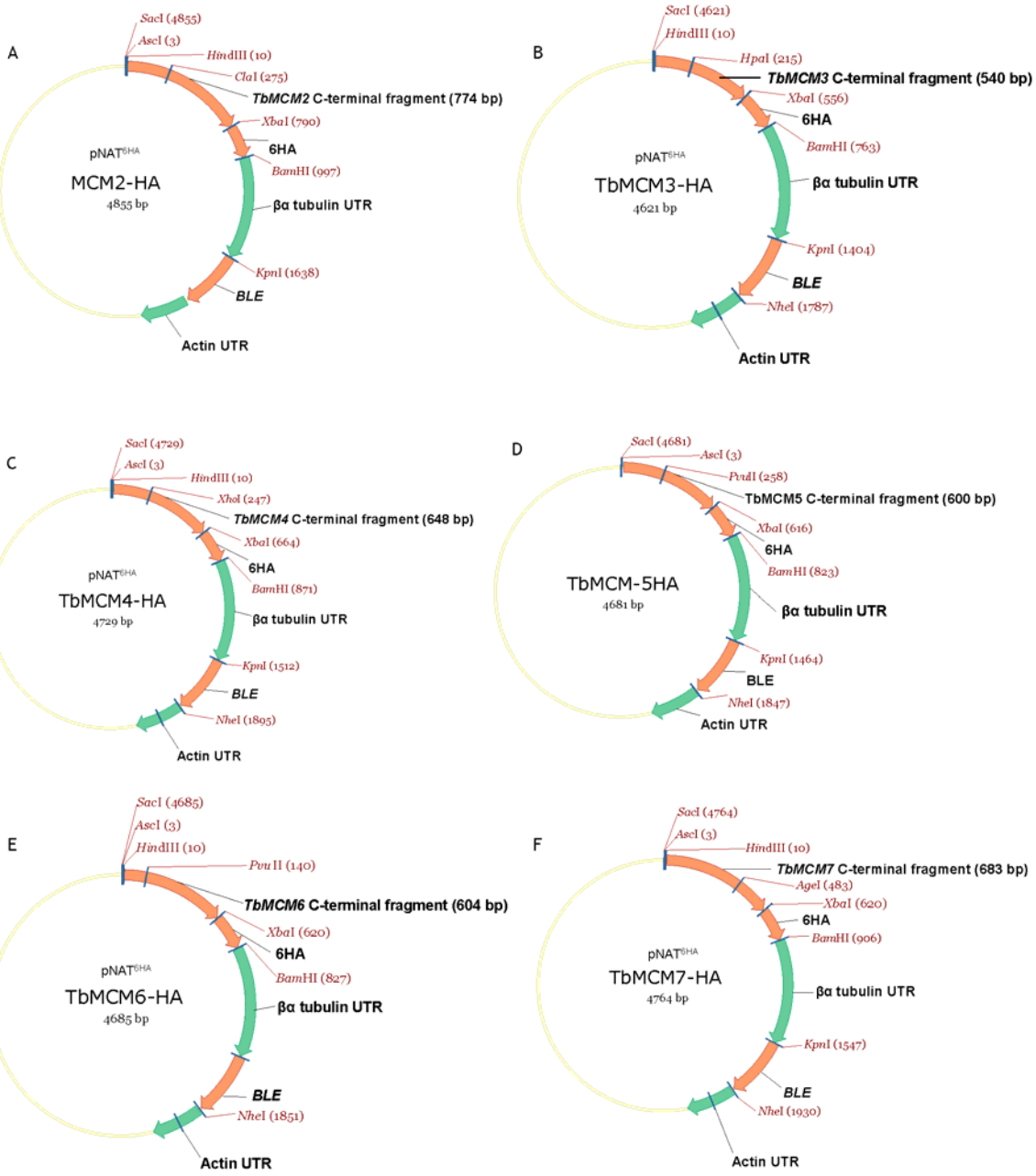


Figure 4-6 – TbMCM-HA constructs for C-terminal tagging

Plasmids for tagging *TbMCM*s at the C-terminus of their endogenous loci: (A) = *TbMCM2-HA,* (B) = *TbMCM3-HA*, (C) = *TbMCM4-HA*, (D) = *TbMCM5-HA*, (E) = *TbMCM6-HA*, and (F) = *TbMCM7-HA.* The C-terminal coding sequence size of each *TbMCM* that was cloned into the plasmid after PCR amplification is shown; the C-terminal fragments of each *TbMCM* was cloned fused to 6 tandem repeats of HA (6HA) to allow the expression of *TbMCMs* inframe with the *6HA*. The sites used for vector linearization to allow integration into the *TbMCM* loci are also shown within the C-terminal fragment of each gene. Selection of *T. brucei* transformant clones was performed using the phleomycin resistance gene (*BLE*) flanked by Actin and tubulin 3' and 5' mRNA processing regions (Actin UTR) and (βα tubulin UTR). All features shown on the vector map are designed using the VectorNTI resource (Invitrogen). The vector is originally from the lab of David Horn (Alsford & Horn 2008)

4.6.2 Confirmation of TbORC1/CDC6-Myc and TbMCM-HA co-expression cell line

After transfection, clones were obtained for all TbMCMs (except for TbMCM5), co-expressing TbORC1/CDC6-Myc, represented by cartoon in Figure 4-7 A and B. Western blots were carried out with anti-HA monoclonal antibody to confirm expression of individually HA-tagged TbMCM subunits in two clones for TbMCM2. three clones for TbMCM3, one clone for TbMCM4, one clone for TbMCM6, two clones for TbMCM7 and, using the same cell extracts, co-expression of TbORC1/CDC6-Myc was also confirmed by western blot and hybridisation with anti-Myc antibody (Figure 4-7 C and D). From the western blots it is evident that the individual HA-tagged TbMCM subunits were expressed at different levels in *T. brucei* procyclic form cells, if TbORC1/CDC6-Myc is considered as a loading control. Visual inspection indicated that TbMCM2 was least expressed, followed by TbMCM7, and TbMCM3, 4 and 6 were probably expressed at the same, higher levels (Figure 4-7 C and D).

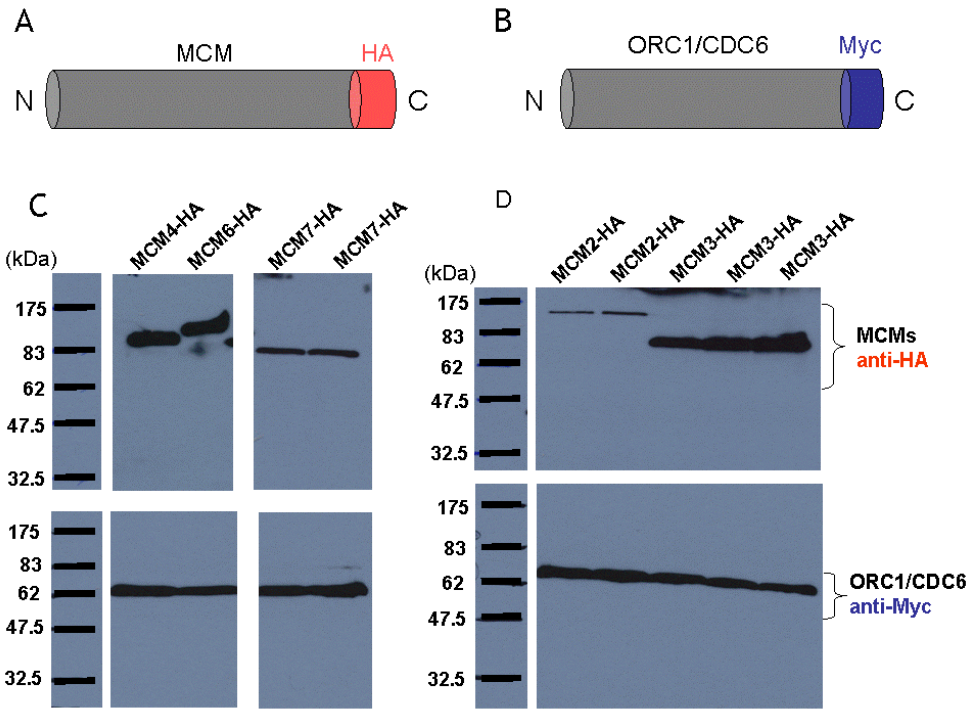


Figure 4-7 – Western blot confirms procyclic TREU 927 cell lines co-expressing TbMCMs-HA and TbORC1/CDC6-Myc

(A) and (B), respectively, show cartoon representations of an MCM subunit tagged at the C-terminus with HA, and TbORC1/CDC6 tagged at the C-terminus with Myc. (C) and (D) show TbMCM-HA expression in whole cell extracts of transformant clones, detected on a western blot using anti-HA antibody (top panel), and TbORC1/CDC6-Myc from the same whole cell extracts detected using anti-Myc antibody (bottom panel). Single clones are shown for TbMCM4-HA and TbMCM6-HA, two clones for TbMCM2-HA and TbMCM7-HA, and three clones for TbMCM3-HA.

4.6.3 Co-immunoprecipitation of TbMCM-HA and TbORC1/CDC6-Myc using anti-HA antibody

To investigate if any of the TbMCM subunits directly or indirectly interact with TbORC1/CDC6 *in vivo*, the doubly tagged procyclic form cell lines described above were used to carried out co-immunoprecipitation (co-IP) experiments with anti-HA to precipitate TbMCM3-HA, TbMCM6-HA, or TbMCM7-HA. Anti-HA antibody was coupled to magnetic anti-mouse conjugated Dynalbeads (Invitrogen; see Materials and Methods for details) and whole cell extracts (input) from 10^8 cells were incubated with anti-HA coupled beads for two hours. After several washes and then elution in high salt, the precipitates from each cell line (eluates) were immunoblotted using either anti-HA (to detect the HA tagged TbMCM subunit) or anti-Myc antibody (to detect TbORC1/CDC6-Myc). The anti-HA blotting determined if IPs were successful, while the anti-myc antibody asked if there were any detectable interactions between the individual TbMCM-HA subunits and TbORC1/CDC6-Myc. Figure 4-8 A shows inputs and eluates from double expressor lines and a cell line expressing TbORC1/CDC6-Myc only (i.e. not transformed with any TbMCM tag construct). The TbORC1/CDC6-Myc cell line served as a control for non-specific binding of TbORC1/CDC6-Myc protein to the anti-HA coupled beads used for IPs. The anti-HA probed western blot confirms that it was possible to IP TbMCM3-HA (96.5 kDa), TbMCM6-HA (104 kDa), and TbMCM7-HA (88.8 kDa) (Figure 4-8 A). However, immunoblotting for TbORC1/CDC6-Myc with anti-myc antibody did not detect co-IP of any of the TbMCM subunits and TbORC1/CDC-6Myc (Figure 4-8 B), suggesting they do not interact, at least in these conditions. Clearly, we cannot rule out interaction between TbORC1/CDC6-Myc and either or both of TbMCM2-HA, TbMCM4-HA and TbMCM5-HA, as these were not analysed. However, if the proteins form part of stable helicase complex, direct interaction between TbORC1/CDC6 and these components of the helicase would allow co-IP of the ORC factor and the three tested MCM subunits. To test this, interactions between the TbMCM subunits is examined below.

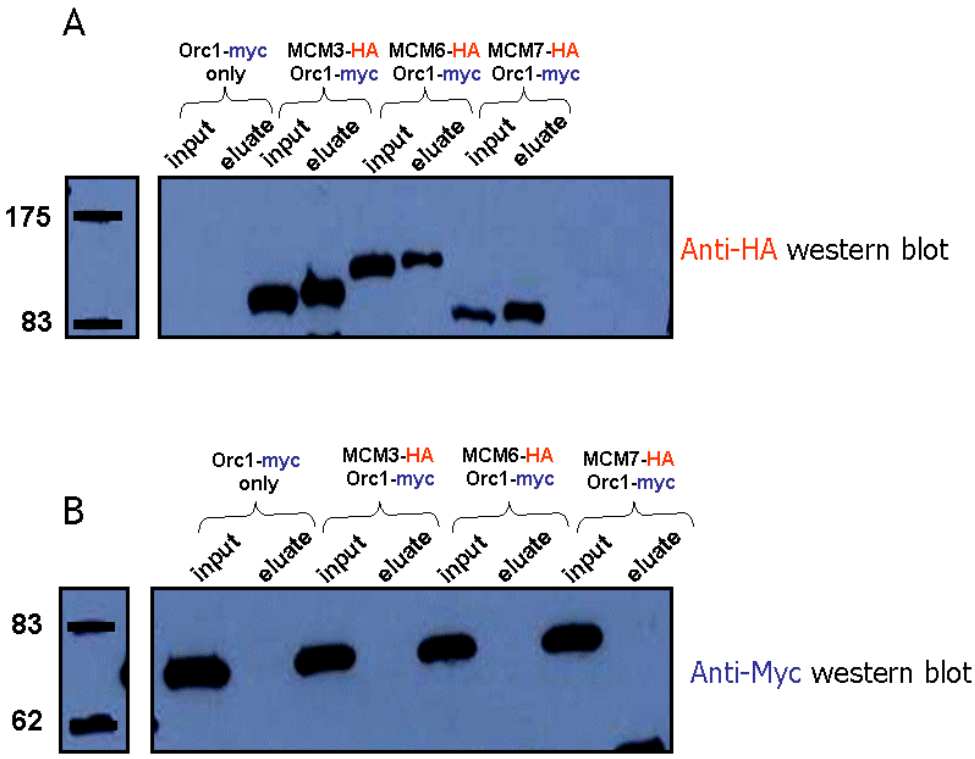


Figure 4-8 – Western blot analysis of the inputs and eluates following co-immunoprecipitation using anti-HA antibody

(A) The input and eluate from co-IPs from whole cell extracts of a control cell line (Orc1-Myc, expressing TbORC1/CDC6-Myc), and from co-expressor cell lines (expressing TbMCM3-HA, TbMCM6-HA, or TbMCM7-HA and TbORC1/CDC6-Myc), were separated on a 10 % SDS-PAGE gel and then transferred to a nylon membrane. This was then probed with anti-HA antibody (A) or with anti-Myc antibody (B).

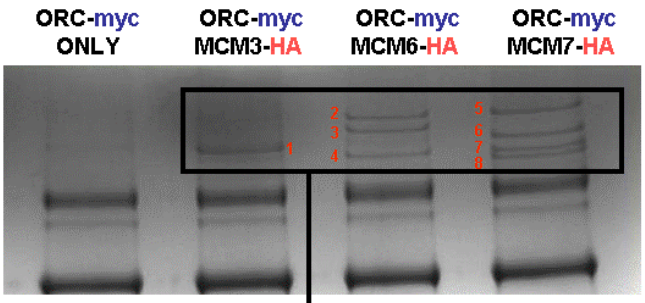
4.6.4 Identification of TbMCM subcomplexes from co-IP of TbMCM-HA and Mass spectrometry

Since IP of HA-tagged TbMCM3, TbMCM6, and TbMCM7 were successful, yet no interaction with Myc-tagged TbORC1/CDC6 was identified by anti-Myc blotting, we wondered if the IP eluates, when separated on an SDS-PAGE gel, could reveal discreet bands in the co-expressor cell lines relative to each other and to the TbORC1/CDC6-Myc alone control. Such bands may indicate interacting factors that are not HA- or Myc-tagged, and would not therefore have been seen in the blots. The eluate from the IP of each TbMCM subunit examined in Figure 4-8, and from the TbORC1/CDC6-Myc control, was therefore separated on a 10 % SDS-PAGE gel and stained with colloidal coommassie. Imaging of the gel revealed distinct patterns of bands for TbMCM7-HA (4 bands), TbMCM6-HA (3 bands), and

TbMCM3-HA (1 band) that were absent from the control (Figure 4-9 A). Each band was excised and analysed by protein fingerprinting by Liquid Chromatography-Electrospray Tandem Mass Spectrometry (LC-ES MS/MS) at the University of Glasgow Proteomics facility. The resulting MS/MS spectra were used to interrogate the TritryDB database using the MASCOT software (http://www.matrixscience.com/). All protein hits identified had at least 11 unique peptides that confidently matched a single ORF within the database. Only protein hits with an overall MASCOT score of greater than 30 ($p < 0.05$) were considered significant. A summary of the results from the mass spectrometry is shown in Figure 4-9 B.

These data show the following: IP of TbMCM6-HA also immunoprecipitated TbMCM2 and TbMCM4; IP of TbMCM7-HA also immunoprecipitated TbMCM2, TbMCM4, and TbMCM6; and the single band excised from the IP of TbMCM3-HA was identified as TbMCM3 itself. Thus, we find that a subcomplex can be detected containing TbMCM2, TbMCM4, TbMCM6 and TbMCM7, which is likely to represent biologically significant *in vivo* protein-protein interactions as the co-IPs were performed from procyclic form whole cell extracts and no cross-linking agents were used prior to the experiment. We did not detect interaction between the above subcomplex and TbMCM3 or TbMCM5, nor did IP of TbMCM3-HA reveal interaction with TbMCM5.

A



B

Bands excised and sent for Mass spectrometry

	Protein Identification by 1D nLC-MS-MS			
Label on gel	Hit - Gene ID	Annotation in genome	No. of peptides	
1	Tb927.2.3930	TbMcm3	22	MCM3-HA IP
2	Tb11.02.5730	TbMcm2	29	MCM6-HA IP
3	Tb11.01.3510	TbMcm6	25	
4	Tb11.02.4070	TbMcm4	31	
5	Tb11.02.5730	TbMcm2	31	MCM7-HA IP
6	Tb11.01.3510	TbMcm6	20	
7	Tb11.02.4070	TbMcm4	11	
8	Tb11.01.7810	TbMcm7	26	

Figure 4-9 – Mass spectrometric characterisation of TbMCM-HA precipitates reveals TbMCM subcomplexes

(A) Eluates are shown from immunoprecipitations (IP) using anti-HA antibody from *T. brucei* cell lines expressing TbMCM3-HA, TbMCM-6HA, TbMCM7-HA or TbORC1/CDC6-Myc (Orc1-Myc ONLY) as a control; proteins were separated by SDS-PAGE and visualised by colloidal commassie staining. Bands that were excised and analysed by mass spectrometry are numbered; the results of the analysis of these bands is shown in (B). The number of peptides indicated for each band (No. of peptides) are unique peptides.

4.6.5 Co-immunoprecipitation of TbORC1/CDC6-Myc and TbMCM-HA using anti-Myc antibody

In other eukaryotes, it is known that once the ORC complex binds origin DNA, two further factors, Cdc6 and Cdt1, are required to recruit the MCM complex (Mcm2-7) for local unwinding of DNA at the origin (Bell & Dutta 2002). Since there is no clear Cdt1 homologue in *T. brucei* and IP assays of TbMCM subunits did not reveal any interaction with TbORC1/CDC6, we considered it possible that TbORC1/CDC6 must first bind to chromatin prior to recruitment of TbMCM. Therefore a substantial portion of the cellular pool of TbMCM may be unbound to DNA/TbORC1/CDC6, masking any interactions being detected by IP of an MCM subunit. We therefore sought to carry out the reciprocal IP from that performed

in Section 4.6.3, using anti-Myc antibody; in this case, TbORC1/CDC6-Myc will be immunoprecipitated and co-IP of TbMCM-HA tested using anti-HA antibody. The same experimental procedures described above and the same co-expression cell lines were used. In this case, single-tagged cell lines expressing only TbMCM6-HA or TbMCM7-HA were used as co-IP controls. IP of TbORC1CDC6-Myc was successful when co-expressed with either TbMCM6-HA or TbMCM7-HA (Figure 4-10 A), and the control cell lines showed no bands, as expected. This demonstrated that IP of TbORC1/CDC6-Myc is successful with the anti-Myc antibody used. However, immunoblotting for TbMCM6-HA or TbMCM7-HA using anti-HA antibody did not detect the proteins in the eluate from any of the IPs (Figure 4-10 B), suggesting that interaction with TbORC1/CDC6-Myc is undetectable in these conditions. This result complements the previous observations made in Section 4.6.3. As before, we cannot exclude from this analysis that TbORC1/CDC6-Myc IP would reveal interaction with TbMCM3, TbMCM4 or TbMCM5, but deem it unlikely.

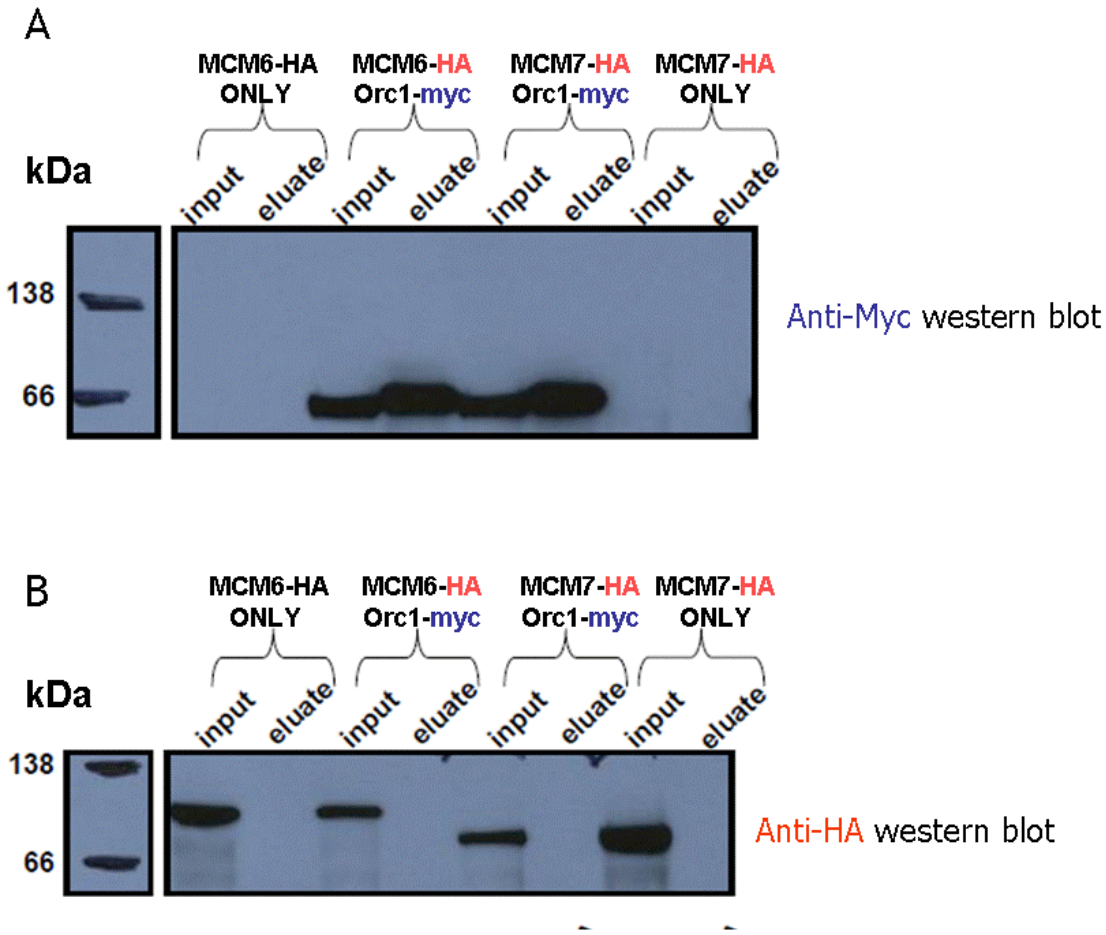


Figure 4-10 – Western blot analysis of the inputs and eluates following co-immunoprecipitation using anti-Myc antibody

(A) The input and eluate from co-IPs from whole cell extracts of control cell lines (MCM6-HA and MCM7-HA, expressing TbMCM6-HA and TbMCM7 respectively), and from co-expressor cell lines (expressing TbMCM6-HA, or TbMCM7-HA and TbORC1/CDC6-Myc), were separated on a 10 % SDS-PAGE gel and then transferred to a nylon membrane. This was then probed with anti-Myc antibody (A) or with anti-HA antibody (B).

4.7 Identification of putative novel pre-RC components in *T. brucei*

So far, we have shown that amongst the pre-RC replication machinery of *T. brucei,* only TbORC1/CDC6 can be identified bioinformatically, and may represent the functional analogue of the six protein ORC complex in higher eukaryotes. Also, all subunits of the MCM complex in higher eukaryotes are represented in *T. brucei* and some have been shown to interact, while interactions between TbORC1/CDC6 and three members of the TbMCM complex have eluded detection. We therefore wondered if *T. brucei* contains novel proteins, or highly divergent ORC subunits, that form part of the pre-RC machinery. To attempt to identify any such factors, we adopted a proteomic approach. Whole cell extracts were prepared from PCF *TbORC1/CDC6-Myc* cells and from untagged (TREU927 WT) cells and subjected to IP using Dynalbeads coupled with anti-myc antibody. The WT cell line served as a control for proteins that non-specifically bind to the magnetic beads. A further control cell line, expressing a distinct myc-tagged protein, to allow identification of *T. brucei* proteins that interact with the myc epitope and not TbORC1/CDC6, was not available and could not therefore be used. After several washes (see material and methods), eluates were run on a 10 % SDS PAGE gel, proteins throughout the lanes excised in 5 distinct bands for each cell line, and analysed by LC-ES MS/MS. The resulting MS/MS spectra were used to interrogate the TritryDB database using the MASCOT software (http://www.matrixscience.com/). In total, 405 proteins were identified from the TbORC1/CDC6-Myc IP and 285 hits from the WT IP. These data were first filtered by excluding all proteins (using the Microsoft Excel VLOOKUP function) which were common in both IP samples. From this, proteins unique to the TbORC1/CDC6 IP tallied ~150 protein IDs, which were absent from the WT control IP. To further filter these proteins, as a rudimentary gauge of abundance, we used the number of unique peptide hits recovered for each protein to rank them. The bait, TbORC1/CDC6 itself, was

identified by 10 unique peptides hits in the TbORC1/CDC6-Myc IP, none of which were found in the IP from the WT control. This “proof of principle” demonstrated that the IP was successful, and that at least some of the filtered proteins are likely to be significant. Table 4-3 shows protein IDs for proteins with 3 unique peptide hits or greater that were recovered from this filtering and are putative TbORC1/CDC6-interacting factors. They are described via their annotation in TriTryDB, where each is considered a hypothetical protein.

Table 4-3 – Potential TbORC1/CDC6-interacting proteins, showing the number of peptide hits for each from mass spectrometry analysis and a summary of bioinformatic analysis of potential function; TbORC1/CDC6 is not included in the list, but was identified by 10 peptides

>Tb927.6.1120 hypothetical protein,

Peptide hits: 14 peptides
CLC PFAM search: Zinc finger (zf - C2H2 type) - PF00096, Tetratricopeptide repeat - PF07719
NCBI psi BLAST: No clear hits; very weak homology to NTPase (DNA PolIII protein) on psi BLAST
PHYRE search: 70% precision hit to virion RNA polymerase

>Tb09.160.3120 hypothetical protein,

Peptide hits: 10 peptides
CLC PFAM search: No clear hits
NCBI psi BLAST: No clear hits; very weak homology to NTPase (mismatch repair and DNA Pol III proteins)
PHYRE search: low precision 5% to archaeal Orc1/Cdc6 protein (2/3 top hits)

>Tb10.389.0050 tbrucei_v5:Tb10.389.0050/Tb927.10.13380 hypothetical protein,

Peptide hits: 10 peptides
CLC PFAM search: AMP binding enzyme - PF00501, Tetratricopeptide repeat - PF00515
NCBI BLAST: very low level homology to *D. melonogaster* Orc4 (E-value = 0.055), and to Ploop NTPase (E-value = 0.002)
PHYRE search: top hits to archaeal Orc1/Cdc6

>Tb11.46.0009 tbrucei_v5:Tb11.46.0009 hypothetical protein,

Peptide hits: 5 peptides
CLC PFAM search: Helicase conserved C-terminal domain - PF00271
NCBI BLAST: DEAD box helicase; Zn finger
PHYRE: Very high precision hits to ATP-dependent RNA helicases

>Tb927.2.5810/1F7.275 tbrucei_v5:Tb927.2.5810/1F7.275 hypothetical protein,

Peptide hits: 4 peptides
CLC PFAM search: Zinc Knuckle (zf-CCHC) - PF00098
NCBI BLAST: Transcription accessory factor, Tex
PHYRE search (N-term 1200aa): 100% precision hit to Pseudomonas Tex RNA binding factor; ARM (armadillo/HEAT) repeats

>Tb927.10.7980 hypothetical protein, (Tb10.6k15.2570)

Peptide hits: 4 peptides
CLC PFAM search: ATPase associated family - PF00004, ABC transporter - PF00005, ATP synthase alpha/beta family - PF00006
NCBI psi BLAST: replication factor C, and polymerases on 1st and 2nd iterations
PHYRE: top hits to archaeal ORC1/Cdc6

>Tb11.02.5725 hypothetical protein,

Peptide hits: 3 peptide hits
CLC PFAM search: No hits
psiBLAST: No strong hits
PHYRE: No convincing hits

>Tb927.2.5130 hypothetical protein,

Peptide hits: 3 peptides
CLC PFAM search: Glycosyl transferase group 1 - PF00534, Short chain dehydrogenase - PF00106
NCBI BLAST: Very low homology to plasmid replication proteins (top hits)
PHYRE: High precision (90-95%) hits to ARM (armadillo/HEAT) repeats, and archaeal topoisomerase V

>Tb10.389.1520 (Tb927.10.12220) hypothetical protein,

Peptide hits: 3 peptides
CLC PFAM search: No hits
NCBI BLAST: No strong hits
PHYRE: no strong hits (5% precision to alpha actinin)

>Tb10.6k15.1050 (Tb927.10.9350) hypothetical protein,

Peptide hits: 3 peptides
CLC PFAM search: ATPase associated family - PF00004, Helix-turn -helix -PF00512, GGDEF domain - PF0090
NCBI BLAST: No strong hits
PHYRE: High precision (60-70%) hits, mainly to ARM (armadillo/HEAT) repeats, including in virion RNA polymerase

>Tb927.7.1060 hypothetical protein,

Peptide hits: 3 peptides
CLC PFAM search: alpha/beta hydrolase fold -PF00561, ATP synthase family -PF00006, CBS domain pair -PF00571
psiBLAST: Very weak BLAST homology; helicases, including MCM4
PHYRE: low precision to Pleckstrin homology (PH) domain

>Tb927.6.3190 hypothetical protein, | protein | length=548

Peptide hits: 3 peptides
CLC PFAM search: Radical SAM superfamily -PF04055, protein Kinase domain -PF00069, Glycosyl transferase family -PF00535
NCBI BLAST: No clear hits in BLAST
PHYRE: low precision (5%) to viral capsid protein (short polypeptide), and to virion RNA polymerase (long polypeptide)

To attempt to gain further insight into the potential functions of the above “top mass spec hits” from the TbORC1/CDC6-Myc IP, position-specific iterative BLAST (PsiBLAST; http://blast.ncbi.nlm.nih.gov/Blast.cgi), Protein Homology/analogY Recognition Engine Search (PHYRE search; http://www.sbg.bio.ic.ac.uk/~phyre/) and PFAM searches (using CLC genomics Workbench v4) were carried out for each. Based on these protein primary amino acid sequence and/or protein structural or domain predictive algorithms,

similarities could be associated to two classes of proteins (Table 4-3). First, three of the proteins provided evidence that they could potentially be related to eukaryotic ORC components or Cdc6. Second, at least four of the others provided evidence that they may be RNA-associated factors. *T. brucei* protein IDs Tb927.6.1120 (14 peptides), Tb11.46.009 (5 peptides), Tb927.2.5130 (3 peptides) and Tb927.10.9350 (3 peptides) belonged to the putative RNA-associated group, while Tb09.10.3120 (10 peptides), Tb927.10.13380 (10 peptides) and Tb927.10.7980 (4 peptides) were putative ORC- or Cdc6-like proteins. Table 4-3 summarises the basis for this these assertions.

Given the ORC-focus of this project, further characterisation of the above proteins was limited to those that have ORC-like characteristics. To do this, Tb927.10.13380 and Tb927.10.7980 were chosen to test if they directly or indirectly interact with TbORC1/CDC6, and RNAi was used to examine their functions in BSF cells. Tb927.10.13380 and Tb927.10.7980 were chosen because using the CLC Workbench PFAM domain search algorithm and PHYRE these two proteins appeared to be more closely related to ORC proteins (see below for a detailed analysis).

4.8 Characterisation of Tb927.10.13380; a putative novel component of the pre-RC machinery in *T. brucei*

This section provides further experiments that illuminate the role of Tb927.10.13380 (referred to hereafter as Tb13380 for the purpose of brevity) as a putative member of the *T. brucei* pre-RC machinery; if not a divergent *T. brucei* ORC member.

4.8.1 Confirmation of interaction between Tb13380 and TbORC1/CDC6

4.8.1.1 Cloning of Tb13380 C-terminal HA tagging construct

Using the Tb13380 sequence information derived from the TriTrypDB database (http://tritrypdb.org/tritrypdb/), a construct was generated that allows *Tb13380* to be expressed as a C-terminal fusion with six copies of the HA eptiope

after integration into the endogenous locus. The strategy for this is as described in Section 4.5.2, and a 628 bp C-terminal fragment of *Tb13380* was PCR-amplified with primers CT_OL65/CT_OL66 (Table 2-2) for this purpose. The PCR product was cloned into the restriction sites *Hin*dIII and *Xba*I of the pNAT^{6HA} vector (Figure 4-11). The construct was linearised with *Pvu*II and transfected into the TbORC1/CDC6-Myc tagged line cell line (Section 4.5) and two phleomycin-resistant clones selected that should co-express TbORC1/CDC6-Myc and Tb13380-HA (see Figure 4-12 A and B for cartoon representation). As a control, the construct was also transfected into WT TREU927 procyclic form cells and two clones recovered. In each case, single clones validated for expression of the expected proteins were taken for further analysis.

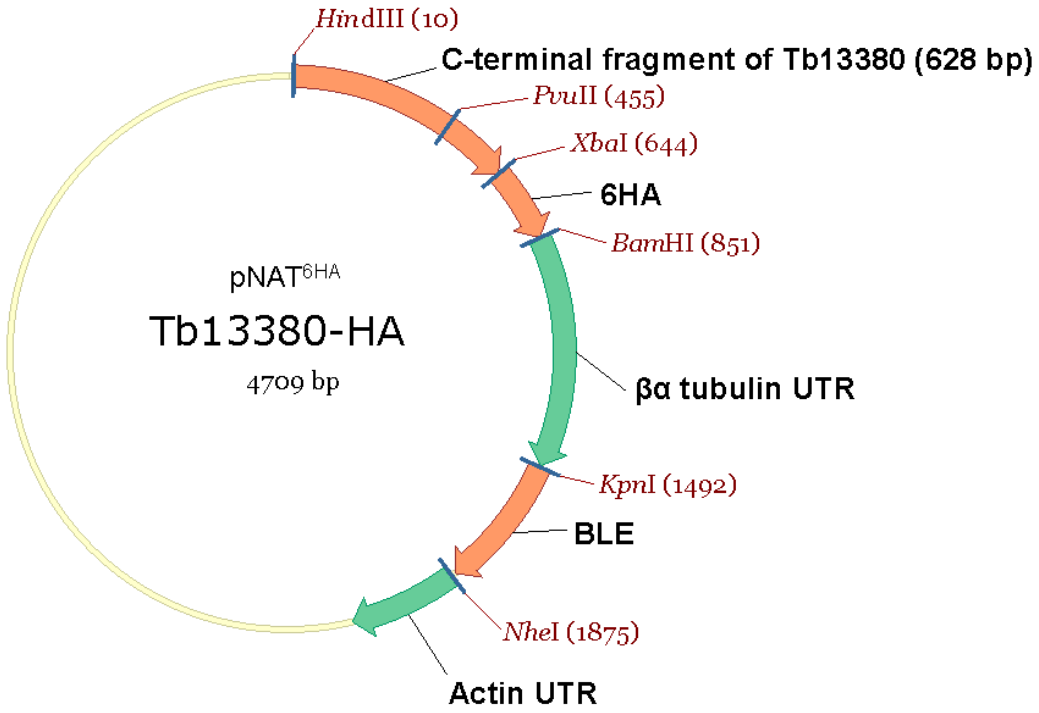


Figure 4-11 – The pNAT^{6HA} vector for *Tb13380-HA* epitope tagging

A C-terminal coding sequence of *Tb13380* (628 bp) that was cloned into *Hin*dIII and *Xba*I sites in the plasmid after PCR amplification is shown; the C-terminal *Tb13380* fragment was cloned fused to 6 tandem repeats of HA (*6HA*) to allow the expression of *Tb13380* inframe with the 6HA. The *Pvu*II site used for vector linearization to allow integration into the *Tb13380* locus is also shown. Selection of *T. brucei* transformant clones was performed using the phleomycin resistance gene (*BLE*) flanked by Actin and tubulin 3' and 5' mRNA processing regions (Actin UTR) and (βα tubulin UTR) respectively. All features shown on the vector map are designed using the VectorNTI resource (Invitrogen). The vector is originally from the lab of David Horn (Alsford & Horn 2008)

4.8.1.2 Co-immunoprecipitation of Tb13380-HA and TbORC1/CDC6-Myc

To test directly whether or not there is an interaction between TbORC1/CDC6 and Tb13380 in *T. brucei* PCF cells, we asked whether the two proteins could be co-immunoprecipitated. Whole cell extracts were prepared (as described previously, Section 4.6.3) from the TbORC1/CDC6-Myc/Tb13380-HA double expressor cells, and from Tb13380-HA and TbORC1/CDC6-Myc single expressors. Expression of both tagged proteins was checked by western blot using the corresponding antibodies (Figure 4-12, input lanes). To test for interaction, co-IP was then performed. Firstly, anti-Myc antibody was used to immunoprecipitate TbORC1/CDC6-Myc and Tb13380-HA interaction was checked by western blotting using anti-HA antibody. The reciprocal experiment was then carried using anti-HA antiserum to precipitate Tb13380-HA and checking for TbORC1/CDC6-Myc interaction with anti-Myc antiserum.

As shown on Figure 4-12 (TbORC1/CDC6-Myc IP box), anti-Myc IP of TbORC1/CDC6-Myc was successful from the TbORC1/CDC6-Myc/Tb13380-HA double expressor cells, as seen by a band of the expected size in the eluate that was absent from the Tb13380-HA control. Probing the same eluate with anti-HA antibody confirmed interaction between the proteins, as a band of the size expected for Tb13380-HA was also present in the eluate from the double expressor cell. No such band was seen in the control Tb13380-HA expressor lane, showing that the interaction is not an artefact of Tb13380-HA non-specifically binding to the anti-Myc antibody-coupled beads. The reciprocal experiment, where IP was used to 'pull-down' Tb13380-HA using anti-HA antibody from TbORC1/CDC6-Myc/Tb13380-HA double expressor or TbORC1/CDC6-Myc single expressor cells, confirmed interaction between the two proteins. In Figure 4-12 (Tb13380-HA IP box), the anti-HA western blot confirmed IP of Tb13380-HA from the double expressor, while the anti-Myc blot revealed a band of the expected size for TbORC1/CDC6-Myc, with no such band in the control lane (IP from the TbORC1/CDC6-Myc expressor cell).

These data support the previous IP-mass spectrometry data of interaction between these proteins and indicate that TbORC1/CDC6 and Tb13380 can be found in the same complex in procyclic *T. brucei* cells. Whether TbORC1/CDC6 and Tb13380 directly or indirectly interact is unknown, as is whether they

constitutively or transiently associate. Nevertheless, this may suggest the presence of a putative divergent subunit of ORC. To further understand the role of Tb13380 in *T. brucei*, we performed RNAi experiments in BSF cells and bioinformatic analyses to try to define a role for Tb13380.

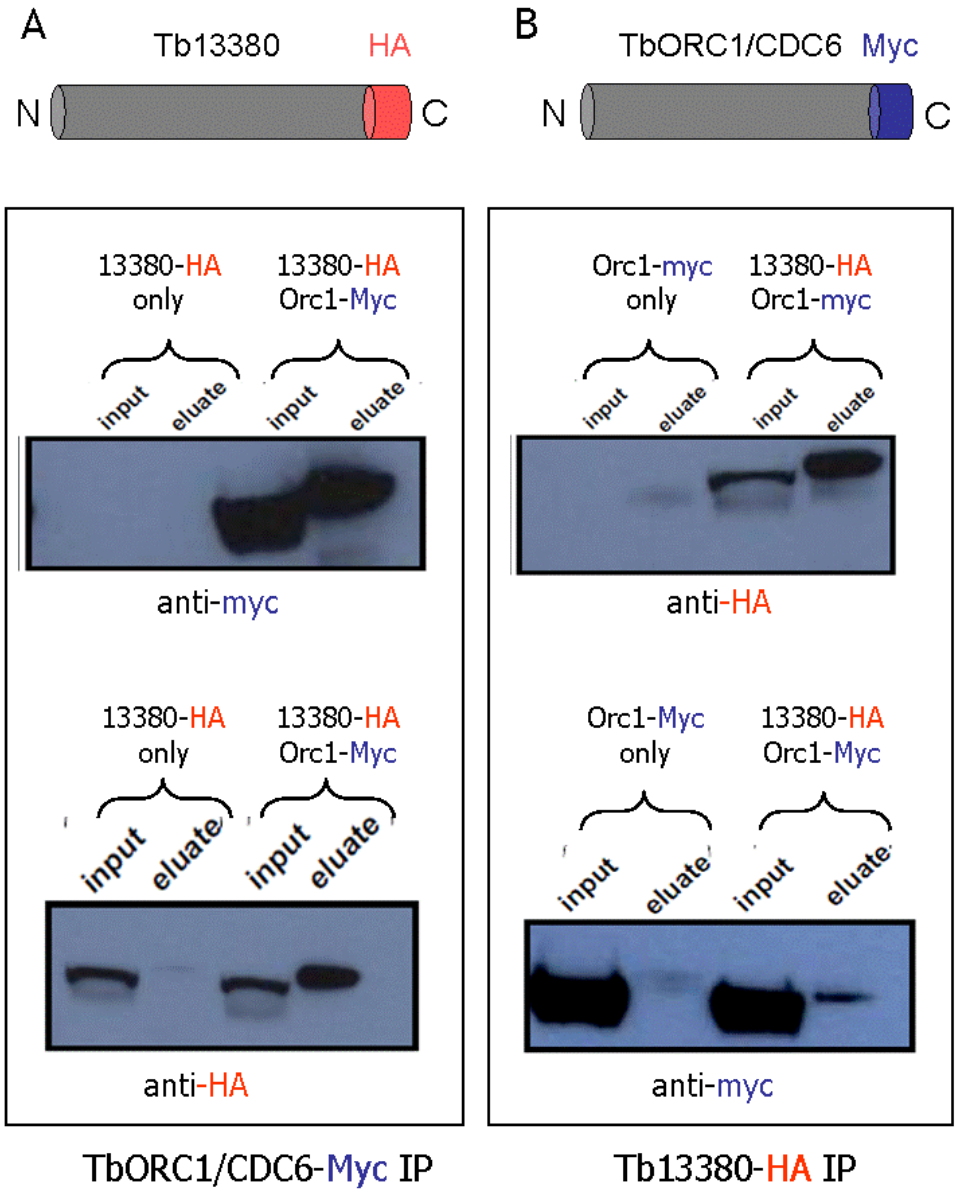


Figure 4-12 – Western blot analysis of the inputs and eluates following co-immunoprecipitation using anti-HA and anti-myc antibodys

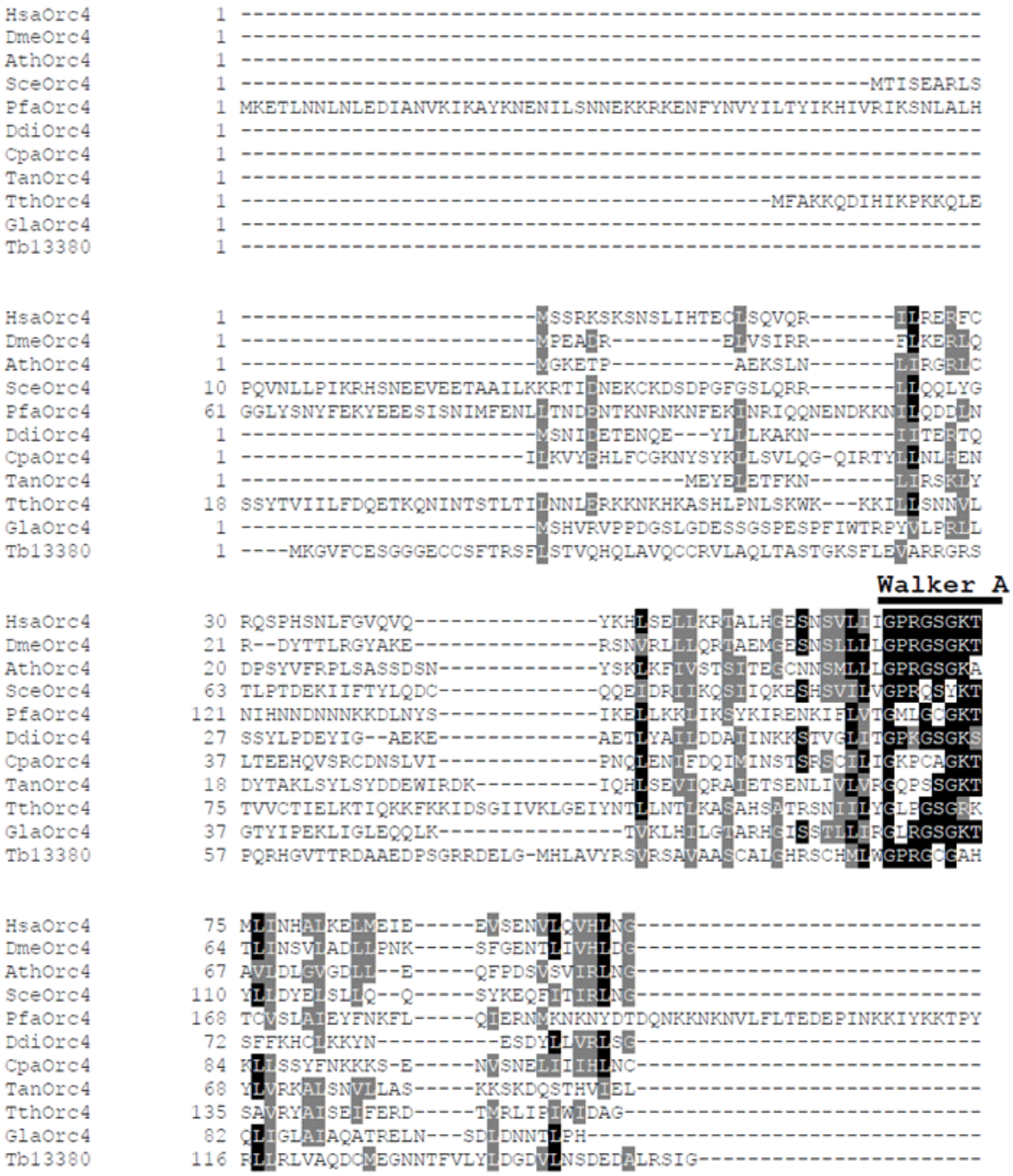
(A) and (B) show cartoon of epitope-tagged constructs of Tb13380 and TbORC1/CDC6. Whole cell extracts from cell lines expressing proteins from both constructs were subjected to co-IP using antibodies to the Myc or HA tags. “TbORC1/CDC6-Myc IP Box”: The input and eluate from co-IPs from whole cell extracts of a control cell line (13380-HA, expressing Tb13380-HA), and from co-expressor cell lines expressing Tb13380-HA (13380-HA), and TbORC1/CDC6-Myc (Orc1-Myc), were separated on a 10 % SDS-PAGE gel and then transferred to a nylon membrane. This was then probed with anti-Myc antibody (anti-myc) or with anti-HA antibody (anti-HA). “Tb13380-HA IP Box”: The input and eluate from co-IPs from whole cell extracts of a control cell line (Orc1-myc, expressing TbORC1/CDC6-Myc), and from co-expressor cell lines expressing Tb13380-HA (13380-HA), and TbORC1/CDC6-Myc (Orc1-Myc), were separated on a 10 % SDS-PAGE gel and then transferred to a nylon membrane. This was then probed with anti-HA antibody (anti-HA) or with anti-Myc antibody (anti-myc).

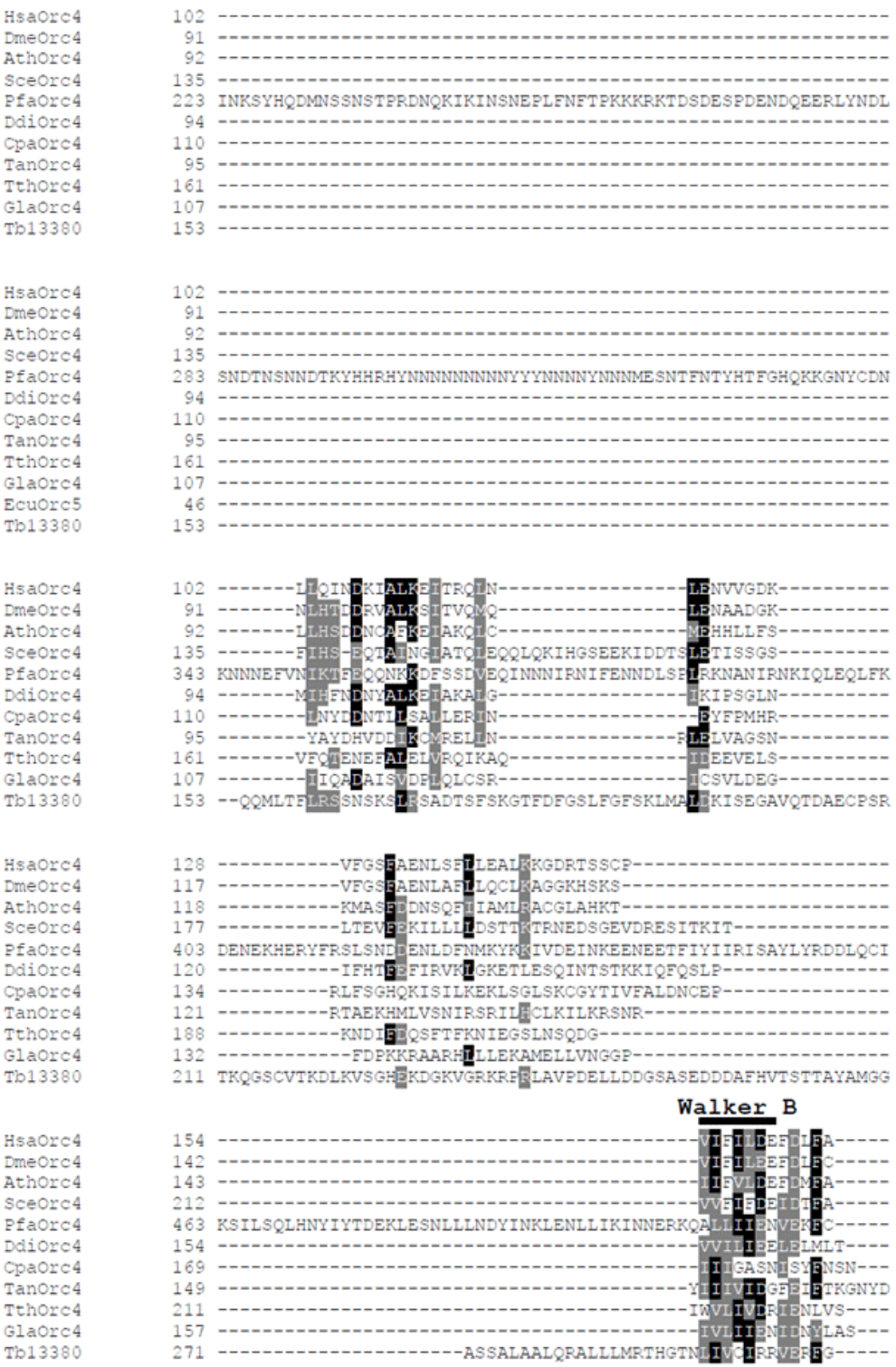
4.8.2 Amino acid sequence and structural analysis of Tb13380

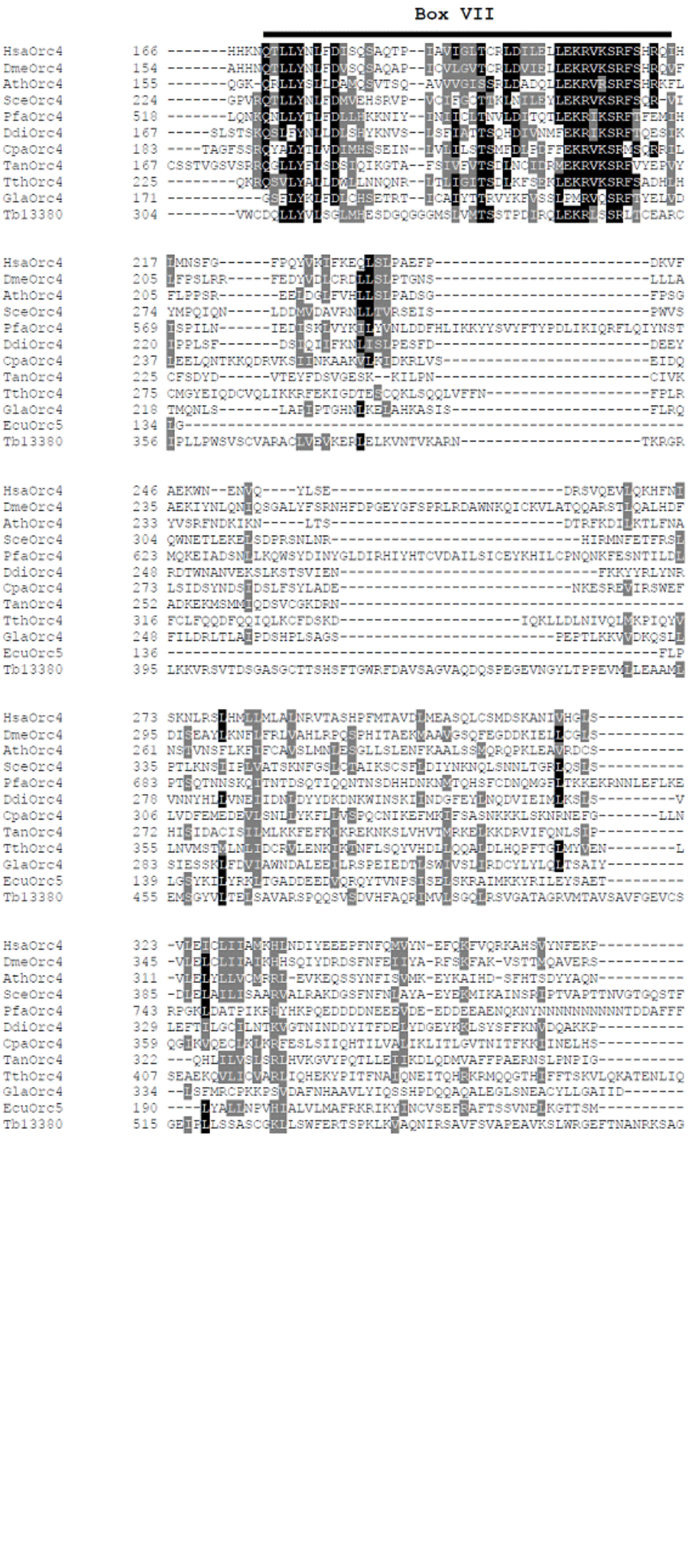
Tb13380 (also named Tb10.389.0050) is annotated as a hypothetical protein in TritrypDB (http://tritrypdb.org). The protein is predicted to be 713 amino acids long and 77.3 kDa in size. Second iteration psiBLAST searches of the NCBI database using the Tb13380 sequence from TriTrypDB as query retrieved *D. melanogaster* Orc4 (DmeOrc4) as a top hit (E-value = 9e-14) and a series of other Orc4 subunits from other organisms (0.003 ≤ E-value ≤ 1e-13). This provided a hint that this protein could be related to eukaryotic Orc4. The reciprocal BLAST search, using DmeOrc4 sequence to query TritryDB and asking if any meaningful hits emerged, was then conducted. To our surprise, considering that previous searches for ORC subunits had not yielded consistent results, the top hit (E-value 0.0021) was Tb13380. Tb13380 shows 34 % amino acid identity to DmeOrc4 for the region for which alignment was recovered by BLAST. To examine if broader sequence homology to eukaryotic Orc4 subunits was apparent, the complete predicted sequence of Tb13380 was aligned with a range of Orc4 polypeptides (see Figure 4-13). Overall this protein displays limited amino acid sequence homology to Orc4 subunits from the organisms in Figure 4-13 (for example: 33% similarity and 16% identity to DmeOrc4; and 33% similarity and 15% identity to *Arabidopsis thaliana* Orc4). Based on primary amino acid sequence homology to Orc4 proteins from higher eukaryotes, Tb13380 appears to possess homology to Walker A and B boxes, and a sensor 1 motif, found with AAA+ ATPase proteins (Neuwald et al, 1999). However, Tb13380 lacks a crucial lysine residue (conserved Lysine at position 62 of DmeOrc4; position 108 of *S. cerevisiae* Orc4 replaced by Alanine at position 114 in Tb13380) in the Walker A motif (Figure 4-13), which binds the nucleotide, and may also have a degenerate Walker B box (Figure 4-13), which binds Mg++, suggesting it may not possess ATPase or a GTPase activity, a critical regulatory function of members of the ORC family of proteins and AAA+ ATPases in general (Speck *et al.* 2005; Neuwald *et al.* 1999). Despite this, there is quite considerable sequence homology between the *T. brucei* protein and eukaryotic Orc4 proteins, in particular around the Walker B box and Box VII motif, perhaps indicative of Orc4 orthology (Figure 4-13). The Box VII motif possesses a key arginine residue at position 281 of *E. coli* DnaA that has been proposed to interact with the γ-phosphate of the adenine nucleotide to sense its bound state (Erzberger, Pirruccello, & Berger 2002b; Neuwald *et al.*

1999). This interaction is meant to coordinate ATP hydrolysis with a conformational change of the ATPAse protein (Erzberger, Pirruccello, & Berger 2002a). It is worth mentioning that although Tb13380 lacks the canonical Walker A and Walker B boxes, the Box VII motif appears to be conserved with other eukaryotic Orc4 proteins (Figure 4-13).

Tb13380 is also conserved and syntenic across the sequenced kinetoplastids: orthologues exist in all the Tritryps, are annotated as hypothetical proteins and are of a similar size. *T. cruzi* has two proteins (Tc00.1047053506357.20 and Tc00.1047053511277.92) that are 97.1 % identical to each other and 54% identical to "Tb13380", while *L. major* encodes a protein (LmjF18.0720) with 38% identity to Tb13380.


Walker A


Walker B


Box VII
HsaOrc4 166 -------HHKNQTLLYNLFDISQSAQTP--IAVIGLTCRLDILELLEKRVKSRFSHRQIH
DmeOrc4 154 -------AHHNQTLLYNLFDVSQSAQAP--ICVLGVTCRLDVIELLEKRVKSRFSHRQVF
AthOrc4 155 -------QGK-QRLLYSLLDPMQSVTSQ--AVVVGISSRLDADQLLEKRVRSRFSHRKFL
SceOrc4 224 -------GPVRQTLLYNLFDMVEHSRVP--VCIFGCTNKLNILEYLEKRVKSRFSQR-VI
PfaOrc4 518 -------LQNKQNLLYTLFDLLHKKNIY--INIICLTNVLDLIQTLEKRIKSRFTFEMIH
DdiOrc4 167 -----SLSTSKQSIEYNLLDLSHYKNVS--ISFIATTSQHDIVNMFEKRIKSRFIQESIK
CpaOrc4 183 ----TAGFSSRQYPLYTLVDIMHSSEIN--IVLILSTSMFDLEDFFEKRVKSRMSQRRIL
TanOrc4 167 CSSTVGSVSPRQGLLYFLSDSIQIKGTA--FSIVFVTSILNCIDRMEKRVKSREVYEPVY
TthOrc4 225 --------QMRQSVLYALLDKLLNNQNR--VTLIGITSILKFSPKLEKRVKSRFSADHLH
GlaOrc4 171 -----------GSELYKLFDLCHSETRT--ICAIYTTRVYKEVSSLPMRVQSRFTYELVD
Tb13380 304 --------VWCDQLLYVLSGLMHESDGQGGGMSLVMTSSTPDIRQLEKRLSSRLICEARC
HsaOrc4 217 LMNSFG------FPQVVKIFKEQLSLPAEFP-----------------------DKVF
DmeOrc4 205 LFPSLRR-----FEDYVDLCRDLLSLPTGNS-----------------------LLLA
AthOrc4 205 FLPPSR-------EEIDGIFVHLLSLPADSG-----------------------FPSG
SceOrc4 274 YMPQIQN-----LDDMVDAVRNLLTVRSEIS-----------------------PWVS
PfaOrc4 569 ISPILN------IEDISKIVVKILYVNLDDFHLIKKYYSVYFTYPDLIKIQRFLQIYNST
DdiOrc4 220 IPPLSF-------DSIQILFKNLISLPESFD-----------------------DEEY
CpaOrc4 237 LEELQNTKKQDRVKSLINKAAKVLKIDKRLVS----------------------EIDQ
TanOrc4 225 CFSDYD------VTEYFDSVGESK--KILPN----------------------CIVK
TthOrc4 275 CMGYEIQDCVQLIKKRFEKIGDTESCQKLSQQLVFFN-----------------FPLR
GlaOrc4 218 TMQNLS-------LAFIPTGHNIKEIAHKASIS----------------------FLRQ
EcuOrc5 134 IG----------
Tb13380 356 IPLLPWSVSCVARACLVEVKEPLELKVNTVKARN---------------------TKRGR
HsaOrc4 246 AEKWN--ENVQ----YLSE-----------------------DRSVQEVLQKHFNI
DmeOrc4 235 AEKIYNLQNIQSGALYFSRNHFDPGEYGFSPRLRDAWNKQICKVLATQQARSTVQALHDF
AthOrc4 233 YVSRFNDKIKN-----LTS-----------------------DTRFKDILKTLFNA
SceOrc4 304 QWNETLEKEISDPRSNLNR-----------------------HIRMNFETFRSF
PfaOrc4 623 MQKEIADSNILKQWSYDINYGLDIRHIYHTCVDAILSICEYKHILCPNQNKFESNTILDL
DdiOrc4 240 RDTWNANVEKSLKSTSVIEN----------------------FKKYYRLYNR
CpaOrc4 273 LSIDSYNDSIDSLFSYLADE----------------------NKESREVIRSWEF
TanOrc4 252 ADKEKMSMMIQDSVCGKDRN-----------------------
TthOrc4 316 FCLFQQDFQQIQLKCFDSKD----------------IQKLLDLNIVQLVKPIQYV
GlaOrc4 240 FILDRLTLAIPDSHPLSAGS----------------FEPTLKKVVDKQSLL
EcuOrc5 136 ----------
Tb13380 395 LKKVRSVTDSGASGCTTSHSFTGWRFDAVSAGVAQDQSPEGEVNGYLTPPEVMILEAAML
HsaOrc4 273 SKNLRSLHMLLMLAINRVTASHPFMTAVDIMEASQLCSMDSKANIVHGLS----------
DmeOrc4 295 DISEAYLKNFLFRLVAHLRPQSPHITAEKVAAVGSQFEGDDKIELLCGLS----------
AthOrc4 261 NSLVNSFLKEIFCAVSLMNIPSGLLSLENFKAALSSVQRQPKLEAVRDCS----------
SceOrc4 335 PTLKNSIIPLVATSKNFGSLQIAIKSCSFLDIYNKNQLSNNLTGFLQSLS----------
PfaOrc4 683 PTSQTNNSKQITNTDSQTIQQNTNSDHHDNKNVTQHSFCDNQMGFLTKKEKRNNLEFLKE
DdiOrc4 278 VNNYHLLVNEIIDNIDYYDKDNKWINSKIINDGFEYINQDVIEIMLKSLS---------V
CpaOrc4 306 LVDFEMEDEVISNLIYKFLIVSPQCNIKEFMKIFSASNKKKLSKNRNEFG-------LLN
TanOrc4 272 HISIDACISLLMLKKFEFKIKREKNKSLVHVIVRKEIKKDRVIFQNLSIP----------
TthOrc4 355 LNVMSTVLNLIDCFVLENKIPINFLSQYVHDLIQQALDLHQPFTGLMYQEN--------L
GlaOrc4 283 SIESSKLFDVIAWNDALEELLRSPEIEDTLSWIVSLIRDCYLYLQLTSAIY---------
EcuOrc5 139 LGSYKILYRKLTGADDEELVQRQYTVNPSISDLSKRAIMKKYRILEYSAET---------
Tb13380 455 EMSGYVLTELSAVARSPQQSVSDVHFAQRIMVLSGQIRSVGATAGRVMTAVSAVFGEVCS
HsaOrc4 323 -VFELCLIIAMKHINDIYEEEPFNFQMVYN-EFQKFVQRKAHSVYNFEKP----------
DmeOrc4 345 -VLELCLIIAIKHHSQIYDRDSFNFELIYA-RFSKFAK-VSTTVQAVERS----------
AthOrc4 311 -VFELYLLVQMRRI-EVKEQSSYNFISVMK-EYKAIHD-SFHTSDYYAQN----------
SceOrc4 385 -DIELAILISAARVALRAKDGSFNFNLAYA-EYEKMIKAINSFAPTVAPTTNVGTGQSTF
PfaOrc4 743 RPGKLDATPIKRHYHKPQEDDDDNEEEVDE-EDDEEAENQKNYNNNNNNNNNNTDDAFFF
DdiOrc4 329 LEFTILGCHLNTKVGTNINDDYITFDELYDGEYKKLSYSFFKNVDQAKKP----------
CpaOrc4 359 QGIKMQECFKLKRFESLSIIQHTILVALIKLITLGVTNITFKKIINELHS----------
TanOrc4 322 ---QHLILVSLSRLHVKGVYPQTLLEIIKDLQDMVAFFPAERNSLPNPIG----------
TthOrc4 407 SEAEKQVLICVARLIQHEKYPITFNAIQNEITQHRKRMQQGTHIFFTSKVLQKATENLIQ
GlaOrc4 334 --ESFMRCPKKPSVDAFNHAAVLYIQSSHPDQQAQALEGLSNEACYLLGAIID--------
EcuOrc5 190 ----LYALINPVHIALVLMAFRKRIKYINCVSEFRAFTSSVNEIKGTTSM----------
Tb13380 515 GEIPLLSSASCGKLLSWFERTSPKLKVAQNIRSAVFSVAPEAVKSLWRGEFTNANRKSAG

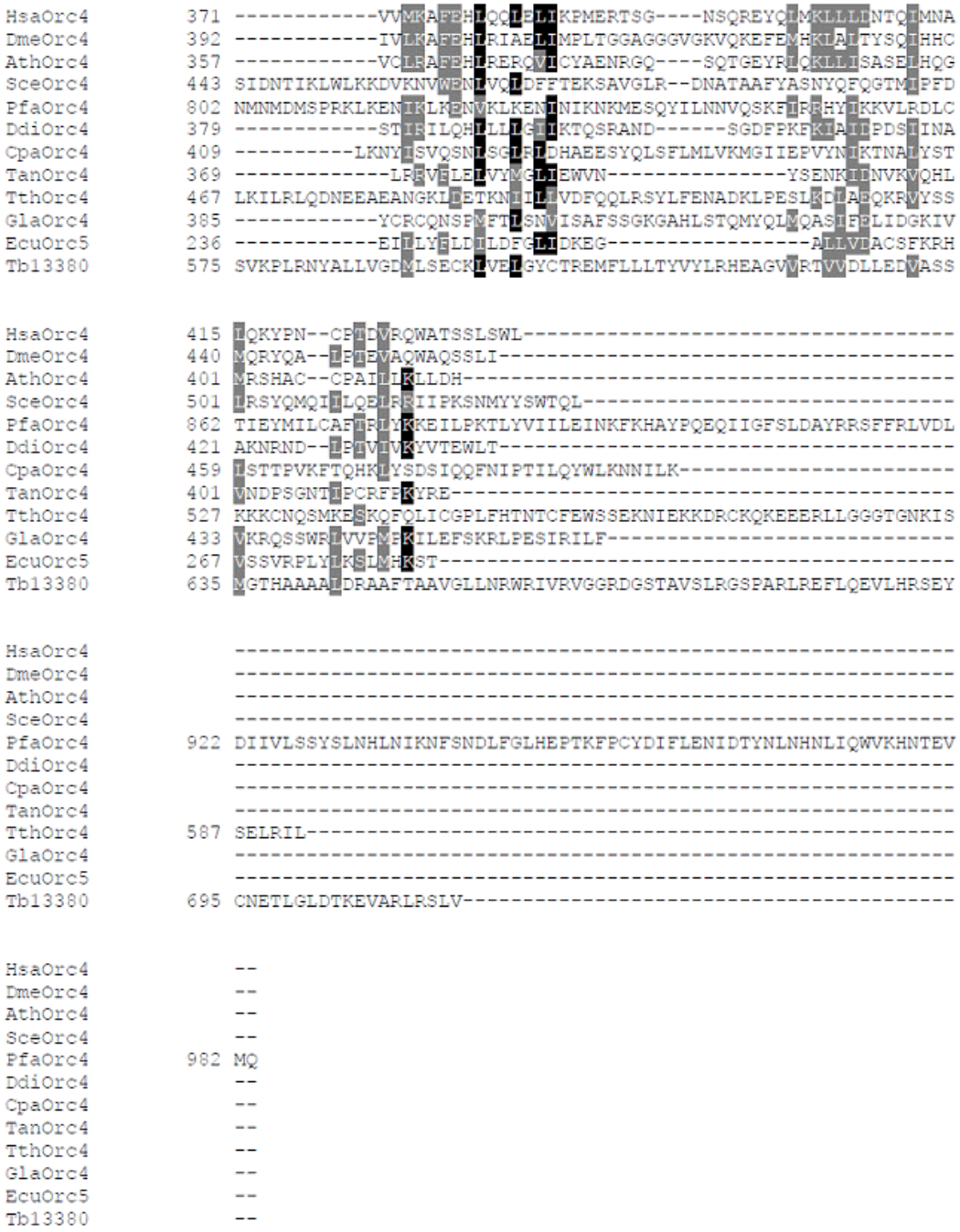


Figure 4-13 – Amino Acid sequence alignment of eukaryotic Orc4 sequences with Tb13380

Sequence alignment of eukaryotic Orc4 proteins from Hsa, Homo sapiens; Dme, *Drosophila melanogaster*; Ath, Arabidopsis thaliana; Sce, Saccharomyces cerevisiae; Pfa, *Plasmodium falciparum*; Ddi, *Dictyostelium discoideum*; Cpa, *Cryptosporidium parvum*; Tan, *Theileria annulata*;; Tth, *Tetrahymena thermophila*; Gla, *Giardia lamblia*; Ecu, and *Encephalitozoon cuniculi*; with Tb13380. The alignment was generated using Clustal X and was coloured (in black and grey) in boxshade (http://www.ch.embnet.org/software/BOX_form.html). Black bars indicate conserved sequence motifs: Walker A; Walker B; and Box VII signature sequence. Black indicates identical residues in 50% of sequences; grey indicates conserved in 50% of sequences. Genbank accession numbers are provided Appendix 2.

To ask if further functional elements of Tb13380 could be identified, domain prediction using a PFAM prediction tool in CLC Genomics workbench v4 (www.clcbio.com) was performed. This confirmed that Tb13380 belongs to the AAA^+ superfamily of proteins, a critical feature of the ORC family of proteins, except for Orc6 which lacks the domain (Erzberger & Berger 2006; Speck *et al.* 2005; Neuwald *et al.* 1999). The PFAM domain searches also revealed that Tb13380 belonged to the tetratricopeptide repeat (TPR) family of proteins, which are known to mediate protein-protein interactions and are often involved in the assembly of multi-protein complexes (Das, Cohen, & Barford 1998; Lamb, Tugendreich, & Hieter 1995). When Orc4 subunit PFAM domain searches were done in CLC using the same tool, the TPR repeat was predicted in *P. falciparum* Orc4 family, while *A. thaliana, H. sapiens* also showed conservation of the AAA^+ motif (Figure 4-14). These observations may support a role for Tb13380 in ORC complex formation in *T. brucei*.

A Phyre search (web-based prediction algorithm for PDB-derived protein fold recognition using Psi-BLAST) revealed two predictions of Tb13380 being structurally similar to archaeal Orc1/Cdc6 (*Aeropyrum pernix* and *Solfolobus Solfataricus*). This lent further support for Tb13380 being a *T. brucei* ORC-related factor.

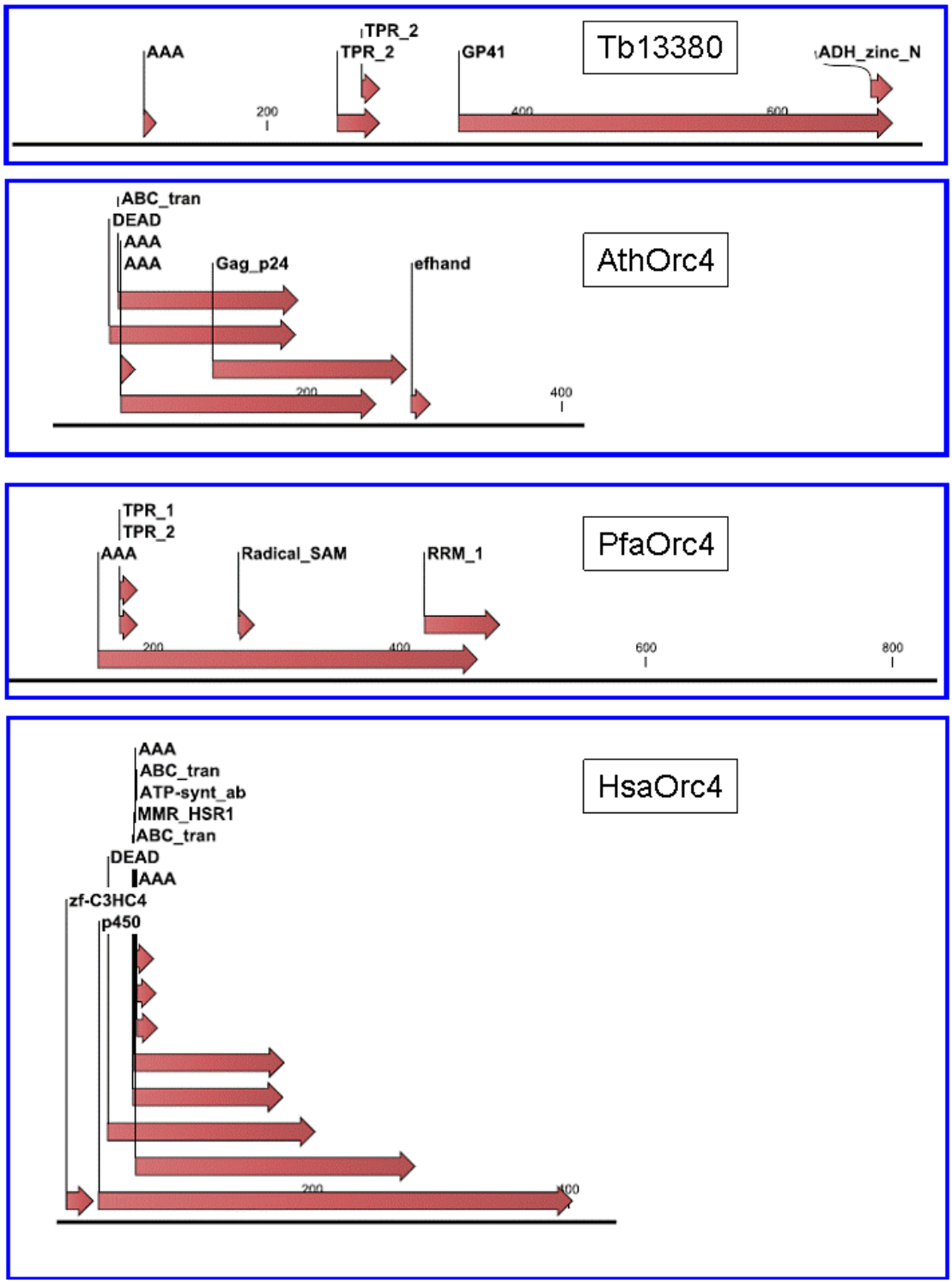


Figure 4-14 – PFAM domain prediction of Tb13380 compared with selected eukaryotic Orc4p

Using the "PFAM 100 most common" domain prediction tool in CLC Genomics Workbench v4, putative functional domain organisation of Tb13380 (Tb13380 box) is shown relative to the same analysis of *A. thaliana* Orc4 (AthOrc4 box), *P. falciparum* Orc4 (PfOrc4 box) and *Homo sapiens* Orc4 (HsaOrc4 box). An ATPase motif is indicated by AAA (PFAM no. PF00004); tetratricopeptide repeats are indicated by TPR1 and TPR2 (PFAM no. PF07719); Zinc-binding dehydrogenase indicated by ADH_zinc_N (PFAM no. PF00107); Radical SAM superfamily is indicated by Radical SAM (PFAM no. PF04055); ABC transporters indicated by ABC_tran (PFAM no. PF00005); GTPase motif indicated by MMR_HRS1 (PFAM no. PF01926); RNA recognition motif indicated by RRM_1 (PFAM no. PF00076); DEAD/DEAH box helicase indicated by DEAD (PFAM no. PF00270); Zinc finger, C3HC4 type indicated by zf-C3HC4 (PFAM no. PF00097); PFAM numbers (PFAM no.) obtained from the Sanger Institute webpage available at (http://pfam.sanger.ac.uk/).

Finally, having potentially identified Tb13380 domains indicative of an ORC-like protein, we questioned if the protein has structural motifs that may implicate it having a nucleic acid binding role. An example is seen in *Schizosacchomyreces pombe* where its Orc4 subunit has been shown to bind to specific sites within its DNA replication origins (Kong & DePamphilis 2001a). PsiPred (http://bioinf.cs.ucl.ac.uk/psipred/) and CLC Genomics predicted structural folds for Tb13380 that are also found in winged helix proteins, typical of all ORC members (Orc1, Orc4, and Orc5) which are nucleic acid binding proteins.

4.8.3 RNAi of Tb13380 results in phenotypes analogous to TbORC1/CDC6 RNAi in BSF T. brucei

4.8.3.1 Cloning of *Tb13380* RNAi construct

In order to understand the role played by Tb13380 in *T. brucei*, we carried out targeted depletion of *Tb13380* transcript by RNAi in BSF cells. *T. brucei* Lister 427 pLew90-pLew13 BSF cells were transfected with the pZJM vector containing a 562 bp fragment of Tb13380. The selection of RNAi target region, primer design and cloning procedure was performed as described in Section 3.3.1. The primers CT_OL63/CT_OL64 (Table 2-2) containing restriction sites *Bam*HI (forward primer) or *Hin*dIII (reverse primer) were used to clone the *Tb13380* RNAi fragment into the pZJM vector (Figure 4-15). Linearization, transfection procedures and selection of clones are same as described in Section 3.3.1.

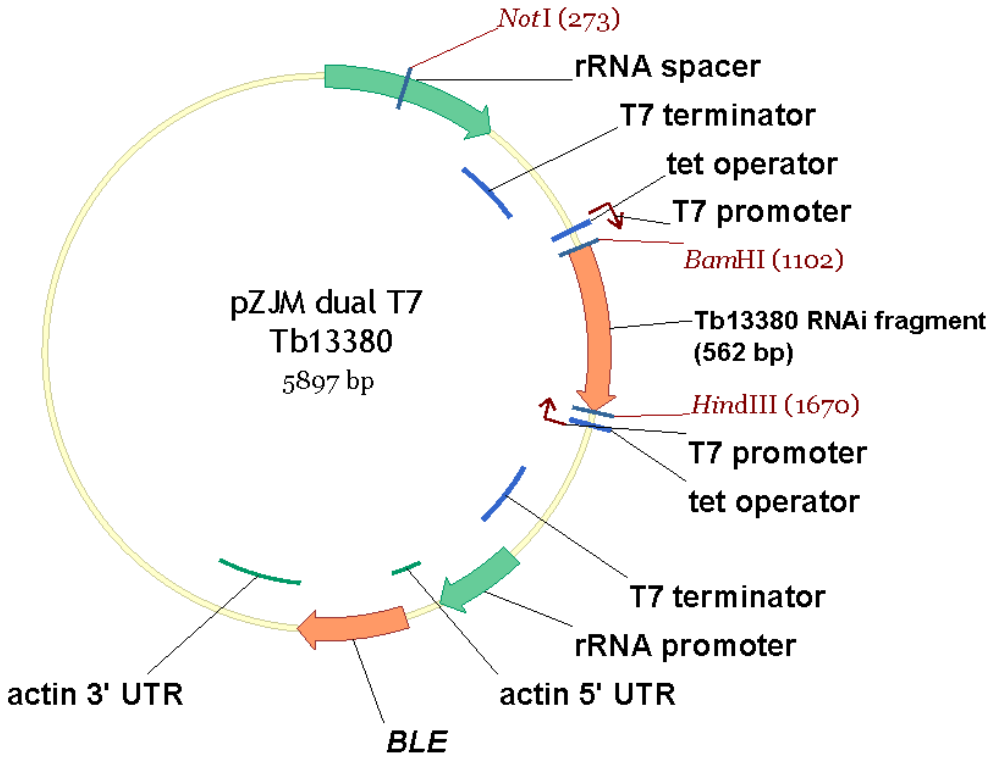


Figure 4-15 – The pZJM dual T7 vector for *Tb13380* RNAi

A fragment of *Tb13380* (562 bp) that was cloned into *Bam*HI and *Hin*dIII sites in the plasmid, after PCR amplification, is shown; Vector properties are same as described previously in Figure 3-1

4.8.3.2 Effect of Tb13380 RNAi on growth in bloodstream form cells

After selection with phleomycin, two antibiotic-resistant clones were selected and growth was monitored up to 36 hrs post induction of RNAi with 2 $\mu g.ml^{-1}$ of tetracycline. Comparing the growth curves of RNAi-induced and uninduced cells showed that each clone exhibited a severe growth defect, with reduced cell density visible as early as 8 hrs post induction of RNAi (Figure 4-16). After 18 hrs, the RNAi-induced cell numbers dropped, indicating the cells were dying.

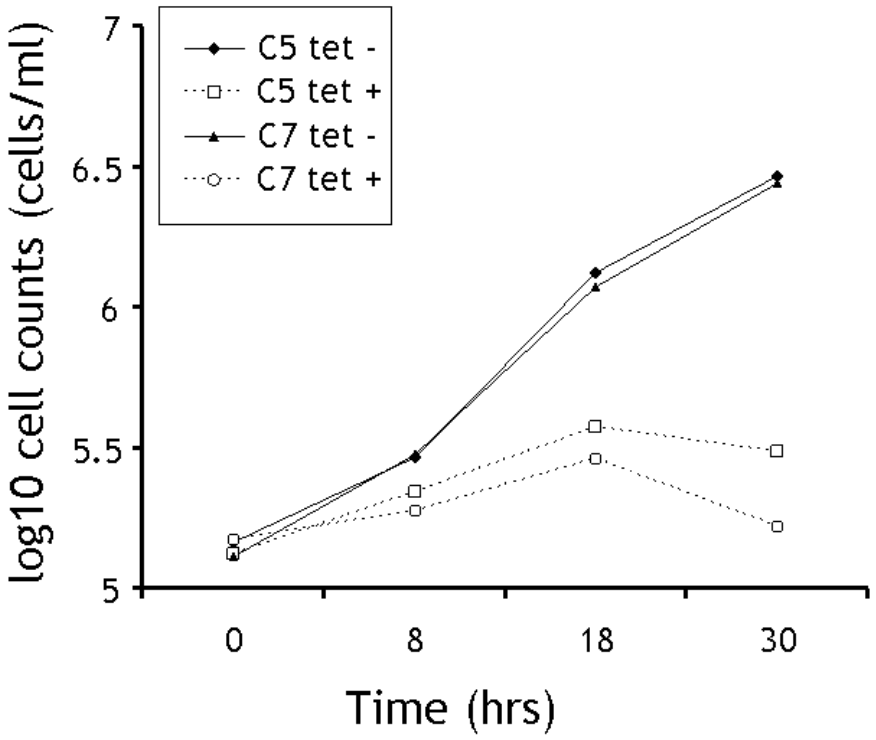


Figure 4-16 – Effect of *Tb13380* RNAi in bloodstream form cells

Growth was monitored in bloodstream cell lines in which *Tb13380* RNAi was induced (+ Tet, broken line) or without induction (-Tet, solid line) for 2 clones; C5 and C7

4.8.3.3 Effect of Tb13380 RNAi on cell cycle and cell morphology in bloodstream form cells

DAPI-staining of RNAi induced and non-induced cells (Figure 4-17) showed the accumulation of multi-nucleate and multi-kinetoplast cells 18 and 26 hrs post induction of RNAi, with a concomitant loss of 1N1K cells. In contrast, the distribution of 1N1K, 1N2K and 2N2K cells remained normal in non-induced samples up to 26 hrs, and aberrant cells represented only a small fraction of the population. To further understand the effect of *Tb13380* RNAi on the cell cycle, flow cytometry was performed on PI-stained RNAi-induced versus non-induced cells. FACS of *Tb13380* RNAi BSF cells revealed a decrease in the proportion of cells with 2C and 4C DNA content with a concomitant increase in the proportion of cells with 8C DNA content (Figure 4-18). This observation is in agreement with the accumulation of multinucleate cells observed from DNA DAPI staining experiments.

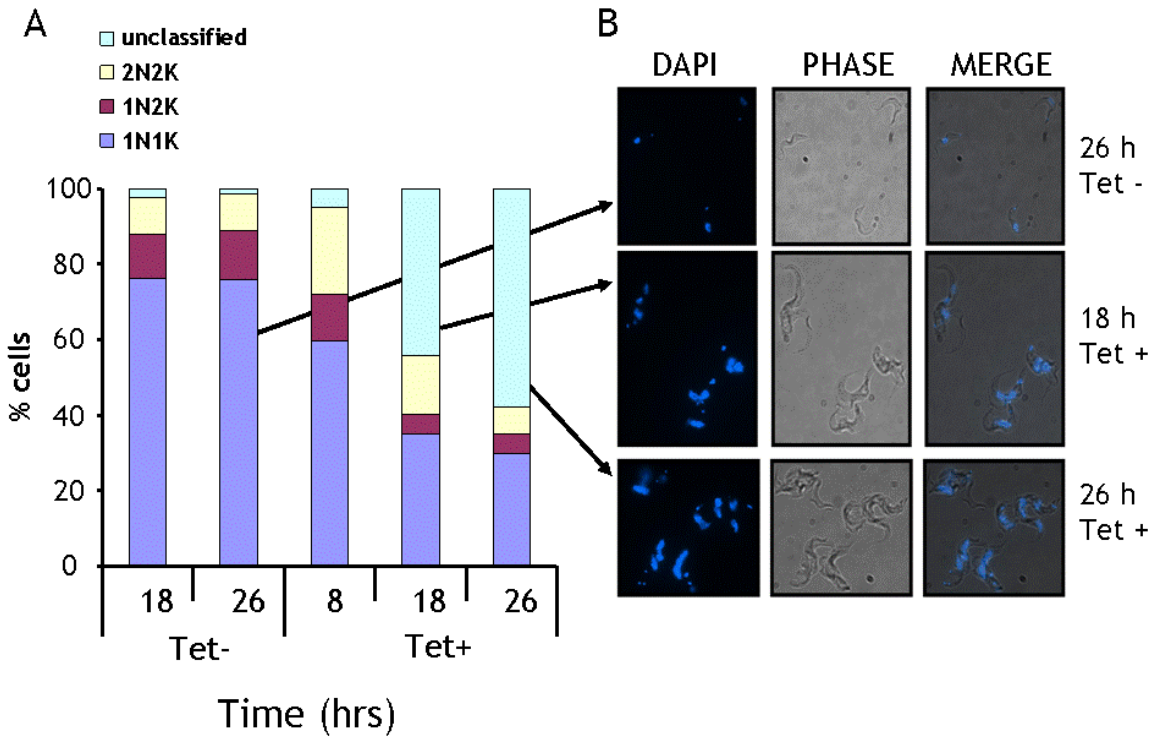


Figure 4-17 – Analysis of nuclear and kinetoplast DNA configuration of *Tb13380* RNAi cells for a single clone

(A) Microscopic of counting nuclear (N) and kinetoplast (K) DNA content after DAPI staining. ~ 100 bloodstream form *T. brucei* cells were counted at each time point for non RNAi-induced control cells (A: 18, 26 hrs; Tet -), or cells depleted in Tb13380 by RNAi (A: 8, 18, 26 hrs; Tet +). (B) Images of normal cells (Tet –, 26 hrs) and abnormal cells after RNAi induction (Tet +; 18 and 26 hrs). DNA is shown by DAPI stain (DAPI); phase contrast image (PHASE), and an overlay of the DAPI and phase images (MERGE).

Though no analysis has yet been performed to quantify the RNAi-induced knockdown of Tb13380 transcripts in the above experiments the timing of the growth arrest and subsequent cell death, as well as the nature of the aberrant cells that arise, is highly reminiscent of the phenotypes that result from *TbORC1/CDC6* RNAi in BSF *T. brucei* cells.

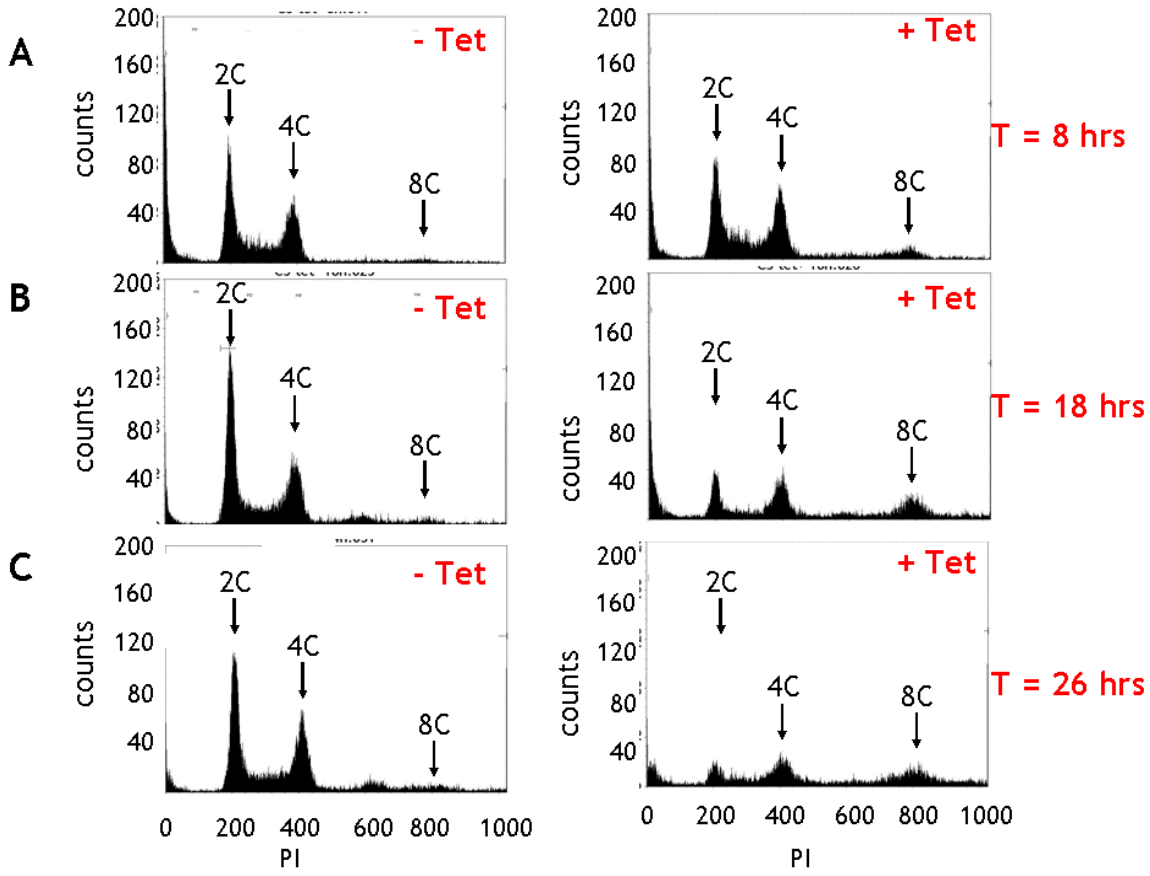


Figure 4-18 – FACS profiles of PI-stained cells after *Tb13380* RNAi

Histograms show propidium iodide-stained *T. brucei* procyclic-form cells after FACS sorting, sampled 8 hrs (A), 18 hrs (B), 26 hrs (C) pre- and post - induction of RNAi for *Tb13380* (- Tet and + Tet ,respectively). The peaks corresponding with cells containing 2C and 4C DNA content are indicated, as is the peak position for cells with 8C content; where C represents a haploid DNA content

4.9 Characterisation of Tb927.10.7890; another putative novel component of the pre-RC machinery in *T. brucei*

This section will describe preliminary analyses that have been carried out to define a role for the *T. brucei* protein Tb927.10.7890 (hereafter referred to as Tb7980 for the purpose of brevity). As earlier mentioned, Tb7980 was identified as a putative TbORC1/CDC6 interacting partner by co-immunoprecipitation and subsequent mass spectrometry analyses.

4.9.1 Confirmation of interaction between Tb7980 and TbORC1/CDC6

4.9.1.1 Cloning of Tb7980-HA tagging construct

To determine if Tb7980 and TbORC1/CDC6 interact *in vivo* in PCF cells, the *Tb7980* DNA sequence was retrieved from the TriTrypDB database (http://tritrypdb.org/tritrypdb/), and a construct was generated that allows *Tb7980* to be expressed as a C-terminal fusion with six copies of the HA epitope after integration into the endogenous loci. The strategy for this is as described in Section 4.5.2. A 633 bp C-terminal fragment of *Tb7980* was PCR-amplified with primers CT_OL67/CT_OL68 (Table 2-2) and the PCR product was cloned into the restriction sites *Hin*dIII and *Xba*I of the pNAT^{6HA} vector (Figure 4-19). The construct was linearised with *Xho*I, transfected into the TbORC1/CDC6-Myc tagged line cell line (Section 4.5) and two phleomycin-resistant clones selected that should co-express TbORC1/CDC6-Myc and Tb7980-HA (A and B). As a control, the construct was also transfected into WT TREU927 procyclic form cells and two clones recovered. In each case, single clones validated for expression of the expected proteins were taken for further analysis

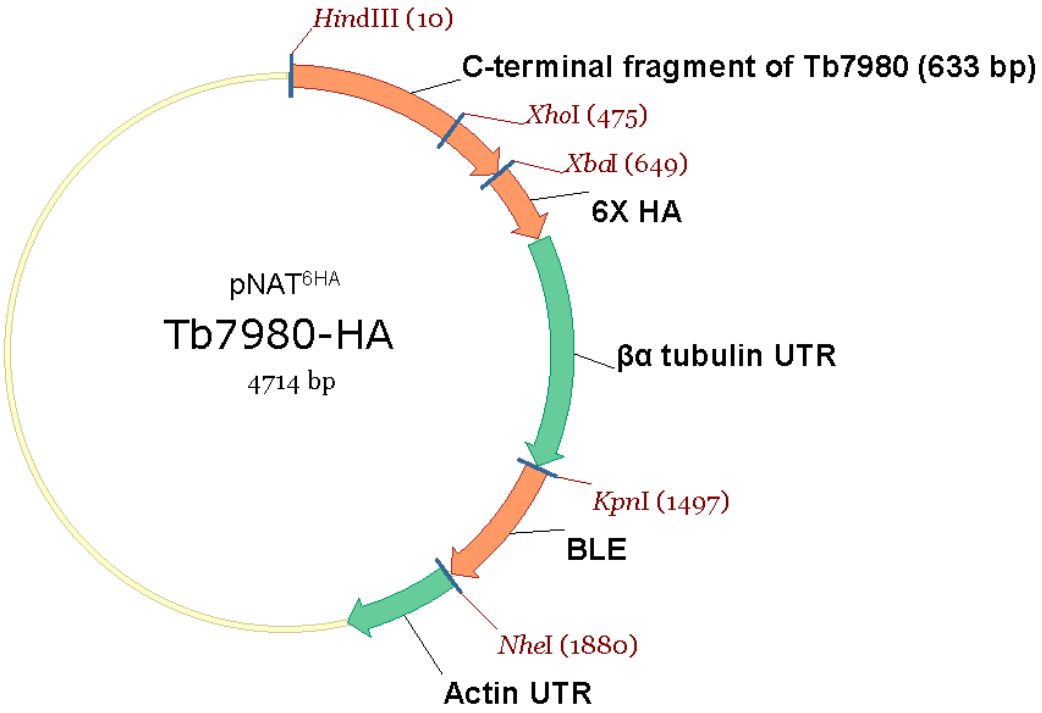


Figure 4-19 – The pNAT^{6HA} vector for *Tb7980-HA* epitope tagging

A C-terminal coding sequence of *Tb7980* (633 bp) that was cloned into *Hin*dIII and *Xba*I sites in the plasmid after PCR amplification is shown. The *Xho*I site used for vector linearization to allow integration into the *Tb7980* locus is also shown. Properties of vectors have been described previously (Figure 4-11)

4.9.1.2 Co-immunoprecipitation of Tb7980-HA and TbORC1/CDC6-Myc

We examined by co-immunoprecipitation whether Tb7980-HA and TbORC1/CDC6-Myc interact. The procedure to validate the interaction between Tb7980-HA and TbORC1/CDC6-Myc was carried out as described for Tb13380-HA and TbORC1/CDC6-Myc in Section 4.8.1.2. Expression of both tagged proteins was checked by western blot using the corresponding antibodies to the HA and Myc tags (Figure 4-20, input lanes).

As shown on Figure 4-20 (TbORC1/CDC6-Myc IP box), anti-Myc IP of TbORC1/CDC6-Myc was successful from the TbORC1/CDC6-Myc/Tb7980-HA double expressor cells, as seen by a band at 66 kDa in the eluate that was absent from the Tb7980-HA control (Figure 4-20; TbORC1/CDC6-Myc IP box; anti-myc panel). The interaction between the proteins was inconclusive when the

same eluate was probed with anti-HA antibody, because a band of the size expected for Tb7980-HA (~48 kDa) was present in both the eluate from the double expressor cell and in the control Tb7980-HA expressor lane (Figure 4-20; TbORC1/CDC6-Myc IP box; anti-HA panel). This band most probably results from the eluted IgG heavy chain fragment from the anti-myc antibody used for the IP. The reciprocal experiment, where IP was used to 'pull-down' Tb7980-HA using anti-HA antibody from TbORC1/CDC6-Myc/Tb7980-HA double expressor or TbORC1/CDC6-Myc single expressor cells, confirmed interaction between the two proteins. In Figure 4-20 (Tb7980-HA IP box; anti-HA panel), from the anti-HA western blot, a definitive conclusion could not be made about the success of the IP of Tb7980-HA from the double expressor, since the same band was present in both the control lane as well as the double expressor lane. However, the anti-Myc blot from the same eluate revealed a band of the expected size (66 KDa) for TbORC1/CDC6-Myc from the TbORC1/CDC6-Myc/Tb7980-HA double expressor cells; with no such band in the control lane. Thus, with the Tb7980-HA IP an interaction with TbORC1/CDC6-Myc was confirmed suggesting that the inconclusive results obtained for the TbORC1/CDC6-IP was due to co-localisation of the putative antibody band and the eluted immunoprecipitated Tb7980-HA at ~48 kDa.

Until the resulting eluates are separated by running the SDS-PAGE for a longer period of time on a gradient gel, a definitive conclusion to confirm the previous IP-mass spectrometry data of interaction between these proteins cannot be reached. However, based on the reciprocal experiment above, it is likely that both data would complement each other. To further understand the potential role of Tb7980 and determine whether it is essential for viability, its expression in bloodstream stage parasites was knocked down using the inducible RNAi system described previously (Section 4.8.3).

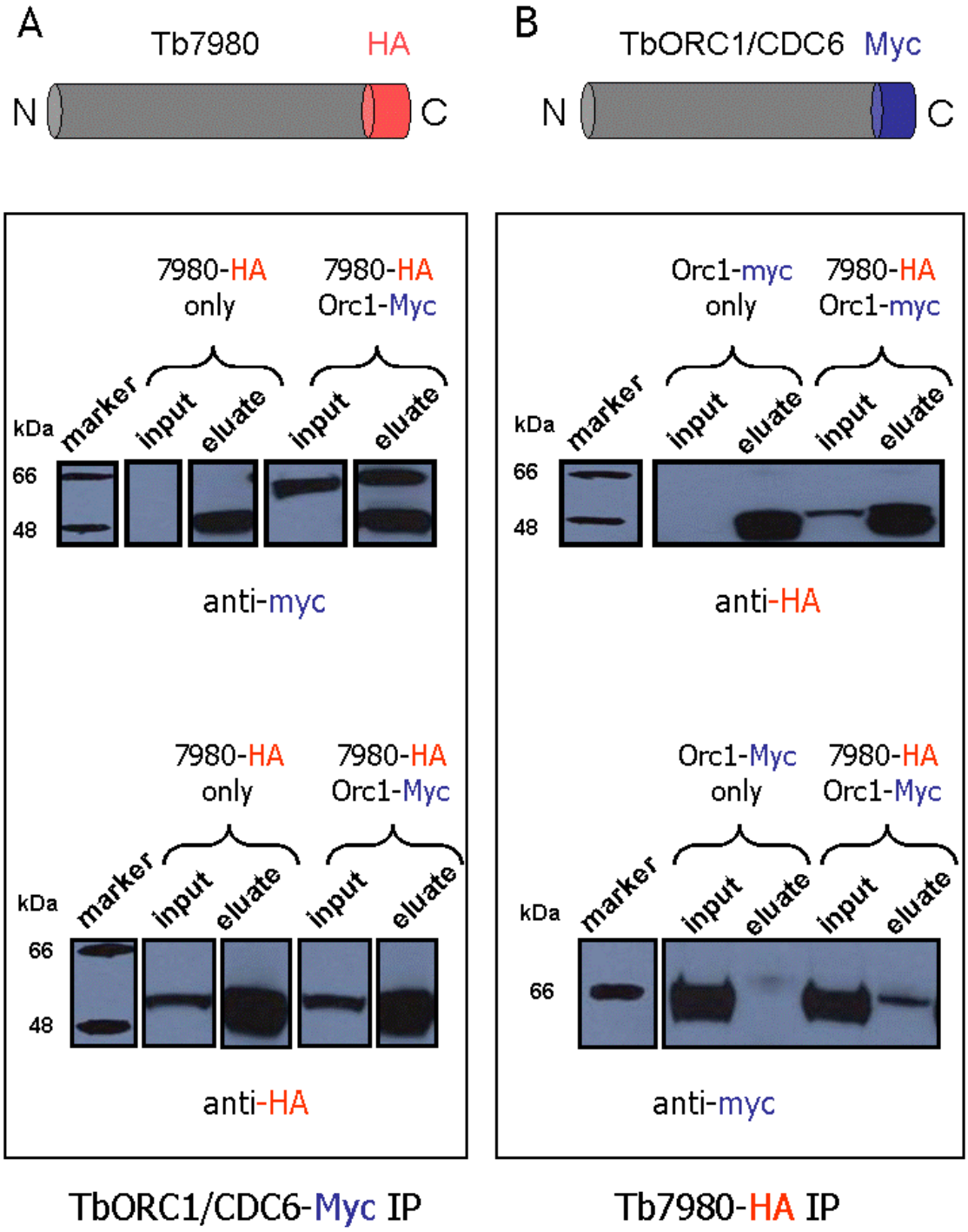


Figure 4-20 – Western blot analysis of the inputs and eluates following co-immunoprecipitation using anti-HA and anti-myc antibodys

(A) and (B) show cartoon of epitope-tagged constructs of Tb7980 and TbORC1/CDC6. Whole cell extracts from cell lines expressing proteins from both constructs were subjected to co-IP using antibodies to the Myc or HA tags. “TbORC1/CDC6-Myc IP Box”: The input and eluate from co-IPs from whole cell extracts of a control cell line (7980-HA, expressing Tb7980-HA), and from co-expressor cell lines expressing Tb7980-HA (7980-HA), and TbORC1/CDC6-Myc (Orc1-Myc), were separated on a 10 % SDS-PAGE gel and then transferred to a nylon membrane. This was then probed with anti-Myc antibody (anti-myc) or with anti-HA antibody (anti-HA). “Tb7980-HA IP Box”: The input and eluate from co-IPs from whole cell extracts of a control cell line (Orc1-myc, expressing TbORC1/CDC6-Myc), and from co-expressor cell lines expressing Tb7980-HA (7980-HA), and TbORC1/CDC6-Myc (Orc1-Myc), were separated on a 10 % SDS-PAGE gel and then transferred to a nylon membrane. This was then probed with anti-HA antibody (anti-HA) or with anti-Myc antibody (anti-myc).

4.9.2 Amino acid sequence analysis of Tb7980

As earlier mentioned, Tb7980 was identified as one of the potential interacting molecular partners of TbORC1/CDC6, including Tb13380. In the preceding sections, data has been shown to suggest that indeed the *T. brucei* machinery may have additional divergent factors; Tb13380. We therefore carried out amino acid sequence analysis of the putative second factor, Tb7980, to specifically ask if obvious ORC motifs could be identified that would justify further biochemical characterisation for inclusion into the *T. brucei* ORC-related family.

Tb7980 is annotated as a hypothetical protein in TritryDB (http://tritrypdb.org). The protein is predicted to be 441 amino acids long and 48.8 kDa in size. Tb7980 is conserved and syntenic across the sequenced kinetoplastids: orthologues exist in all the Tritryps, are annotated as hypothetical proteins and are of a similar size. *T. cruzi* encodes a single protein (Tc00.1047053506247.280) that is 66% identical to Tb7980, while *L. major* encodes a protein (LmjF36.6700) with 44% identity to Tb13380. Second iteration psiBLAST searches of the NCBI database using the Tb7980 sequence from TriTrypDB as query retrieved no meaningful hits with E-value better than the threshold set by the search algorithm. However, third iteration BLAST retrieved a number of DNA helicases from bacteria with the best hit having an E-value of 7e-25. A rather less convincing, but perhaps meaningful, hit was Replication factor C with an E-value of 0.023. In any case the output of the BLAST suggested a probable DNA replication function for the Tb7980, prompting further bioinformatic amino acid analyses.

Examination of the Tb7980 amino acid sequence revealed the presence of a walker A motif (GxxGxGKT; where x = any amino acid residue), which putatively binds nucleotide. Tb7980 has this motif between positions 33 and 40 (inclusive) of its primary amino acid sequence. This motif is known to be present in energy generating proteins that hydrolyse ATP or GTP (Neuwald *et al*. 1999). This observation is consistent with the fact that ORC members (Orc1, Orc4, and Orc5 in budding yeasts) possess this key conserved ATP binding element that is crucial for regulating their function (Speck *et al*. 2005). To ask if the functional nucleotide binding motif present in the primary sequence could be related to the function of the protein, we searched the PHYRE database for related protein

structures. To our surprise, the PHYRE search revealed three predictions of Tb7980 being structurally similar to Cdc6p from the archaeum *Pyrobaculum aerophilum* (85% precision), archaeal Orc1/Cdc6 from *Solfolobus Solfataricus* (80% precision) and archaeal Orc1/Cdc6 *Aeropyrum pernix;* lending further support for Tb7980 being a *T. brucei* ORC/CDC6-related factor. To further support this finding, CLC Workbench v4 PFAM domain searches also revealed that Tb7980 belonged to the AAA^+ superfamily of proteins (Figure 4-21), a critical feature of the ORC family (Orc1, Orc4 and Orc5) of proteins, except for Orc6 which lacks the domain (Erzberger & Berger 2006; Speck *et al.* 2005; Neuwald *et al.* 1999). As shown in Figure 4-21, *S. cerevisiae* Orc1, Orc5 and Cdc6 proteins were also predicted to have the AAA^+ motif in which this function has already been biochemically characterised and reported (Randell *et al.* 2006; Speck *et al.* 2005; Harvey & Newport 2003). These observations put together may support a role for Tb7980 in the formation of an ORC complex formation in *T. brucei*.

Although our co-IP experiments were inconclusive, the bioinformatic analyses prompted to us to further investigate the role of this protein by RNAi of the gene in BSF cells.

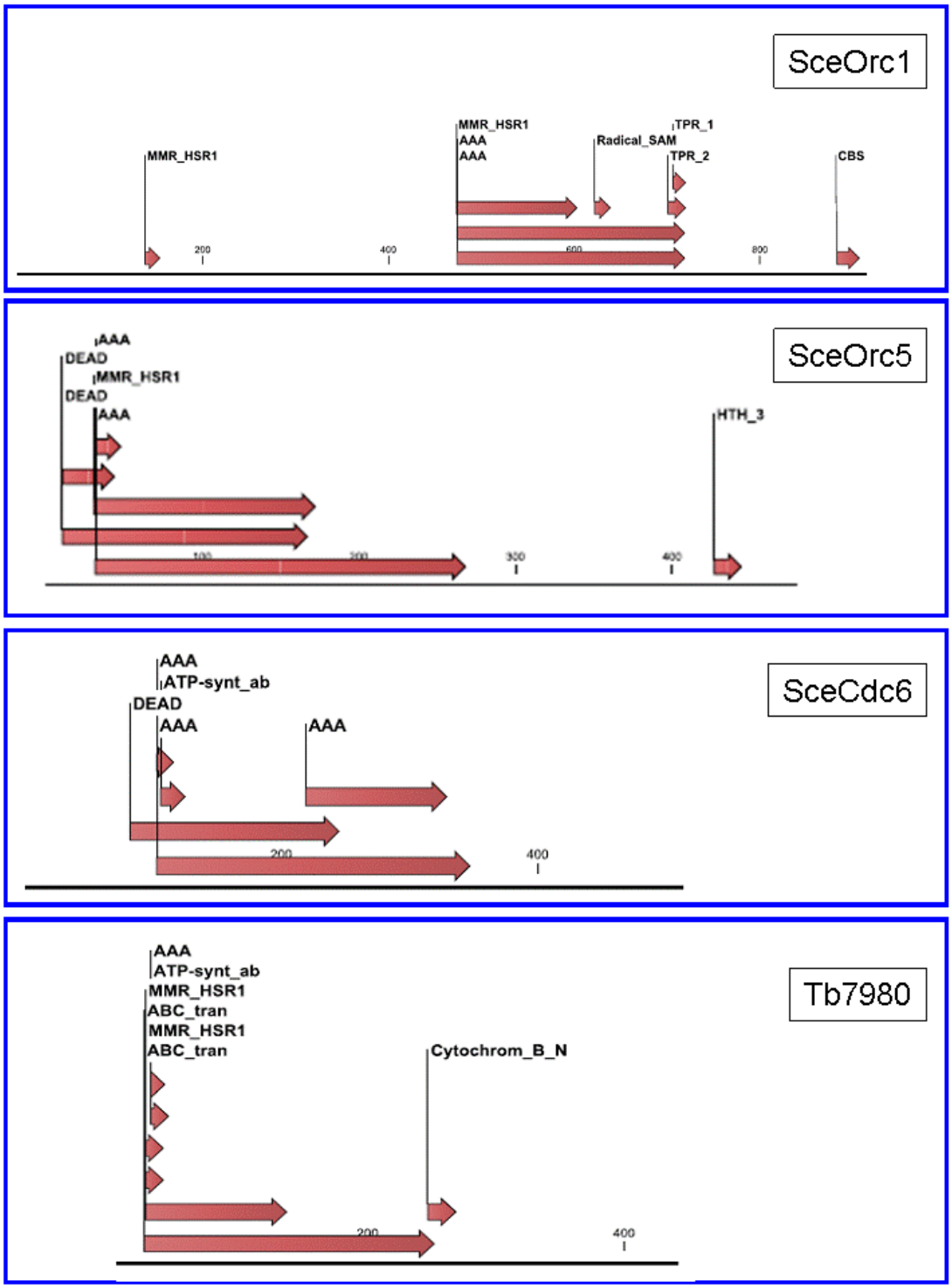


Figure 4-21 – PFAM domain prediction of Tb13380 compared with selected *S. cerevisiae* Orc and Cdc6 proteins

Using the "PFAM 100 most common" domain prediction tool in CLC Genomics Workbench v4, putative functional domain organisation of Tb7980 (Tb798080 box) is shown relative to the same analysis of *S. cerevisiae* Orc1 (SceOrc1 box), *S. cerevisiae* Orc5 (SceOrc5 box) and *S. cerevisiae* Cdc6 (SceCdc6 box). An ATPase motif is indicated by AAA (PFAM no. PF00004); tetratricopeptide repeats are indicated by TPR1 and TPR2 (PFAM no. PF07719); Radical SAM superfamily is indicated by Radical SAM which confer protein-protein interactions and multi-complex formation function; ABC transporters indicated by ABC_tran (PFAM no. PF00005); GTPase motif indicated by MMR_HRS1 (PFAM no. PF01926); cytochrome b (N-terminal) motif indicated by Cytochrom_B_N (PFAM no. PF00033); Helix-turn-helix motif indicated by HTH_3 (PFAM no. PF01381); CBS domain indicated by CBS (PFAM no. PF00571) DEAD/DEAH box helicase indicated by DEAD (PFAM no. PF00270); Zinc finger, C3HC4 type indicated by zf-C3HC4 (PFAM no. PF00097); PFAM numbers (PFAM no.) obtained from the Sanger Institute webpage available at (http://pfam.sanger.ac.uk/).

4.9.3 RNA interference analysis of Tb7980 in bloodstream form cells

4.9.3.1 Cloning of Tb7980 RNAi construct

In order to further investigate the role of Tb7980 *in vivo*, an RNAi approach was used. A 446 bp fragment (Figure 4-22) of *Tb7980* was amplified by PCR using oligonucleotides CT_OL67 and CT_OL68 (Table 2-2), cloned into *Bgl*II/*Hin*dIII of the pZJM dual T7 vector as described in 4.8.3.1 and transfected into the *T. brucei* BSF 427 pLew90-pLew13 cell line. Two independent clones were selected, and analysed further for the effect of inducing *Tb7980* RNAi.

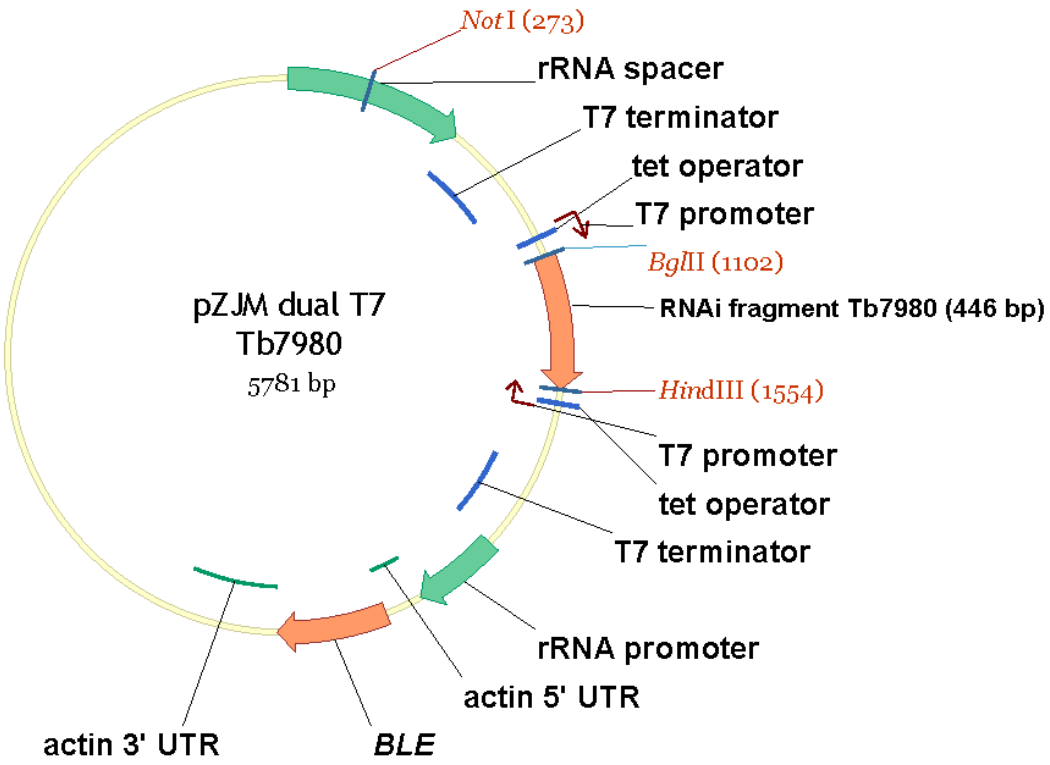


Figure 4-22 – The pZJM dual T7 vector for *Tb7980* RNAi

A fragment of *Tb7980* (446 bp) that was cloned into *Bgl*II and *Hin*dIII sites in the plasmid, after PCR amplification, is shown; Vector properties are same as described previously (Figure 4-15)

4.9.3.2 Effect of TbORC6 RNAi on growth in bloodstream form cells

Growth curves were generated in the absence or presence of tetracycline, which induces expression of the *Tb7980* double-stranded RNA. Induction of *Tb7980*

RNAi resulted in a growth defect in BSF cells observable from 18 h post induction of RNAi in both clones (Figure 4-23), leading to cell death from 30 hours as evidenced by the decreasing cell numbers. These results suggested that *Tb7980* may be essential for viability of BSF *T. brucei* cells.

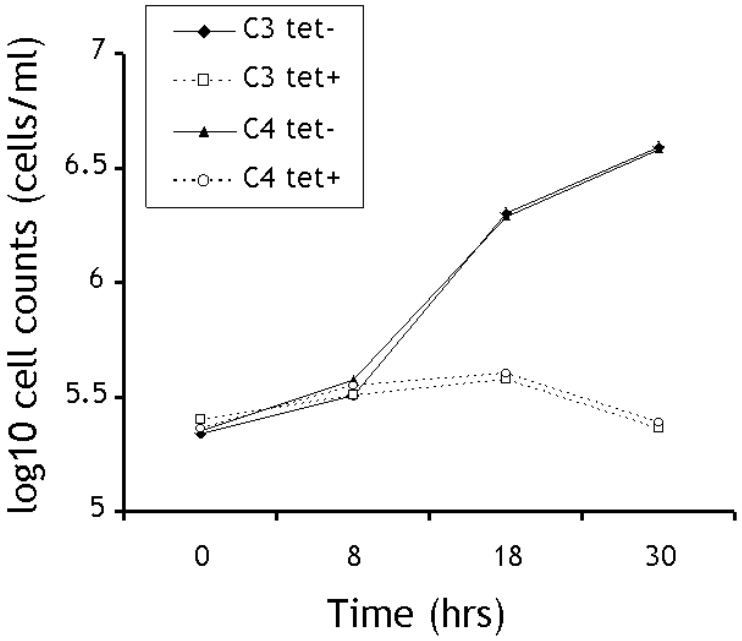


Figure 4-23 – Effect of *Tb798080* RNAi in bloodstream form cells

Growth was monitored in bloodstream cell lines in which *Tb13380* RNAi was induced (+ Tet, solid line) or without induction (-Tet, broken line) for 2 clones; C3 and C4

4.9.3.3 Effect of *Tb7980* RNAi on cell morphology in bloodstream form cells

To determine whether the growth defect observed upon *Tb7980* RNAi occurred as a result of a cell cycle defect, DNA content was monitored as described in Section 4.8.3.3. The results for a single clone after DNA-DAPI staining and microscopically counting were very similar to that described for TbORC1/CDC6 and Tb13380 depletion by RNAi (Sections 3.6and 4.8.3, respectively). As shown in Figure 4-24, after RNAi depletion of Tb7980, DAPI staining revealed an increase in cells with >2 nuclei and >2 kinetoplast DNA content (>2N2K cells, unclassified) from ~ 2% (at 8 hrs post-induction) to 60% (at 24 hrs post induction) of the population, with a concomitant decrease in 1N1K cells from ~ 70% to ~ 20% within the same time period. This cell morphological defect was not observed in the control population (uninduced population) where 1N1K cells

remained constant constituting 80% of the population between 8 hrs and 24 hrs during which the experiment was performed.

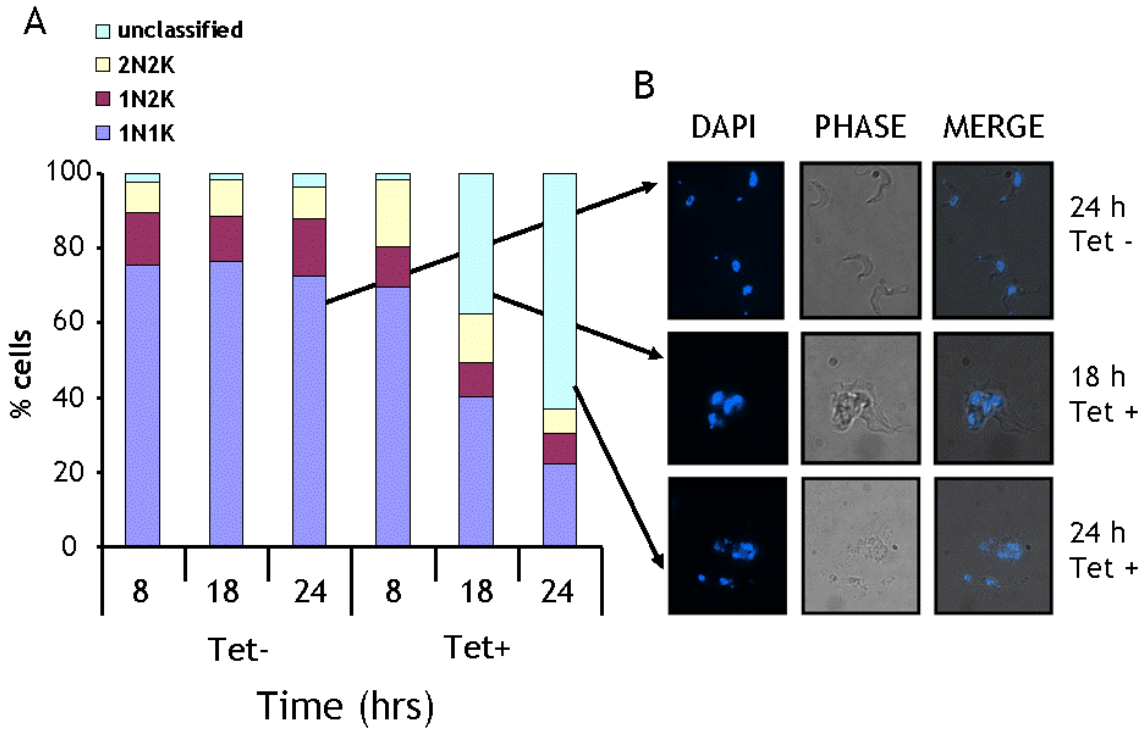


Figure 4-24 – Analysis of nuclear and kinetoplast DNA configuration of *Tb7980* RNAi cells for a single clone

(A) Microscopic of counting nuclear (N) and kinetoplast (K) DNA content after DAPI staining. ~ 100 bloodstream form *T. brucei* cells were counted at each time point for non RNAi-induced control cells (A: 8, 18, 24 hrs; Tet -), or cells depleted in Tb7980 by RNAi (A: 8, 18, 26 hrs; Tet +). (B) Images of normal cells (Tet –, 24 hrs) and abnormal cells after RNAi induction (Tet +; 18 and 24 hrs). DNA is shown by DAPI stain (DAPI); a phase contrast image (PHASE), and an overlay of the DAPI and phase images (MERGE).

4.9.3.4 Effect of *Tb7980* RNAi on cell cycle

Analysis of flow cytometry profiles of the *Tb7980* RNAi-depleted cells showed that the cells re-replicated their nuclei and kinetoplasts, as demonstrated by the appearance of 8C peaks with the concomitant disappearance of 2C and 4C peaks following induction at 18 and 24 hrs with the uninduced RNAi cells remaining unchanged for the same period of time (Figure 4-25). This observation is in agreement with the accumulation of multinucleate cells observed from DNA DAPI staining experiments.

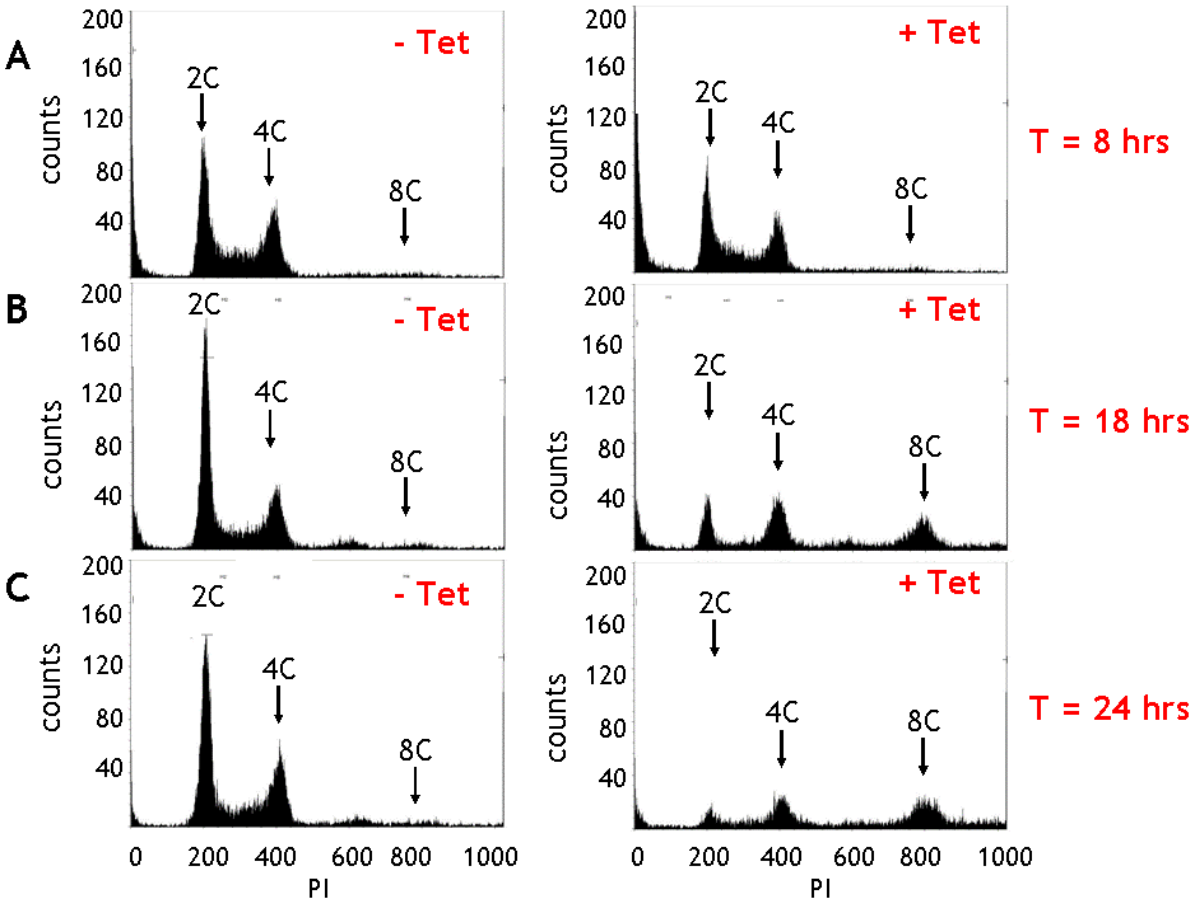


Figure 4-25 – FACS profiles of PI-stained cells after *Tb7980* RNAi

Histograms show propidium iodide-stained *T. brucei* procyclic-form cells after FACS sorting, sampled 8 hrs (A), 18 hrs (B), 26 hrs (C) pre- and post - induction of RNAi for *Tb7980* (- Tet and + Tet ,respectively). The peaks corresponding with cells containing 2C and 4C DNA content are indicated, as is the peak position for cells with 8C content; where C represents a haploid DNA content

While analysis to quantify the RNAi-induced knockdown of *Tb7980* transcripts or protein levels remains to be carried, it is remarkable that in the RNAi analysis of *Tb7980,* the timing of the growth arrest and subsequent cell death, as well as the nature of the aberrant cells that arise, is highly reminiscent of the phenotypes that result from *TbORC1/CDC6* and *Tb13380* RNAi in BSF *T. brucei* cells.

4.10 Discussion

The aim of this chapter was to investigate if TbORC1/CDC6 is the sole functional ORC-related component of the *T. brucei* pre-replication machinery. To do this, targeted IP analysis to test for interactions between known factors, as well as a proteomics approach after IP of TbORC1/CDC6, were adopted. These

experiments required the generation of tagged variants of the proteins under investigation. Ideally, specific antibodies raised and validated against native proteins would be employed; hence the tools used here will be justified. An overview of the proteomic analysis and subsequent bioinformatic approaches used will also be discussed within the context of *T. brucei* replication initiation machinery. To conclude this chapter, our existing hypothesis for what constitutes the *T. brucei* pre-RC machinery will be discussed in relation to the novel candidates identified.

4.10.1 Are in silico approaches for identifying molecular interactors sufficient; a case study of the T. brucei DNA replication initiation machinery

Unprecedented advances in DNA sequencing technologies over the last decade and the increasing availability of whole genome sequences for various organisms has progressively amplified the pace of basic scientific research; reviewed in (Metzker 2010). The revolution in DNA sequencing technologies has also triggered a steady rise in the development of software and web-based tools that facilitate comparative relationships between biological pathways, and even between whole organisms; for review see (Chen, Jorgenson, & Cheung 2009). Various Bioinformatic tools that help for whole genome or specific comparative studies, some freely available and some commercial, are increasingly available to the scientific community for studying patterns in evolution of genes, for ascribing functions to genes, and for understanding interaction pathways (Chen *et al*. 2009). Inference from the presence or absence of a specific catalogue of genes has been used to predict what biological pathways are active or inactive in an organism for which experimental information is not available. An example of this has been observed in the microsporidium *E. cuniculi*, which has been reported to have lost a number of DNA repair genes (Gill & Fast 2007). Nevertheless, to extrapolate function, or absence of function, using homology from other organisms might sometimes be misleading due to biological variation. Thus, experimental analysis still remains vital to substantiate computer-based predictive strategies.

With the publication of the genome sequence for *T. brucei* and related kinetoplastids in 2005, it was possible to use a number of these computer-based algorithms to search the genome database to predict the function of *T. brucei* proteins based on primary amino acid sequence and overall 3D structural homology with those of other organisms. Overall, 9068 ORFs were predicted, and those genes (~ 64%) for which obvious homology could not be ascribed to the products were annotated as hypothetical proteins (Berriman *et al.* 2005). Nuclear DNA replication initiation in *T. brucei* provides a good example, as this was a poorly studied process with essentially no functional data available prior to the release of the whole genome sequence. The TritryDB genome annotation indicates that *Trypanosoma* and *Leishmania* species possess a simplified ORC machinery that is analogous to archaea (Aslett *et al.* 2010). This hypothesis was supported by a recent paper from the lab of M.C. Elias (Godoy *et al, 2009)*. Indeed, our own bioinformatic searches identified only one subunit of ORC, and found that and other replication initiation proteins, including Cdt1 and CDC7-DBF4, were 'missing' in the these parasites. Given the potential evolutionary and therapeutic consequences of such putative novelty, we considered it important to experimentally test the bioinformatic hypothesis that the *T. brucei* ORC machinery is structurally and functionally distinct from its mammalian host. This led us to generate a number of genetic tools, as discussed below.

4.10.2 Is TbORC1/CDC6 the helicase (TbMCM) loader in T. brucei?

One fundamental question that was asked in this chapter was whether the TbMCM helicase is recruited onto origin DNA in *T. brucei* by TbORC1/CDC6. With no obvious homologue of Cdt1, a verified eukaryotic helicase loader (Chen *et al.* 2007; Randell *et al.* 2006; Tanaka & Diffley 2002), or archaeal WhiP, which might function as a helicase loader (Robinson & Bell 2007), present in the genome of *T. brucei,* it seemed possible that a direct interaction between TbORC1/CDC6 and TbMCM might mediate this crucial step in replication. Any such interaction would be highly novel in eukaryotes, and therefore of mechanistic interest. However, if such interaction was detectable further assays (such as *in vitro* reconstitution) would ultimately be needed to corroborate their functionality, since the approach we took only asked if a stable interaction was detectable between

TbORC1/CDC6 and TbMCM, and could not exclude an indirect interaction via additional factors. Nevertheless, stable interaction between ORC and MCM has not previously been described in unmodified eukaryotes; except when Cdt1 is tethered to Orc1-5 (Chen *et al.* 2007).

Studies in bacteria, archaea and eukaryotes (see Section 1.6) suggest that the presence of a loading factor is normally crucial for recruitment of helicase onto origins, and only in some archaea is this role provided by Orc1/Cdc6 itself (Kasiviswanathan *et al.* 2005; De Felice *et al.* 2003). In the analysis described here, we did not find evidence for ORC1/CDC6 binding to the MCM complex in *T. brucei*. We tested for *in vivo* interaction between TbORC1/CDC6 and TbMCM3, TbMCM6 and TbMCM7, without success. Tests of interaction between TbORC1/CDC6 and TbMCM5 or TbMCM4 were not carried out. However, we showed also that TbMCM2, TbMCM4, TbMCM6 and TbMCM7 can form a complex (see below), so it seems likely that if TbORC1/CDC6 and TbMCM4 were to interact, this would have revealed interaction with these other subunits. In *S. pombe* Mcm3 forms a dimer with Mcm5 (Lee & Hurwitz 2000). If this is conserved also in *T. brucei*, the lack of TbORC1/CDC6 interaction with TbMCM3 (albeit only tested via IP of TbMCM3-HA) is likely also to rule out interaction with TbMCM5. We therefore tentatively conclude that *T. brucei* recruitment of the replicative MCM helicase is not mediated through direct interaction with TbORC1/CDC6. However, in the absence of TbMCM5-HA tagged cells, or examination of TbORC1/CDC6-TbMCM4 interaction, we cannot formally exclude that these subunits mediate interaction. A recent paper has shown that in *S. cerevisiae,* absence of Cdt1 results in reduced levels of Mcm3 and Mcm5, suggesting that the helicase loader stabilises these subunits of the MCM complex (Tsakraklides & Bell 2010). We do not see evidence for low abundance of TbMCM3 in *T. brucei*, suggesting that any evolutionary absence of Cdt1 does not result in constitutive loss of this subunit, though again TbMCM5 has not been tested. If TbMCM5 were to be of low abundance, and was critical in direct recruitment of TbMCM by TbORC1/CDC6, small amounts of the TbMCM subunits examined following IP may have been missed. Equally, the analysis has examined interactions in whole cell extracts, and it is possible that functional, direct interaction between TbORC1/CDC6 and TbMCM is only found in the nucleus, and hence our assay may not be sensitive enough to detect this.

4.10.3 MCM helicase subunits in T. brucei form typical eukaryotic MCM subcomplexes

Bioinformatic mining of the kinetoplastid genomes shows that they possess all six eukaryotic MCM paralogues. The *in vivo* IP experiments described here show that the *T. brucei* MCM proteins are able to form subcomplexes that appear typical of the eukaryotic model described in Section 1.6.3.4. Briefly, in eukaryotes a subcomplex of Mcm4/6/7 has been shown to have 3′-5′ helicase activity *in vitro* while addition of Mcm2, Mcm3 and Mcm5 tend to play a regulatory role within the hexameric complex (Lee & Hurwitz 2000; Ishimi 1997). From our MCM IPs, the presence of TbMCM2/4/6/7 subcomplex in *T. brucei* was demonstrated. In contrast to the co-IP of TbMCM2, TbMCM4, TbMCM6 and TbMCM7, IP of TbMCM3 did not co-IP any other MCM subunit. This is consistent with the expectation from other eukaryotes that TbMCM3 does not always interact with a TbMCM2/4/6/7 complex, but is inconsistent with the protein forming a stable dimer with TbMCM5. This may indicate a difference from the eukaryotic model, but it seems unlikely given the conservation of TbMCM5 in all kinetoplastids and its sequence homology with *S. cerevisiae* Mcm5 (42 % identity between *S. cerevisiae* Mcm5 and TbMCM5 from *Leishmania* and *Trypanosomes*).

Immuno-depletion of Cdt1 from budding yeast extracts results in reduced purification yields of Mcm3 and Mcm5, while co-overexpression of Cdt1 improves the purification yield of Mcm3 and Mcm5 (Tsakraklides & Bell 2010). These results have been interpreted as suggesting that Cdt1 may stabilise Mcm3 and Mcm5 in the hexameric complex. The absence of an interaction between TbMCM3 and TbMCM5 may therefore be due to limitations in our purification protocol to retain a *T. brucei* Cdt1-like factor during the IP (assuming *T. brucei* has a homologous protein that has been missed in our *in silico* searches).

Despite the absence of evidence for a TbMCM3/5 subcomplex, the observation of a TbMCM2/4/6/7 complex argues that assembly of a putative *T. brucei* MCM hexameric complex, involving all six protein subunits, follows the eukaryotic paradigm (Figure 4-26)

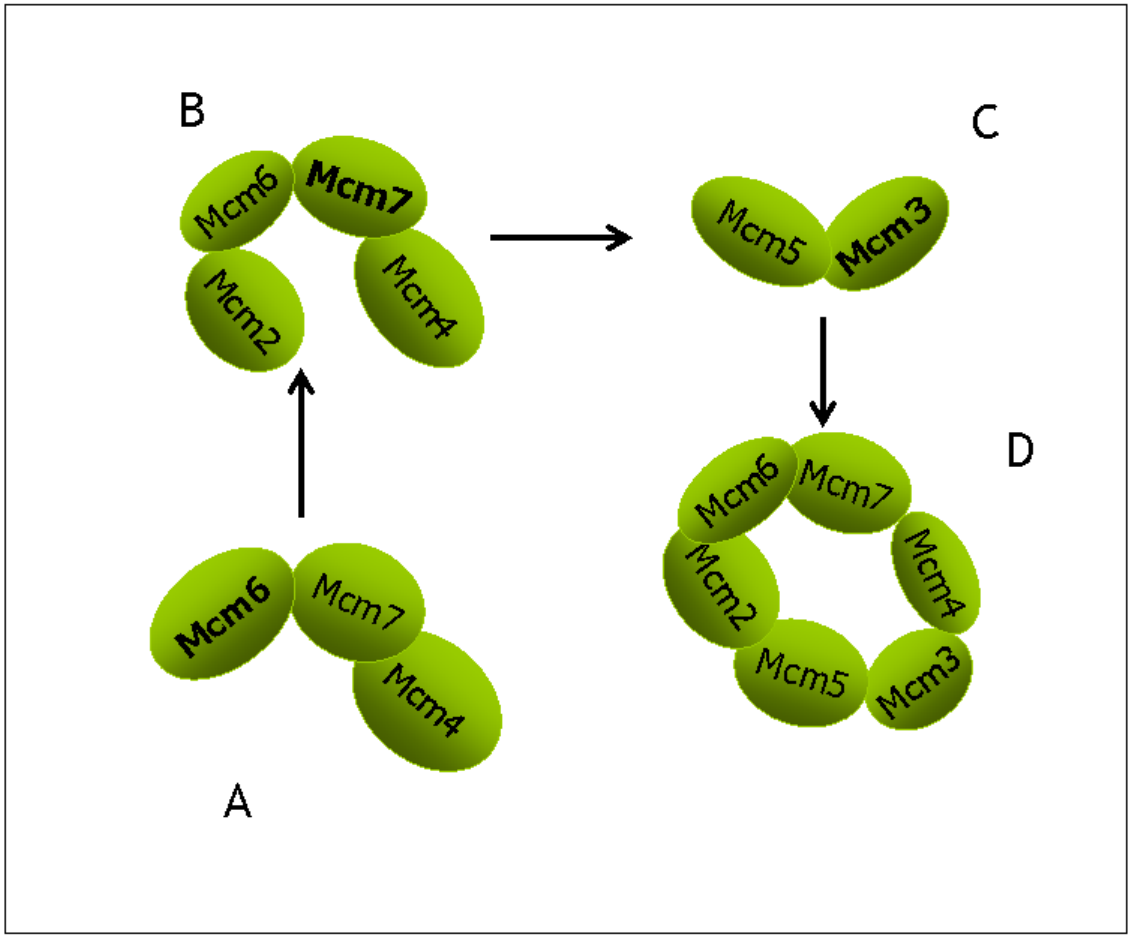


Figure 4-26 –Typical eukaryotic subcomplexes *in vivo* for human MCM proteins

(A) trimer of human Mcm4/6/7; (B) tetramer of human Mcm2/4/6/7; (C) dimer of human Mcm3/5; Arrows indicate direction of interaction of subcomplexes for formation of a heterohexamer. Figure adapted from (Yu, Feng, & Liang 2004)

4.10.4 The T. brucei DNA replication initiation machinery: complex or simplified?

From the above data, we have proposed that 'downstream' events in pre-RC assembly in *T. brucei* are more eukaryotic than archaeal: we confirm bioinformatic predictions that the MCM helicase exists as a complex of distinct subunits, and we can find no evidence that this is recruited to origins directly by TbORC1/CDC6. With this in mind, below is a discussion of our comparative eukaryotic genome mining for the pre-RC machinery in relation to how this aids our understanding of *T. brucei* replication initiation, followed by analysis of the novel proteins obtained from IP studies with TbORC1/CDC6.

DNA replication is an intricate biological process that involves a series of complex reactions tightly regulated at various levels. In humans, from start to finish 165 proteins are currently known to be involved DNA replication (Cotterill & Kearsey 2009), either acting individually or acting in (sub)complexes. The association of ORC, Cdc6, Cdt1, and the MCM complex to form the pre-RC at replication origins has been observed in all characterised eukaryotes. In budding yeast, for example, *in vitro* pre-RC assembly has been shown to be a dynamic and sequential process where there is ordered loading of ORC and Cdc6, then upon recruitment of an Mcm2-7-Cdt1 heptameric complex, there is a concomitant offloading of ORC from the pre-RC (Tsakraklides & Bell 2010). These results suggest that pre-RC formation is a tightly regulated process and involves the concerted participation of all of these factors. What consequences might stem from the putative absence of several of these factors in some eukaryotes? As shown in Figure 4-1, only some subunits of the ORC family are detected in some eukaryotes and in a few (including *T. brucei*) only a single protein is found that is similar to Cdc6 (Orc1/Cdc6). Orc6 and Orc3 appear to be the ORC subunits that are most frequently undetected, being absent from genome searches of nine of the organisms examined, including representatives of each supergroup (Adl et al, 2005) except the Archaeplastida. Orc2 and Orc4, and to a lesser extent Orc5, were more widely detected, either individually or in combination. Orc1 was the sole subunit that was universally detected (albeit in four cases as a potential 'hybrid' with Cdc6). This distribution may not reflect presence or absence, but instead the extent to which the sequences of the proteins are functionally constrained in evolution. For instance, it may be that Orc3 and Orc6 are much diverged in sequence amongst eukaryotes, and therefore escape bioinformatic searches using experimentally characterised members of the family. Alternatively, some or all of the subunits may truly be absent, reflecting later evolutionary formation in some eukaryotes, or loss in many.

There is little functional data to assess if eukaryotic DNA replication could initiate if some ORC components were absent (though, clearly, this is possible in archaea). Orc6 has been shown to be essential for cell viability in *S. cerevisiae* and to be involved in the recruitment of Cdt1 (Chen *et al.* 2007), but some evidence suggests it could be dispensable in some circumstances. This is because

it has been shown not to be required for ORC-DNA binding (Chen *et al.* 2007; Semple *et al.* 2006; Lee & Bell 1997). The more widespread detection of Orc1, Orc2, Orc4 and Orc5 subunits could be explained by the fact that they have each been proposed to possess a winged helix domain fold that acts in binding DNA (Speck *et al.* 2005). Moreover, each (except Orc2) are members of the AAA^+ family of proteins with ATPase activities (Speck *et al.* 2005), which provides a key regulatory component in the complex and in interactions with other factors. However, the same motifs are conserved in Orc3, so why then this might be less conserved is unclear, though it may indicate that there is some functional redundancy within the 'standard' six ORC subunits (whose detailed individual roles are not yet understood; (Duncker *et al.* 2009). As stated in the results, the absence of some detectable ORC subunits in *E. cuniculi*, given its relatively close relationship with fungi (Gill & Fast 2007; Katinka *et al.* 2001), may be the clearest sign that eukaryotic ORC architecture can diverge from the six-subunit paradigm. Ultimately, such questions can only be resolved by biochemical structural analysis. Below we described the first attempt to do so in a protozoan organism.

4.10.5 Tb13380 is a bona fide member of the ORC family

To attempt to ask what proteins interact with TbORC1/CDC6 we adopted the strategy of TbORC1/CDC6-Myc IP and identification of interacting factors by separation of recovered proteins by gel electrophoresis followed by identification by mass spectrometry. As a relative quantitative approach to give fold-enrichment would have required optimisation potentially beyond the time-frame of the project, we utilised this rapid 'one-step' protein purification approach. A drawback of this technique is that we lost the ability to discriminate small levels of enrichment, and/or weak and transient interactors. Another drawback is that it is likely that some proteins will bind to the anti-Myc coupled beads, rather than to TbORC1/CDC6. By comparing a control WT-IP to the TbORC1/CDC6-Myc IP, we were able to filter out these contaminants. Since the same amount of beads was used in both the TbORC1/CDC6-Myc-IP and the WT-IP, this appeared to be a reliable approach because the same amount of bead surface area is exposed to protein extract during experimentation. Nevertheless, it is unlikely that all the ~150 proteins that were found specifically in a

TbORC1/CDC6 - IP were real interactors; it is necessary to confirm interactions by direct co-IP experiments.

Three putative TbORC1/CDC6 interacting proteins stood out from the above analysis as potential ORC-related factors, based on sequence analysis: Tb927.10.13380 (or Tb13380), Tb927.10.7980 (or Tb7980) and Tb09.160.3120. Functional characterisation of Tb13380 and Tb7980 by RNAi showed highly similar phenotypes to those observed following RNAi of TbORC1/CDC6 (chapter 3), consistent with these proteins acting in the same pathway. Co-IP of Tb13380 and TbORC1/CDC6 confirmed that the proteins are *bona fide* interactors. The same experiments for TbORC1/CDC6 and Tb7980 were inconclusive, though do not rule out interaction. Combining the bioinformatic and experimental data, we speculate that at least Tb13380 and TbORC1/CDC6 act together in the initiation of DNA replication, and thus Tb13380 is a member of the pre-RC complex in *T. brucei*.

All ORC protein family members except Orc6 belong to the AAA^+ ATPase family (Chen *et al*. 2007), characterised by the presence of Walker A and Walker B motifs, and associated sensor 1 and 2 motifs, that are necessary for binding and hydrolysing, typically, ATP to ADP (Speck *et al*. 2005). The Walker A motif (consensus, GxxxxGK[T/S]) binds ATP/GTP (Kawakami & Katayama 2010), while the Walker B box (consensus, hhhh[D/E]) binds Mg++. Although Tb13380 appears to lack canonical Walker A and Walker B motifs, structural prediction algorithms predict it to have AAA^+ folds, suggesting it belongs to the AAA^+ ATPase family. We therefore speculate that it is still likely to function as a TbORC family member. This assertion is based on similar patterns observed for budding yeasts Orc2 and Orc3 proteins. Both proteins may also not be functional ATPases, but possess the AAA^+ folds characteristic of this protein family (Clarey *et al*. 2006; Speck *et al*. 2005). Therefore, in *T. brucei,* ATPase regulation of the pre-RC components is likely to focus on TbORC1/CDC6, which possesses clearly conserved ATPase domains. It is possible also that ATPase activity, which is likely to be used for ORC conformational changes to mediate association and disassociation with other pre-RC proteins and with origins (Speck *et al*. 2005), might reside in other ORC factors. One possibility is Tb7980. Though we cannot confirm by co-IP that this protein interacts with TbORC1/CDC6, RNAi gives very

similar phenotypes and the protein has AAA^+ fold homology, though alignment of the individual motifs and assigning ORC subunit orthology has proved elusive.

Based on the evidence presented above, we conclude that the pre-RC machinery contains at least one further factor beyond TbORC1/CDC6, encoded by Tb10.389.0050 (that we have named Tb13380) and we propose to name TbORC4, based on potential orthology with Orc4 in other eukaryotes (Figure 4-27). As yet, definitive tests are needed to ask if Tb7980 can also be considered a *T. brucei* pre-RC, and potential ORC, component. Indeed, further experiments to test this hypothesis, by detailing the nature of the putative interactions amongst these factors and their roles during DNA replication, are needed.

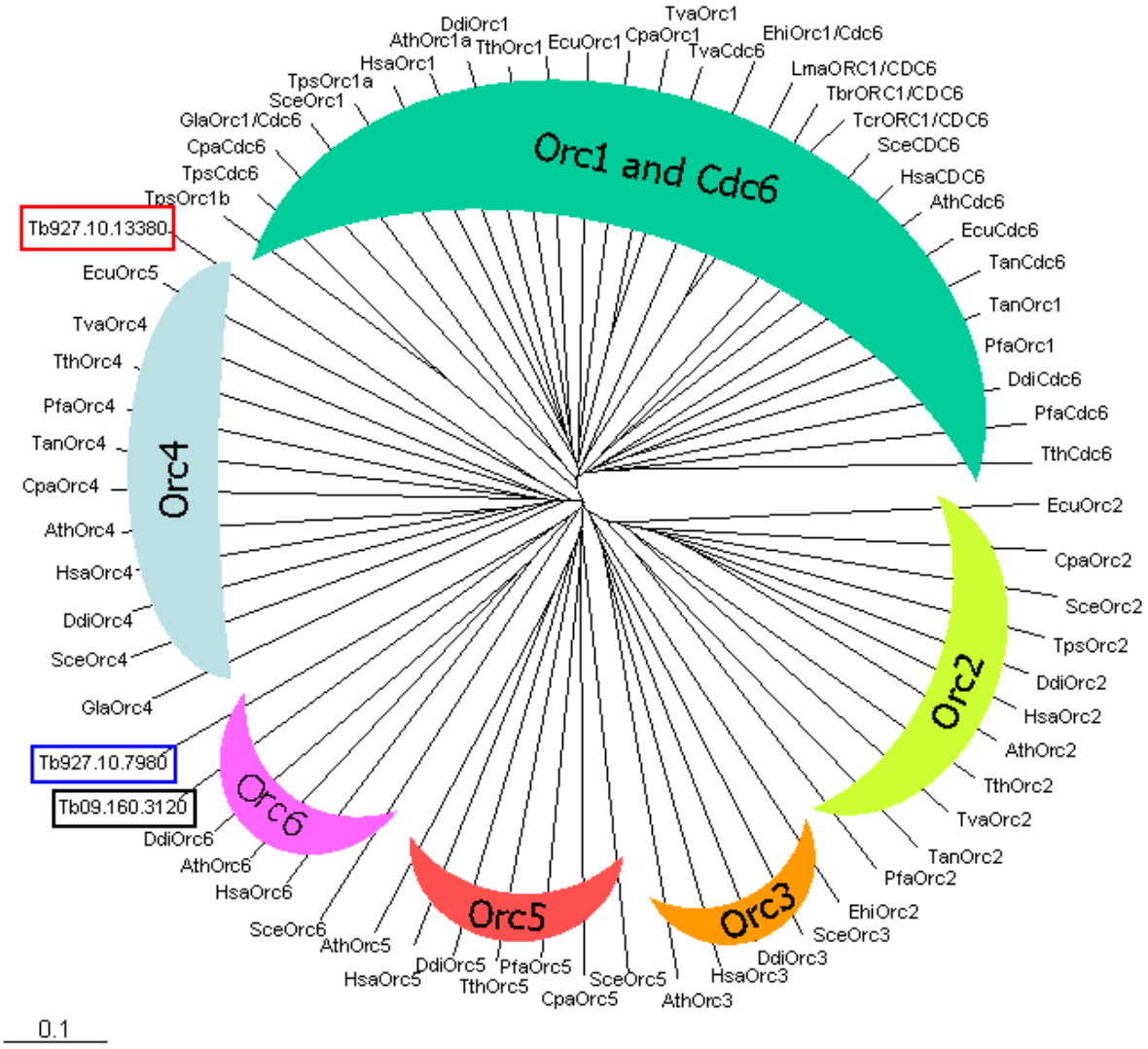


Figure 4-27 – Phylogenetic tree of eukaryotic ORC proteins and putative *T. brucei* ORC proteins

A neighbour-joining phylogenetic tree was generated from the ClustalX alignment in Figure 4-13 (with exception of DmeOrc4); each line refers to the length of the tree arms and indicates 10 amino acid changes per 100 amino acids. Hsa, Homo sapiens; Dme, *Drosophila melanogaster*; Ath, Arabidopsis thaliana; Sce, Saccharomyces cerevisiae; Pfa, *Plasmodium falciparum*; Ddi, *Dictyostelium discoideum*; Cpa, *Cryptosporidium parvum*; Tan, *Theileria annulata*;; Tth, *Tetrahymena thermophila*; Gla, *Giardia lamblia*; Ecu, and *Encephalitozoon cuniculi* ; Tva, *Trichomonas vaginalis;* and Tb, *Trypanosoma brucei.* Genbank accession numbers are provided Appendix 3

4.10.6 How is TbORC1/CDC6 recruited to origins in T. brucei? A hypothetical mechanism

This section considers the specific question of how DNA sites along chromosomes might be selected to act as origins in *T. brucei*. In higher eukaryotes, the mechanism by which ORC identifies and selects a functional origin locus and nucleates it for pre-RC licensing has not been clearly elucidated. Several factors

have been proposed to trigger ORC recognition of origins, although some of these ORC binding sites do not eventually function as active origins (Wyrick *et al*. 2001). Some factors that influence ORC binding include: (1) local chromatin architecture, e.g. presence of euchromatin (MacAlpine *et al*. 2010); (2) recruitment via other interacting proteins, e.g. EBNA1 (Norseen *et al*. 2008); (3) local sequence elements, e.g. AT-rich sequences in *S. cerevisiae* (Theis & Newlon 1997); CpG islands in mouse embryonic stem cells (Sequeira-Mendes *et al*. 2009); (4) topological properties of DNA, e.g. the presence of negative supercoils in *Drosophila* (Remus, Beall, & Botchan 2004). A more detailed discussion of the influence of chromatin and local DNA sequence elements will be covered in Chapter 5.

Recently, in eukaryotic cells three independent reports have proposed an RNA-dependent mechanism of ORC recruitment to DNA replication origins or to other sequences, mediated via interactions with either known RNA-binding proteins or interaction with RNA. In the free-living ciliate protozoan *Tetrahymena thermophila,* ORC is targeted to the ribosomal DNA (rDNA) replication origin via its interaction with a non-coding fragment of 26S rRNA known as 26T RNA. The formation of an RNA-DNA hybrid between the 26T RNA and origin DNA is thought to dictate the recruitment of ORC at rDNA origins in these organisms (Donti *et al*. 2009; Mohammad *et al*. 2007). In Epstein-Barr virus, it has been shown that ORC binding to origin DNA of an episome (oriP) requires interaction of ORC to a viral protein, EBNA1, and this interaction is stabilised by a G-rich RNA sequence, suggesting that RNA molecules are involved in the recruitment of ORC in human cells (Norseen *et al*. 2008). In yeast and mammalian cells, the non-coding telomeric repeat-containing RNA (TERRA) has been implicated to facilitate TFR2 recruitment to telomeres via the formation of DNA-RNA hybrids. This interaction mediated by TERRA is also thought to enhance the recruitment of ORC to telomeres, since ORC interacts with TFR2 (Deng *et al*. 2009). Given these findings, could it also be the case that in *T. brucei* TbORC1/CDC6 recruitment to origins is mediated via an RNA-interaction mechanism?

Analysis of the TbORC1/CDC6-Myc IP and mass spectrometry data above concentrated on putative ORC–like factors that interact with TbORC1/CDC6. However, a wider consideration of the proteins recovered, but not so far analysed functionally, may provide preliminary evidence that supports RNA-

dependent recruitment of ORC at replication origins (or to DNA in general) in *T. brucei*. Among the 'best' hits (Section 4.7), *in silico* analyses suggested that four proteins (Tb927.6.1120, Tb11.46.0009, Tb927.2.5130, and Tb927.10.9350) appeared to have an RNA-associated function. Using Phyre, Tb927.6.1120 (14 peptide hits) showed homology, at 70 % precision, to a virion RNA polymerase, Tb11.46.0009 (5 peptide hits) showed 100 % precision homology with ATP-dependent RNA helicases, Tb927.2.5130 (3 peptide hits) showed 90 - 95 % precision homology to Armadillo/HEAT repeat proteins (Andrade *et al*. 2001), and Tb927.10.9350 (3 peptide hits) showed 60 – 70 % precision homology also to Armadillo/HEAT repeat proteins. Whilst this is very preliminary, these bioinformatic predictions could suggest TbORC1/CDC6 interaction with a protein complex that involves, and perhaps generates, RNA. This is clearly highly speculative, but it would be of value to test, thereby asking if RNA mediates TbORC1/CDC6 recruitment to replication origins, or underlies further functions of the protein in *T. brucei*.

4.11 Highlight of major findings

In summary, the findings in this chapter include: (1) *T. brucei* ORC1/CDC6 interacts with at least one other protein (Tb927.10.13380, previously annotated as hypothetical) that bears sequence homology with eukaryotic ORC subunits, and is likely to act in origin definition and DNA replication initiation; (2) RNAi of a further protein (Tb927.10.7980) is consistent with a role in replication, and preliminary evidence suggests this may also interact with TbORC1/CDC6; (3) TbORC1/CDC6 appears not to interact directly with the TbMCM helicase subunits, consistent with previous observations from a number of eukaryotic organisms, and contrary to reports in some archaeal species; (4) MCM subunits in *T. brucei* form at least one subcomplex homologous to that previously observed for human, yeast, *Drosophila, Xenopus* and mouse MCM proteins. Taken together, these data appears to refute the hypothesis that the DNA replication pre-RC machinery in *T. brucei* is analogous to archaea. Rather, we propose a model for TbORC containing least two-components (TbORC1/CDC6 and TbORC4), more analogous to the eukaryotic model, suggesting that origin designation is not carried out by a single protein as suggested by Godoy *et al* (Godoy *et al*.

2009). More work is needed to substantiate these proposals, and it remains unclear what protein or proteins provide a link between TbORC and TbMCM.

5 Genome-wide localisation of TbORC1/CDC6 DNA binding sites in *Trypanosoma brucei*

5.1 Introduction

It is well established that the budding yeast *S. cerevisiae* has a highly conserved consensus sequence element that defines origins of DNA replication (Newlon 1996). Until 2006 only few origins had been characterised in higher eukaryotes and the DNA sequences showed no known consensus (Aladjem *et al.* 2006). Like in humans, origin usage and origin specification on chromosomes in mouse, rat and chicken, and in the more diverged *Drosophila melanogaster* and *Xenopus laevis,* appears to be more plastic [see (Aladjem *et al.* 2006) for a comprehensive review]. This observation has led to several suggestions that DNA replication origin usage in these organisms may be guided by factors that lie beyond primary DNA sequence elements, such as local DNA topology (Remus et al. 2004), local chromatin architecture (MacAlpine *et al.* 2010) and transcriptional regulatory units (Sequeira-Mendes *et al.* 2009), to name just a few. In the past, the approaches taken to identify origins in these organisms have been more direct, targeting specific loci on chromosomes. Some of these methods include: nascent strand abundance for characterisation of origins near the *c-MYC* genes in HeLa cells (Waltz, Trivedi, & Leffak 1996) and the *Adenine Phosphoribosyl Transferase* gene in Chinese Hamster Ovary (CHO) cells (Delgado *et al*. 1998); two-dimensional (2D) gel electrophoresis for characterisation of rDNA loci in *Xenopus* embryos (Hyrien *et al.* 1995); and single molecule analyses for characterisation of rDNA locus in HeLa cells (Lebofsky & Bensimon 2005), among others. From the aforementioned approaches, it became evident that the selection and utilisation of sites to act as origins of DNA replication could not be limited to specific loci as this would not reflect the presence of origins in a global genome context.

Recently, genome-wide approaches have been adopted and these studies have offered new hope for identifying a complete repertoire of DNA replication origins for a particular organism, overcoming the potential for variation due to factors such as local GC or AT content, gene density, gene architecture, and transcriptional activity, which are likely to influence origin usage for specific loci when selected and characterised individually (Gilbert 2010; Aladjem 2007). Some of these approaches are discussed further below.

In this chapter we focus on the origins of DNA replication in *T. brucei*, since the validation that TbORC1/CDC6 acts in replication means it offers the opportunity to define these sequences, which has so far not been possible. Although the molecular features that define the selection of chromosomal loci to act as origins of DNA replication in higher eukaryotic species (*Xenopus, Drosophila,* and humans) remain poorly understood (Aladjem *et al.* 2006), in bacteria, budding yeast and archaea consensus nucleotide sequence elements are preferentially selected to act as origins of replication [for review see (Sun and Kong 2010).

5.2 Results

5.2.1 Generation and verification of Myc-tagged TbORC1/CDC6

In order to identify DNA sequences to which TbORC1/CDC6 binds, we chose to adopt a chromatin immunoprecipitation (ChIP) approach. To this end, a vector for Myc-tagging of TbORC1/CDC6 at the C-terminus was constructed, as described in Chapter 4. This was linearised and transfected into *TbORC1/CDC6* +/- mutant PCF TREU927 *T. brucei* cells. Western blot analysis was carried out to confirm expression of TbORC1/CDC6-Myc and to test the specificity of the anti-Myc antibody used; Southern blotting was carried out to confirm the genotype of transformants and demonstrate functionality of the Myc-tagged TbORC1/CDC6 protein. All these experiments have been described previously in Section 4.5. Here, the functionally validated TbORC1/CDC-Myc expressing line was used for ChIP experiments to study its interactions with DNA *in vivo*.

5.3 Chromatin immunoprecipitation (ChIP) and Microarray technology (chip) – [ChIP-chip]

5.3.1 Description of methodology

ChIP-chip (also known as ChIP-on-chip) is an *in vivo* technique that combines the isolation of DNA sequences by chromatin immunoprecipitation (ChIP) and their subsequent identification via a microarray-based approach (chip). The ultimate goal of a ChIP-chip experiment is to query a genome in order to determine what DNA sequences a given protein binds (Kim & Ren 2006). A summary of the

methodology is shown in Figure 5-1. Briefly, in a ChIP-chip experiment, a cross-linking step is used to stabilise DNA-protein interactions in cells. The cells are lysed and the chromatin is sheared, the target protein is enriched by immunoprecipitation using a specific antibody or antiserum, DNA-protein interactions are then de-crosslinked, and DNA is purified from proteins and RNA. The generate sufficient material for analysis, the purified DNA is PCR-amplified and labelled differentially from input DNA (refers to whole genomic DNA that is not subjected to ChIP, but treated in the same way as ChIP material prior to IP). Both labelled DNAs are then co-hybridised on a microarray (chip), and the ratio of ChIP DNA versus input DNA is quantified. For TbORC1/CDC6, we adapted this generalised procedure as described below.

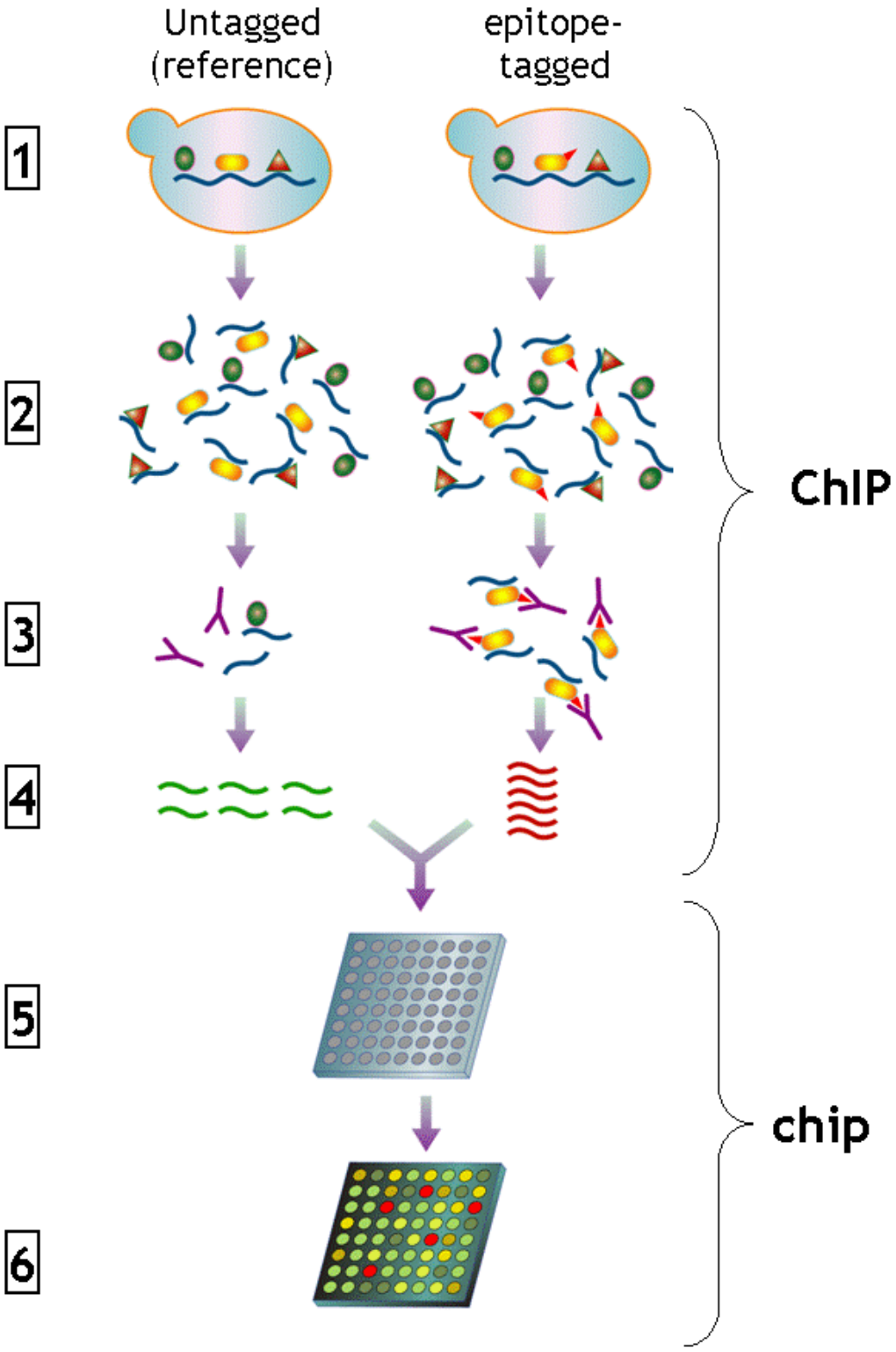


Figure 5-1 – Workflow overview of a ChIP-chip protocol

chIP: (1) Crosslink proteins to DNA; (2) Extract and shear crosslinked DNA; (3) Immunoprecipitate with specific antibody; (4) Reverse crosslink, PCR-amplify and differentially label DNA. Chip: (5) co-hybridize labelled DNA to microarray; (6) Red/Green fluorescence ratio at each position on array is proportional to extent of binding. Taken from (Pollack & Iyer 2002)

For ChIP in PCF *T. brucei* cells, an established protocol from the lab of G.A.M Cross was adapted (See Materials and Methods for a detailed protocol). Briefly, 10^8 PCF cells were fixed with 11 % formaldehyde for 20 min at RT. After cross-linking, formaldehyde was neutralized by addition of Glycine to a final concentration of 125 mM. After five minutes incubation, cells were washed twice in ice-cold PBS, lysed and the chromatin was sonicated using a Bioruptor® (Diagenode). Insoluble material was removed by centrifugation at 4000 g for 20 mins at 4°C, and the supernatant diluted. Samples of this material were retained as 'input' DNA. The sheared chromatin was next incubated with IgG-coated magnetic beads coupled to a ChIP-validated anti-Myc monoclonal antibody (Millipore®) to precipitate TbORC1/CDC6-Myc crosslinked to DNA. As a negative control for non-specific factors binding to the beads or antibody, a TREU927 wild type cell line (not transfected with the TbORC1/CDC6-Myc construct) was treated in exactly the same way. The beads were washed seven times, and samples retained as 'flow through'. IP material was then eluted from the beads, the cross-links reversed to separate DNA from protein, and the DNA purified using a Qiagen® Gel extraction protocol. Nicks introduced in the DNA during sonication were repaired using a Quick Blunting kit from NEB, according to manufacturer's protocol. Contaminating proteins and RNA were purified from the ChIP DNA using Proteinase K and RNase A, the DNA was cleaned again using a Qiagen® Gel extraction kit, PCR-amplified using a Whole Genome Amplification kit (SIGMA) and purified using a PCR purification Kit from Qiagen®.

5.3.2 Confirmation of chromatin immunoprecipitation by western blot

Prior to the purification of TbORC1/CDC6-Myc ChIP DNA from the crosslinked protein-DNA complex, the input sample and ChIP samples were checked by western blotting using an antibody to the Myc tag to test if the anti-Myc antibody was specific to TbORC1/CDC6-Myc tagged protein, and that the ChIP experiment was successful. A 20 µl aliquot was taken from both the input and ChIP samples and separated on by 10 % SDS-PAGE (Invitrogen). The gel was transferred to a nitrocellulose membrane (GE Healthcare), the membrane was probed with an anti-Myc antibody and detected using an ECL western blotting detection kit (described in Materials and Methods).

The results in Figure 5-2 showed that no band was visible in the samples from the TREU927 wild type control cell line where TbORC1/CDC6 is not Myc-tagged, while a single band of 66 kDa visible was visible in the input, to some extent in the flow through (FT), and in the eluate lanes from the TbORC1/CDC6-Myc cell line (Figure 5-2). This indicated that TbORC1/CDC6-Myc could be successfully immunoprecipitated from whole cell extract after cross-linking to chromatin and sonication, and that the anti-Myc antibody did not cross-react significantly with other antigens in the whole cell lysate. From Figure 5-2 it is also apparent that not all crosslinked TbORC1/CDC6-Myc bound to the beads, as seen by a band of the same size in the flow-through (FT) material.

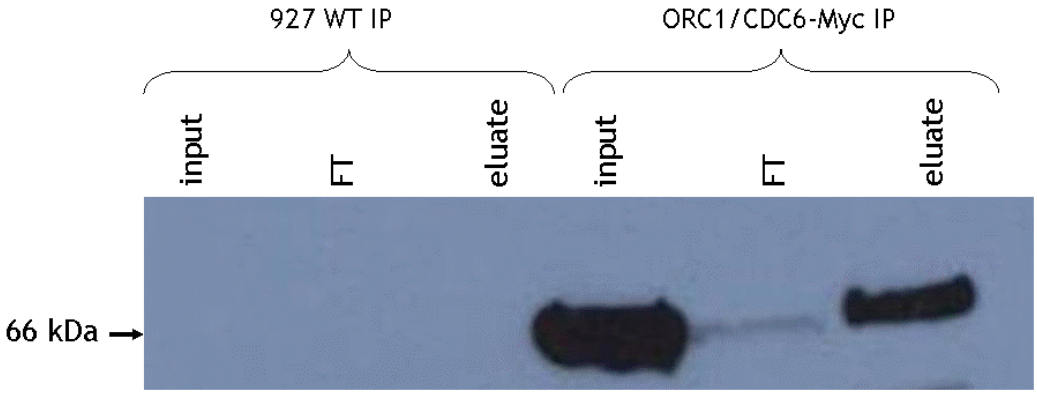


Figure 5-2 – Western blot of TbORC1/CDC6Myc during chromatin imunoprecipitation (chIP)

Western blot of whole cell extract after sonication step before IP (input), unbound material (FT), and after IP (eluate) using anti-Myc antibody against TbORC1/CDC6-Myc (ORC1/CDC6-MYC IP) or in a control cell line (TREU 927 wild type (WT) IP).

5.3.3 Confirmation of chromatin immunoprecipitation by PCR

After confirmation of TbORC1/CDC6Myc IP by western blot in this chIP protocol, DNA was purified and whole genome amplification employed as described above. The amount of DNA recovered after this step, from both WT and TbORC1/CDC6-Myc-tagged input and ChIP material, was measured using a NanoDrop analyser and 5 µl of DNA was resolved on a 1.5 % agarose gel, stained with SYBR safe (Invitrogen) and imaged with a transilluminator (BioRad). The results showed that the purified ChIP and input DNA had a product size range of ~200 bp to ~1 Kb (Figure 5-3). After quantification by NanoDrop, ~ 5 µg of chIP DNA was sent for hybridisation to a custom-designed *T. brucei* microarray while 100 ng was

sent for Solexa sequencing (Illumina). Samples from the WT experiments were not sent as the amount of DNA recovered from ChIP and subsequent whole genome amplification did not produce sufficient DNA for co-hybridisation on the microarray (Figure 5-3).

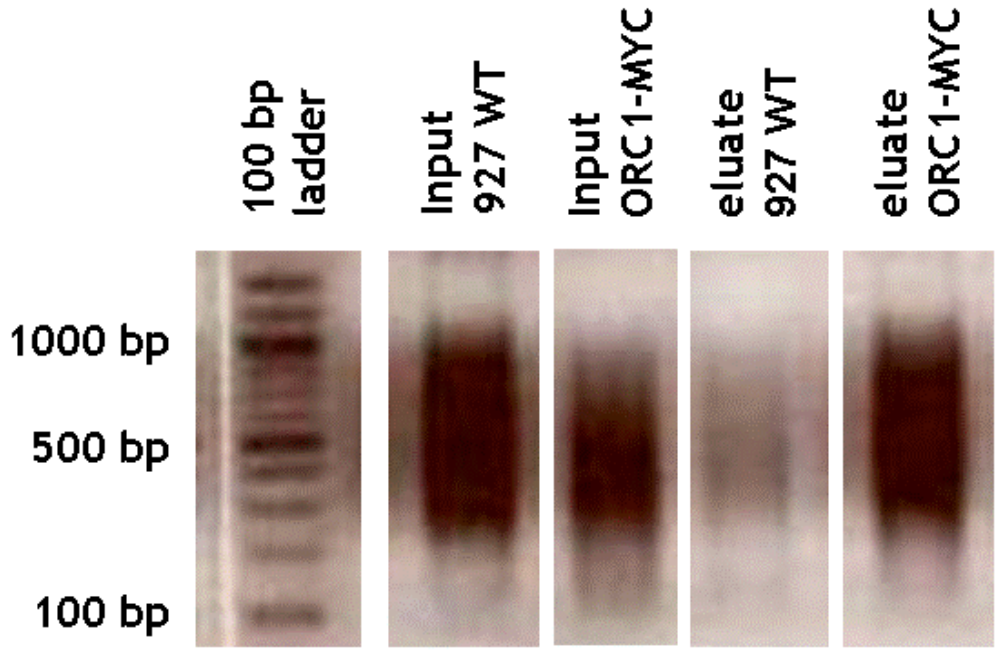


Figure 5-3 – Purified input and ChIP DNA after whole genome PCR-amplification

An Image is shown of a 1.5 % agarose gel of purified DNA before (input) and after IP (eluate) using a TbORC1/CDC6-Myc cell line (ORC1-MYC) and a control cell line (927 WT). Size markers are shown on the far left lane (100 bp ladder).

5.4 Microarray (chip) design, output and data analysis

This section will cover the analyses of the microarray data post co-hybridisation of TbORC1/CDC6-Myc IP (labelled with Cy5) and input sample (labelled with Cy3). After co-hybridisation, the data was extracted using NimbleScan software and the ratio of ChIP versus input fluorescent intensity signal of the corresponding Cy dyes was used to call peaks. This was performed by NimBleGen, and part of the data analysis described below was carried out in conjunction with L. Marcello (WTCMP, University of Glasgow), a collaborator on the DNA replication project.

5.4.1 Defining chromosome regions in the context of ChIP-chip analysis

The majority of the *T. brucei* genome consists of 11 pairs of megabase chromosomes (that vary in size from 1 – 6 Mb). The core of the megabase chromosomes, which are known to harbour most of the transcribed 'housekeeping' genes, is diploid, while the *VSG* array-containing subtelomeres are aneuploid (Berriman *et al*. 2005; Melville *et al*. 1998). Together with these megabase chromosomes, the genome is also composed of a few linear intermediate size (200-900 kb in length) and ~100 linear minichromosomes (~ 50-150 kb) (Wickstead *et al*. 2004; Melville *et al*. 1998). In this chapter our analysis has focused only on the megabase-sized chromosomes since ~95 % of this component of the genome has sequence available and was used for our microarray design (Berriman *et al*. 2005). To facilitate data analysis, the genome was divided into three categories: chromosome core, subtelomere-proximal and *VSG* array. Whereas the first category comprises housekeeping genes, the second has been defined as containing *INGI* retrotransposon sequences, *ESAG* genes, VR (*VSG*-related) genes, and *retrotransposon hot spot* loci (*RHS)*. The third category, *VSG* arrays, has a much more defined structure, consisting mostly of tandem repeats of what has been termed a *VSG* cassette, the fundamental unit that is copied into the *VSG* expression sites by gene conversion during antigenic variation (Figure 5-8). The *VSG* cassette starts at a 70 bp repeat and includes a *VSG5'* flanking sequence (hereafter referred to as 5' *VSG* component), the *VSG ORF* itself, and a *VSG 3'* flanking sequence (referred to as 3' *VSG* component); the latter is in turn flanked by a downstream 70 bp repeat, which become the start of the next *VSG* cassette. With these categories in mind, a program was written by L. Marcello in Perl to demarcate the megabase chromosomes into "core", "subtelomere-proximal" and "*VSG* array" sections. Beginning from one chromosome end, genes along the length of each megabase chromosome were analysed sequentially and upon arrival at the coordinates of the first *VSG* along the chromosome, a "*VSG* array" feature would be generated while *RHS*, *VSG*-related and *ESAG* genes (but not *ESAG3*, as this is also in *VSG* arrays) would generate a "subtelomere-proximal" feature (referred to as the "subtelomere" feature, for short). All hypothetical genes were skipped, and a "core" feature generated, which corresponds to the region encompassing all other non-

hypothetical genes. Gaps were then joined. Some short regions containing subtelomeric genes present at strand-switch regions (SSRs) in the core of chromosomes were identified, and these correspond with where chromosome fusions were shown to have occurred in the past (El Sayed *et al.* 2005a). These were left as having a "subtelomere" feature, even though they are in the core of the chromosome. A schematic representation of these features on Chr6, as an example, is shown in Figure 5-4.

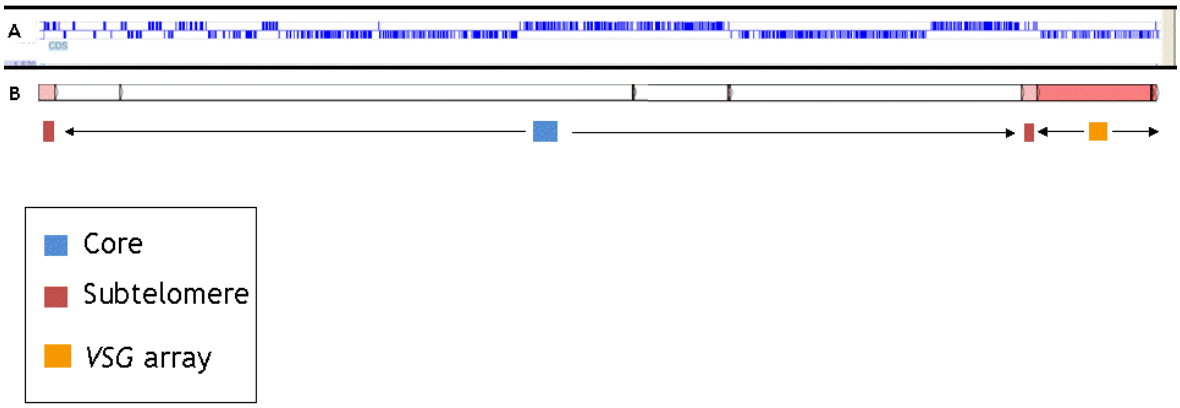


Figure 5-4 – Demarcation of chromosome features on megabase chromosomes

(A) Schematic representation of *T. brucei* chromosome 6 coding sequence. Blue blocks indicate genes in directional gene clusters along the chromosome, either in the forward strand (top) or in the reverse strand (bottom). (B) Chromosome features predicted by Perl script to demarcate the chromosome "core" (blue box), the "subtelomere" (red box), and the "*VSG* array" (orange box)

5.4.2 Design of Microarray and hybridisation

Chip design was performed by Roche Diagnostics NimbleGen (USA) using a selected set of *T. brucei* TREU927 DNA sequence files (Appendix 4), obtained from the Pathogen Sequencing Group at the Wellcome Trust Sanger Institute (Cambridge) and chosen to provide the best coverage possible of megabase chromosomes and some examples of telomeric sequences (only Chromosome 1 and 10 have been assembled in their entirety) (Renauld *et al.* 2007). The NimbleGen Tiling ChIP Service (design ID 19441, design name 090109_Tbru_LM_CHIP) was used. 385,816 probes were tiled across all unique regions of the *T. brucei* genome at an average spacing of 61 bp. The coverage of probes relative to chromosome features described above is as shown on Table 5-1.

Table 5-1 – Probe coverage compared to chromosome architecture (ng = NimbleGEN)

	Total length of region (bp)	Overlaps with ng probes (bp)	% Probe coverage
"Core"	17713126	12933634	73.0%
"Subtelomere"	1111257	319507	28.8%
"*VSG* array"	2188340	1344036	61.4%

Labelling, hybridization and data acquisition was performed by NimbleGen, according to standard procedures. Two independent ChIP DNA samples from the TbORC1/CDC6myc PCF cell line were evaluated. However, only the input DNA from one sample was sent for co-hybridisation - hence the IP sample which did not have a corresponding input DNA sample was not analysed further. The control ChIP and input DNA samples from 927 WT cells were not sent for co-hybridisation on the array. The data sets were viewed using NimbleGen SignalMap software either as a scaled log2 ratio or as filtered peak data generated after normalisation of the log2 ratio data.

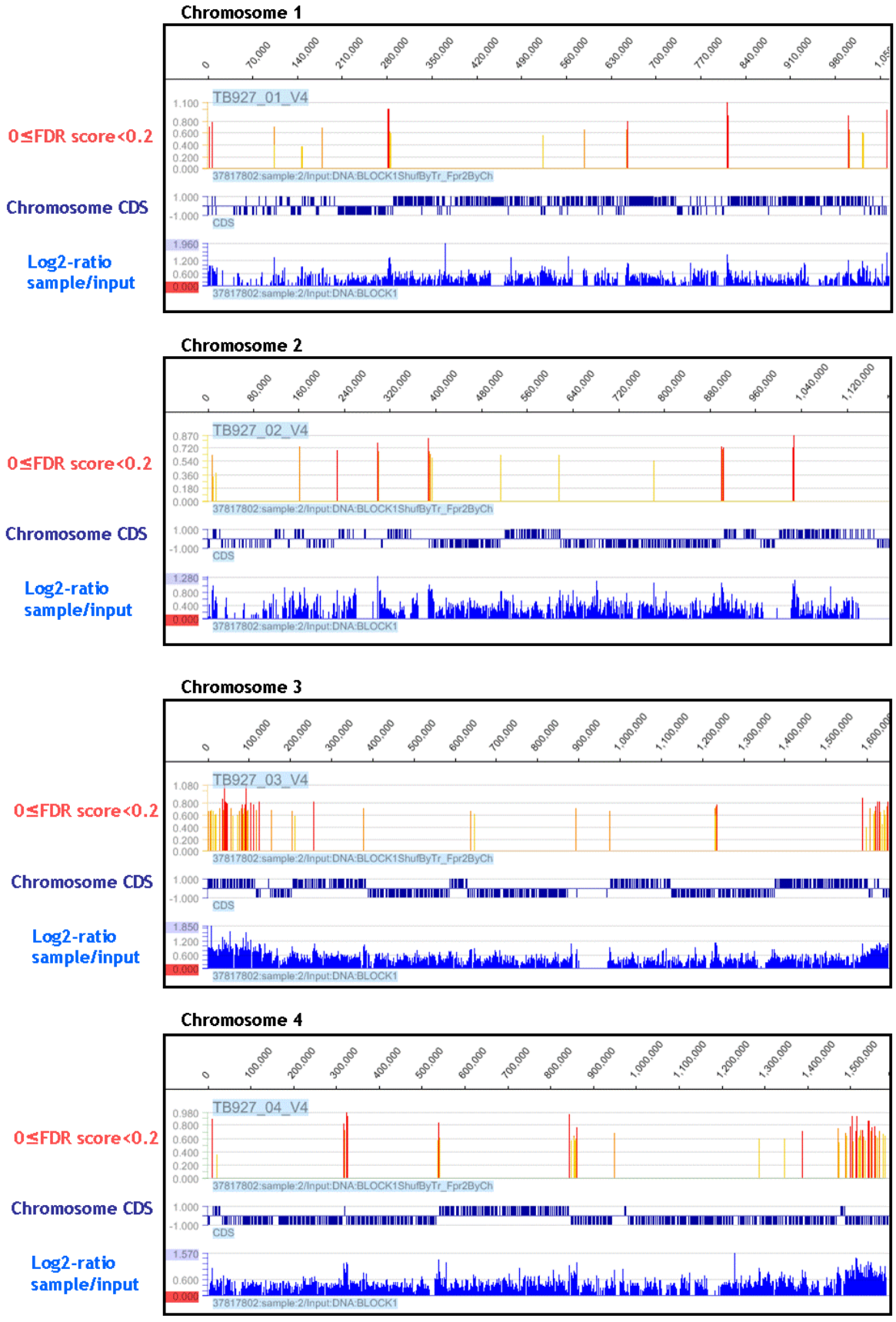
From the NimbleGen ChIP-chip guide, the scaled log2 ratio is defined as "the ratio of the input signals for the experimental and test samples that were co-hybridized to the array. The log2 ratio is computed and scaled to center the ratio data around zero. Scaling is performed by subtracting the biweight mean for the log2-ratio values for all features on the array from each log2-ratio value". In our case, the experimental sample refers to the input DNA (input) while the test sample refers to the ChIP DNA (sample). According to the same guide, for peak detection, "the scaled log2 ratio is normalised by searching for four or more probes whose signals are above the specified cut-off values, ranging from 90% to 15%, using a sliding window of 500 bp. The cut-off values are a percentage of a hypothetical maximum, which is the mean + 6[standard deviation]. The ratio data is then randomized 20 times to evaluate the probability of "false positives". Each peak is then assigned a false discovery rate (FDR) score based on the randomization". FDR scores ≤ 0.05 indicate the highest confidence of binding sites, while FDR scores > 0.2 are generally not considered as high confidence binding sites.

5.5 TbORC1/CDC6 binding site analysis:

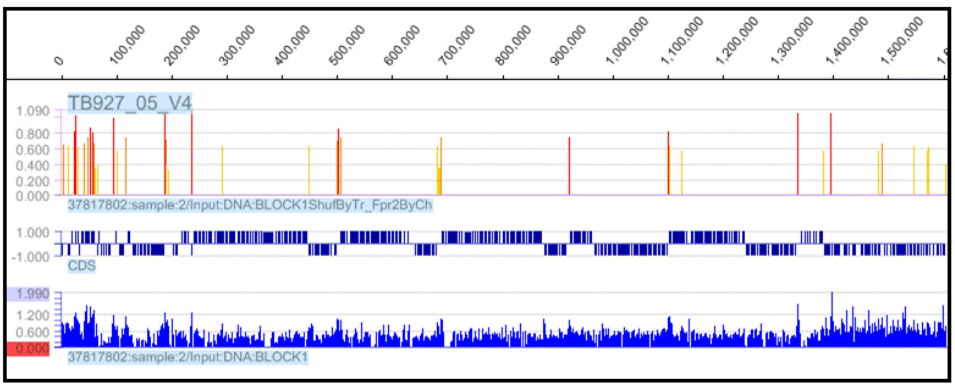
5.5.1 Chromosomes overview; schematic representation of TbORC1/CDC6 binding sites

After co-hybridisation, data acquisition and analysis by NimbleGen, the scaled log2 ratio and the peak data sets were aligned with the annotated genome sequence coordinates for ten chromosomes (Chr1-10) from *T. brucei* strain TREU927 using the NimbleGen SignalMap software. The contiguous sequences for Chr11 version 4 that was used for probe design is in the process of being assembled, hence could not yet be included in the analysis. As shown in Figure 5-5, for each chromosome, the scaled log2 ratio data is manually adjusted to a minimum value of zero on the y-axis to show the fluorescent signal intensity of the Cy5-labelled TbORC1/CDC6-Myc ChIP DNA (sample) enrichment relative to the fluorescent signal intensity of the Cy3-labelled input DNA (input). Positive log2 ratios of Cy5/Cy3 are indicated by peaks and these correspond to a relative excess of the DNA in the Cy5-labelled sample, while log2 ratios of Cy5/Cy3 near zero correspond to an equal abundance of the Cy5-labelled sample and Cy3-labelled input DNA (Figure 5-5, log2-ratio sample/input). In Figure 5-5, for each chromosome, the middle panel shows the chromosome coding sequence (chromosome CDS) with blue boxes representing ORFs for both the forward and reverse coding strands above and below a horizontal line, respectively. The chromosome CDS coordinates are aligned with the log2 ratios for each probe on the array to illustrate the coordinates of loci within chromosome regions where Cy5-labelled sample DNA is enriched for on the array. The top panel shows assigned peaks (assignment described above) with 0≤FDR score<0.2, which indicate putative TbORC1/CDC6 binding sites identified. The 0.2 FDR score was empirically chosen as a cut-off, as many binding sites with an FDR score higher than 0.05 still appeared to show a significant, non-random distribution. At FDR score < 0.2, 685 binding sites were identified for the ten chromosomes analysed while at a more stringent FDR score cut-off (≤ 0.05) a total of 242 binding sites were identified. The global distribution pattern of the reduced number of binding sites remains unaltered whether the FDR score cut-off is set at 0.05 or at 0.2 (discussed below).

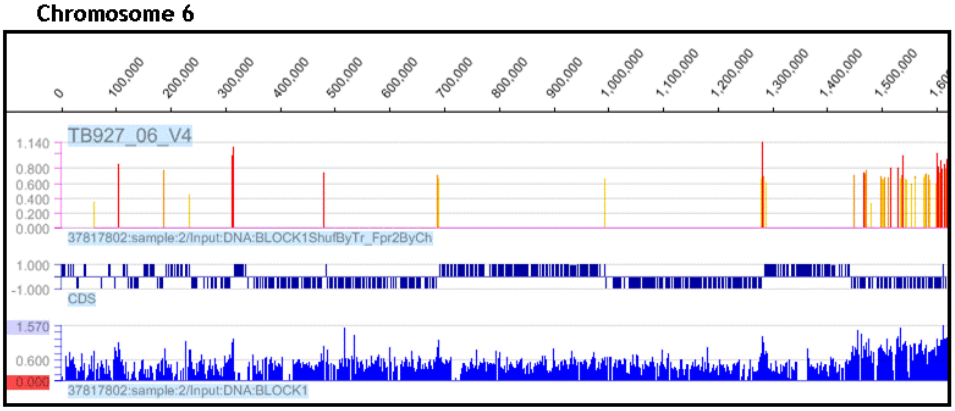
In summary, the results revealed that TbORC1/CDC6 binding within the chromosome interior (“core”) localised mainly to regions separating directional gene clusters (strand switch regions, SSRs), while the subtelomeric ends of chromosomes showed high density clusters of TbORC1/CDC6 binding (Figure 5-5). A detailed analysis of binding sites within the chromosome core and chromosome ends is described below.


Chromosome 1
0≤FDR score<0.2
Chromosome CDS
Log2-ratio sample/input
TB927_01_V4
37817802:sample:2/Input:DNA:BLOCK1ShufByTr_Fpr2ByCh
CDS
37817802:sample:2/Input:DNA:BLOCK1
Chromosome 2
0≤FDR score<0.2
Chromosome CDS
Log2-ratio sample/input
TB927_02_V4
37817802:sample:2/Input:DNA:BLOCK1ShufByTr_Fpr2ByCh
CDS
37817802:sample:2/Input:DNA:BLOCK1
Chromosome 3
0≤FDR score<0.2
Chromosome CDS
Log2-ratio sample/input
TB927_03_V4
37817802:sample:2/Input:DNA:BLOCK1ShufByTr_Fpr2ByCh
CDS
37817802:sample:2/Input:DNA:BLOCK1
Chromosome 4
0≤FDR score<0.2
Chromosome CDS
Log2-ratio sample/input
TB927_04_V4
37817802:sample:2/Input:DNA:BLOCK1ShufByTr_Fpr2ByCh
CDS
37817802:sample:2/Input:DNA:BLOCK1

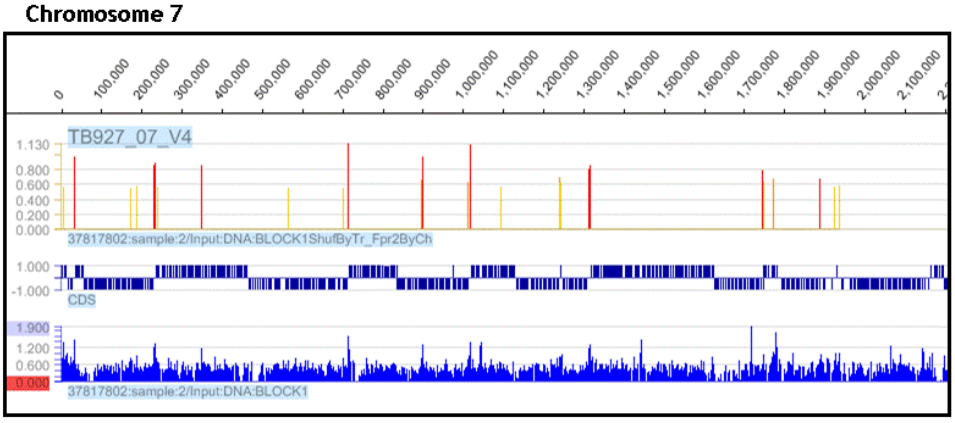
Chromosome 5



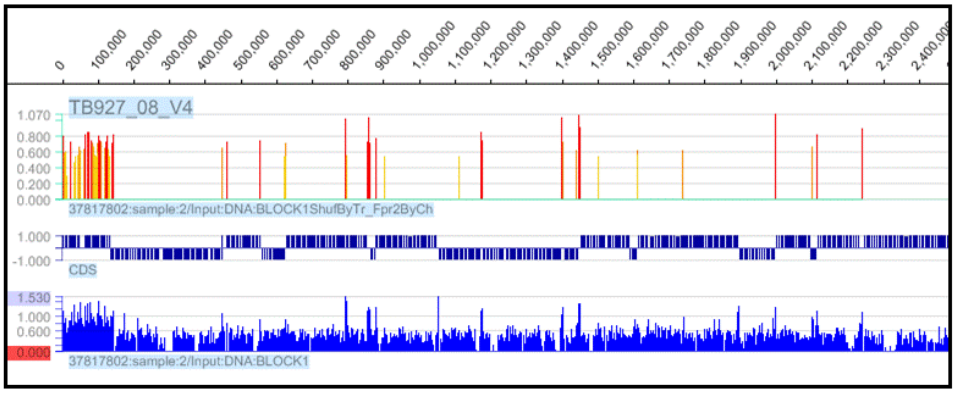
Chromosome 6



Chromosome 7



Chromosome 8



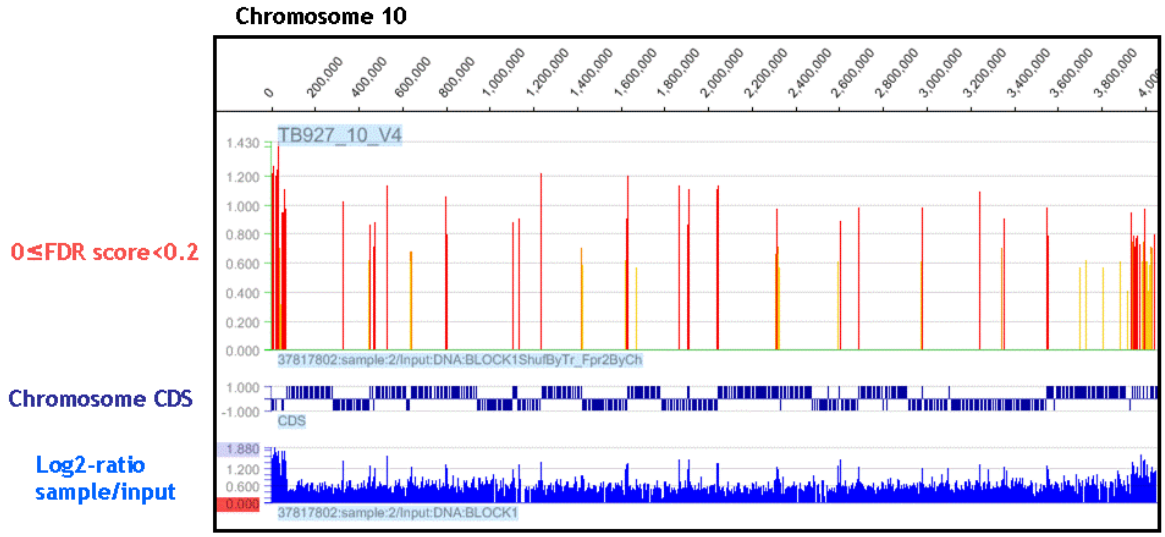


Figure 5-5 – Distribution of TbORC1/CDC6 binding sites from ChIP-chip

TbORC1/CDC6-Myc chIP and control input DNA samples were labelled with Cy5 and Cy3, respectively, and co-hybridized to a *T. brucei* 385K tiled array. Chromosomes 1-10 are shown, not to scale. Each chromosome is shown in the same way. Bottom panel: raw data were analysed using NimbleScan software and visualised using SignalMap software (NimbleGen) producing signal ratio values over a 500 bp sliding window; indicated by (log2-ratio sample/input); the scale on the y-axis indicates the ratio of the signals derived from the corresponding co-hybridisation, and the x-axis indicates the genomic coordinates of the microarray tiles. Middle panel: Chromosome coding sequence (Chromosome CDS) is depicted by blue boxes showing both the forward coding strand above (top strand, 1) and the reverse coding strand below (bottom strand, -1) a horizontal line. Top panel: a statistical method was used to generate p-value enrichment from the log2 ratio data with each probe assigned a false discovery rate (FDR score set here at < 0.2) for assignment of peaks which are indicated by vertical red and orange lines (0≤FDR score<0.2); FDR scores ≤ 0.05 indicate the highest confidence of binding sites are depicted by red vertical lines, while FDR scores <0.2 indicate low confidence of binding sitesand are depicted by orange vertical lines. The *T. brucei* chromosome coordinates (bp) are indicated by the numbers at the top

5.5.2 Chromosomes features and TbORC1/CDC6 binding

A quantitative analysis of TbORC1/CDC6 ChIP-chip data for chromosomes 1 to 10 (21 Mb in length) was performed on the chromosome features predicted by the Perl script described in Section 5.4.1 to investigate the degree of overlap between TbORC1/CDC6 occupancy with the three assigned chromosome features: “core”, “subtelomere” and “*VSG* array”. A representative chromosome (Chr6) is shown in Figure 5-6A to illustrate the distribution of TbORC1/CDC6 peaks and the assigned chromosome features are shown beneath it. The results revealed that, at a p-value of 0≤FDR score<0.2, a total of 685 TbORC1/CDC6 putative binding sites were present across the ten chromosomes. The sum of all “core” regions, with a total size of 17.7 Mb and 73 % probe coverage, housed 40 % (278) of TbORC1/CDC6 binding sites (Figure 5-6B). The “subtelomere”

regions, with a total size of 1.1 Mb and 29 % probe coverage, had 12 % (81) TbORC1/CDC6 binding sites (Figure 5-6B). The "*VSG* array" regions, which had a total size of 2.1 Mb, and 61 % coverage by array probes, had 48 % (330) of the total number of TbORC1/CDC6 binding sites (Figure 5-6B). Within the "core", the spacing of unique binding sites (binding sites that are closer than 2 kb are grouped together) is ~ 110 kb, compared with ~ 30 kb within the "subtelomere", and ~ 7kb with the "*VSG* array". In total, the putative TbORC1/CDC6 binding sites cover a region of 450 kb, which corresponds with ~ 2 % of the genome sequence contained within chromosomes 1-10.

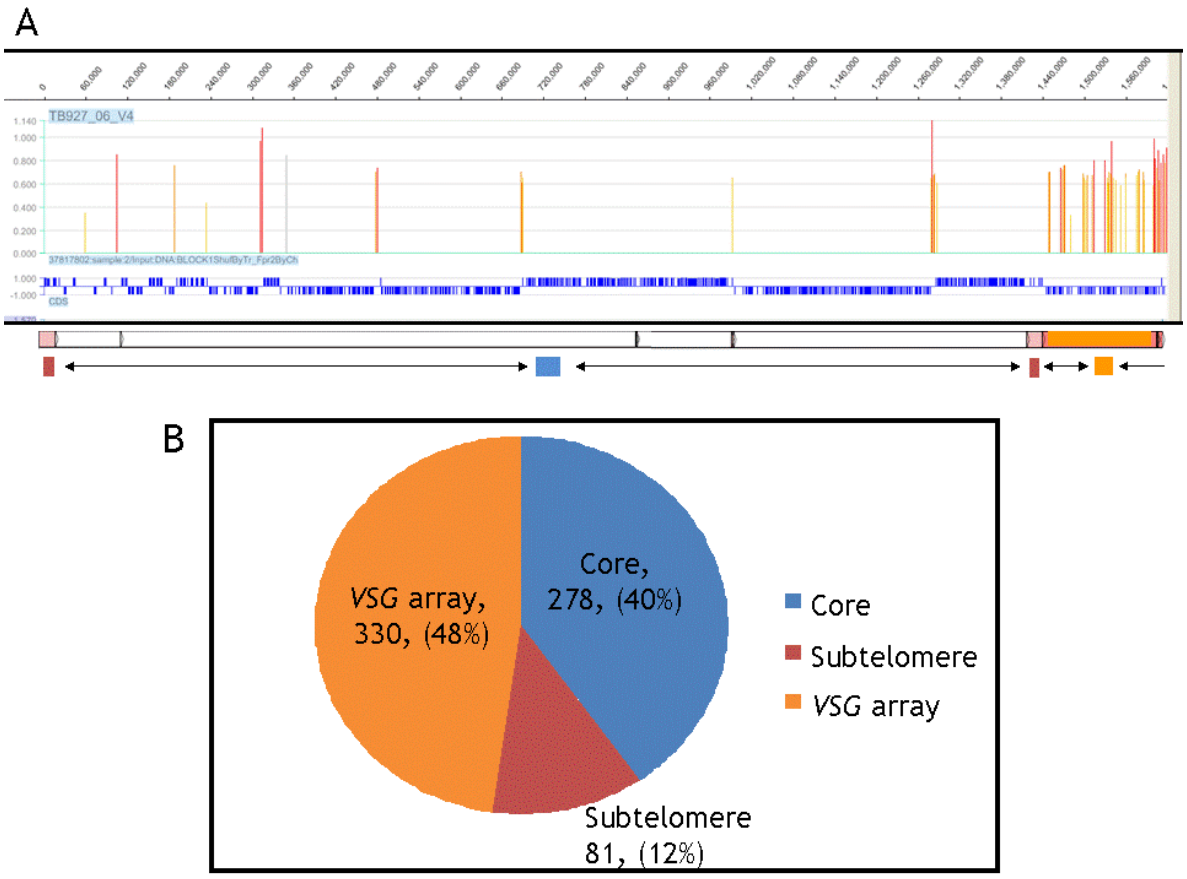


Figure 5-6 – Overlap of TbORC1/CDC6 binding sites and chromosome features

(A) SignaMap output graphic representation of chromosome 6, showing genome coordinates (top), TbORC1/CDC6 binding sites with p-value 0≤FDR score<0.2 indicated by vertical red and orange lines, and chromosome forward and reverse coding strands (CDS) indicated by blue boxes above and below a horizontal line, respectively. Chromosome features predicted by Perl script to demarcate chromosome "core" (blue box), from "subtelomere" (red box) and from "*VSG* array" (orange box) regions are shown. (B) A Pie chart depicting the overlap between TbORC1/CDC6-associated binding site sequences and Perl script-defined chromosome features (core,

subtelomere and *VSG* array) for Chr1-10 combined; numbers indicate total number of binding sites for each region with a corresponding percentage of overlap with features indicated in brackets.

5.5.3 Polycistrons, strand switch regions (SSRs) and TbORC1/CDC6 binding

A quantitative analysis of TbORC1/CDC6 binding within the "core" features of Chr1-10 was performed to investigate the degree of overlap with polycistrons (on forward or reverse strand) and SSRs (either convergent or divergent). Chromosome (Chr) 6 is shown in Figure 5-7A as an example to illustrate the distribution of TbORC1/CDC6 binding sites within the chromosomes and their relationship with the organisation of the genes in directional gene clusters, indicated by the direction of arrows in Figure 5-6B. The results revealed that, within the "core" of the ten chromosomes analysed, the forward coding polycistrons had a total of 68 (24 %) TbORC1/CDC6 binding sites, while the reverse coding polycistrons had a total of 84 (29%) TbORC1/CDC6 binding sites (Figure 5-7C). The comparable distribution of binding sites between forward and reverse polycistrons is not surprising, as they are functionally equivalent. More interestingly, TbORC1/CDC6 binding within the "core" mainly localised to divergent SSRs: 114 (38%) binding sites compared to 23 (8%) binding sites for convergent SSRs. In nine out of ten chromosomes analysed there are a total of 46 divergent SSRs, 37 (80%) of which had at least one TbORC1/CDC6 binding site. Of the 43 convergent SSRs, only 6 (13%) had TbORC1/CDC6 binding sites. It appears, therefore, that TbORC1/CDC6 binding in PCF *T. brucei* cells appears to favour divergent SSRs.

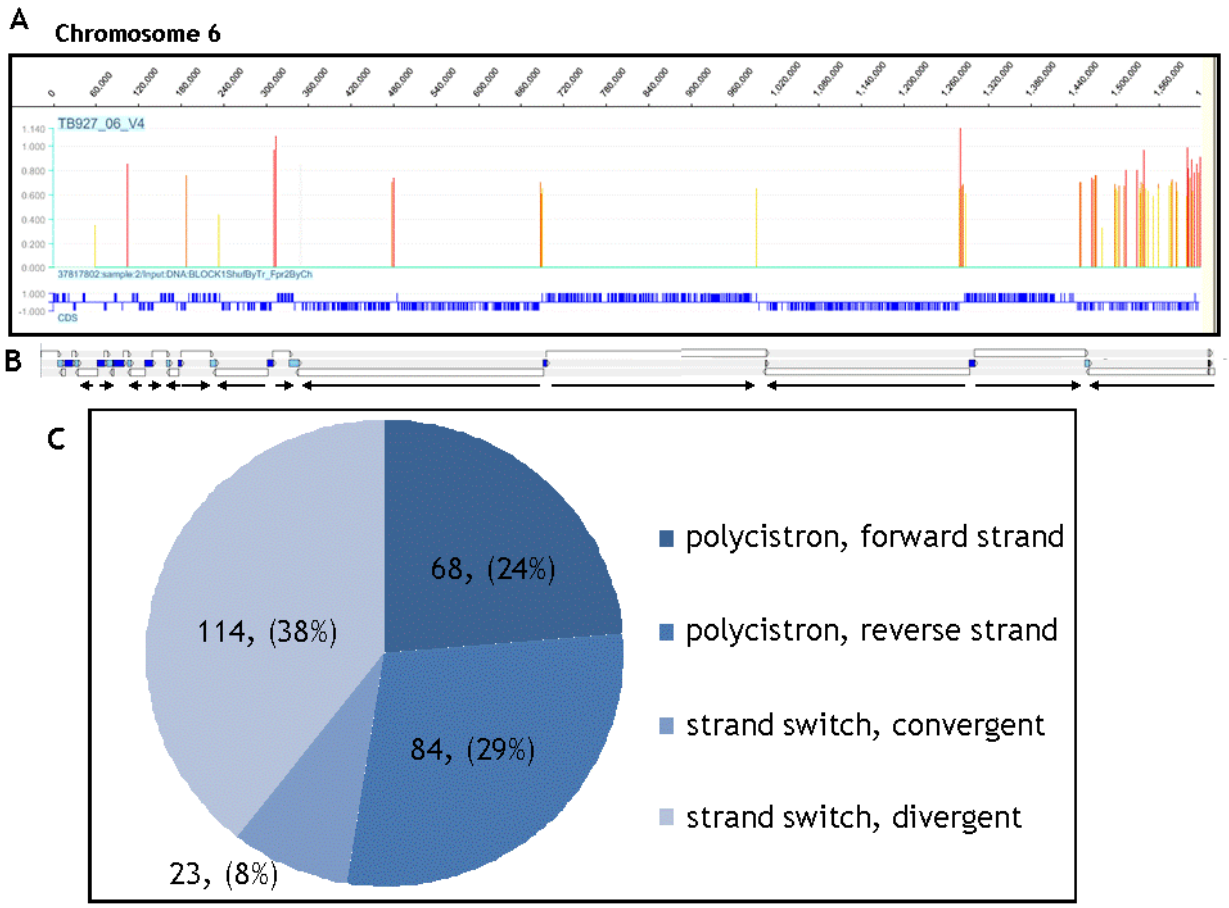


Figure 5-7 – Distribution of TbORC1/CDC6 binding sites within polycistrons and strand switch regions

(A) SignaMap output graphic representation of chromosome 6 showing genome coordinates (top), TbORC1/CDC6 binding sites with p-value 0≤FDR score<0.2 indicated by vertical red and orange lines, and chromosome forward and reverse coding strands (CDS) indicated by blue boxes above and below a horizontal line, respectively. (B) The distribution of directional gene clusters on chromosome 6. Clusters are boxed to reveal those transcribed from left to right (above) and those transcribed from right to left (below); direction of transcription is indicated by black arrows. (C) Pie chart depicting the overlap of TbORC1/CDC6-associated binding site sequences with either the forward coding polycistron (polycistron, forward strand), the reverse coding polycistron (polycistron, reverse strand), or with either convergent strand switches (strand switch, convergent) or divergent strand switches (strand switch, divergent) for Chrs 1-10 combined; numbers indicate total number of binding sites for each region with a corresponding percentage relating to the total number of core binding sites indicated in braces

5.5.4 VSG arrays and TbORC1/CDC6 binding

A *VSG* array is defined by a tandemly repeated unit structure as described above (section 5.4.1) and illustrated in Figure 5-8A (Barry *et al*. 2005). To investigate any potential relationship between TbORC1/CDC6 binding and the *VSG* array components, a program was written in Perl by L. Marcello to mark out the components that make up the *VSG* array: 70 bp repeat, 5' *VSG* component, *VSG* ORF, 3' *VSG* component and 70 bp repeat (Figure 2-8A). Overall, across the ten

chromosomes analysed, 8.6 % of 70 bp repeat sequence, which had a total size of 47 kb, was occupied by TbORC1/CDC6, corresponding with 87 binding sites (Figure 5-8B). 14.8 % of 5' *VSG* component sequence, which had a total size of 774 kb, was bound by TbORC1/CDC6 (261 binding sites). 5.2 % of *VSG* sequence, which had a total size of 785 kb, was occupied by TbORC1/CDC6 (112 binding sites). Finally, 9.6 % of 3' *VSG* component sequence, which had a total size of 135 kb, was bound by TbORC1/CDC6 (80 binding sites). In summary, the 5' *VSG* component appeared to be enriched for TbORC1/CDC6 binding; 14.8 % of the total feature length overlapped with TbORC1/CDC6 binding sites, ~5 % more than any other part of the *VSG* cassette. Of all the TbORC1/CDC6 binding sites predicted to fall within the *VSG* arrays (540 in total), 48 % mapped to the 5' *VSG* component.

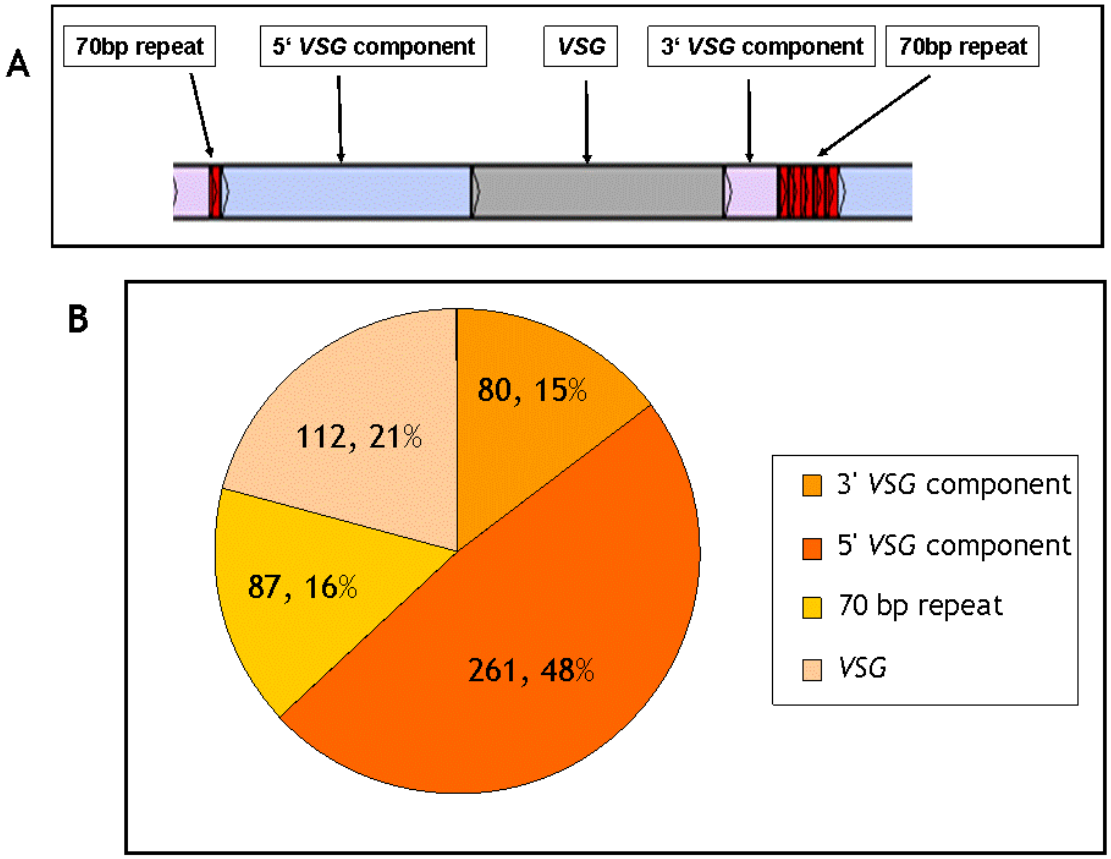


Figure 5-8 – Schematic representation of a typical *VSG* array cassette

(A) *VSG* array showing a typical cassette structure consisting of 70bp repeat, 5' *VSG* component, *VSG*, 3' *VSG* component, 70bp repeat. (B) Pie chart depicting the overlap of TbORC1/CDC6-associated binding site sequences with *VSG* array components (70bp repeats, 5' *VSG* component, *VSG*, and 3' *VSG* component) for Chr1-10 combined; numbers indicate total number of binding sites for each region with a corresponding percentage of overlap with feature indicated in braces

5.5.5 VSG expression sites and TbORC1/CDC6 binding sites

In Section 3.8.6, we proposed an involvement of TbORC1/CDC6 in the regulation of *MVSG* expression, and perhaps also in *BVSG* expression. Here, we have used Artemis (a bioinformatics software available from the Sanger Institute at http://www.sanger.ac.uk/resources/software/artemis/) to manually inspect *BES* and *MES* for which telomeric sequences are available from the TREU927 strain, as these were also included into the microarray design. Most of these sequences (with the exception of the GUTat10.1 BES and the ES present within assembled megabase chromosomes) were obtained from C. Hertz-Fowler (Sanger Institute, UK), and represent the best current assembly of TAR-cloned 927 chromosome ends, which has not been published [for TAR cloning of trypanosome telomeres, see Section 3.8.6, and the publication of *T. brucei* Lister 427 chromosome ends (Hertz-Fowler *et al.* 2008)]. It is important to highlight here that in this context when we refer to *VSGs,* we mean a telomeric *VSG* present at the end of a telomeric expression site, and not to a *VSG* present in the subtelomeric arrays described above. Telomeric *VSGs* are preceded by a much larger tract of 70 bp repeats when compared with *VSG* array 70 bp repeats, and are flanked upstream by *ESAGs* and pseudogenes (LaCount *et al*. 2001) (see Section 3.8.6 for detailed description of a BES).

5.5.5.1 TbORC1/CDC6 binding to Bloodstream expression sites (BES)

All the contigs for the analysis of BES and MES used for localising TbORC1/CDC6 binding sites were obtained from C. Hertz-Fowler, and have not yet been published, as they are still in the process of being assembled. We have analysed one BES and two 'pseudo-BESs', and each of the three will be outlined in turn.

Before the full genome sequence of *T. brucei* strain TREU927 was published, the DNA sequence of Chr1 had been sequenced and assembled; extending into the subtelomeric regions at both ends of the chromosome (Hall *et al.* 2003). At the left subtelomeric end of Chr1 there is a BES promoter that does not seem to be part of a canonical BES, as it is divergent in sequence and lacks BES-associated *ESAG6* and *ESAG7* (Hall *et al.* 2003). Although this subtelomeric end might not represent a typical BES (Pham, Rothman, & Gottesdiener 1997), we examined the loci for the presence of TbORC1/CDC6 binding sites. The results revealed

that TbORC1/CDC6 binding sites in this region do not co-localise with the putative promoter elements. Rather, the closest binding sites were located on either side of the 70bp repeats, 5-8 kb downstream of the promoter (FDR score = 0.05; Figure 5-9). No binding sites were found within 90 kb upstream.

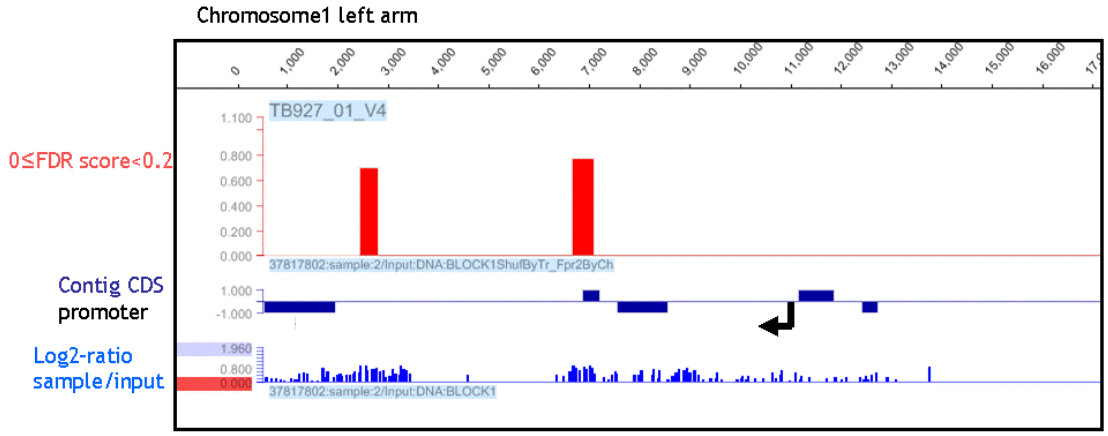


Figure 5-9 – TbORC1/CDC6 binding at the left subtelomeric BES of Chromosome 1 left arm

Signalmap out profile for a contig DNA sequence of the left arm of chromosome 1; All features are same as described above (Figure 5-5) except that in this case the location of the promoter for the BES is indicated by an arrow, with the direction of transcription shown by the arrow head; red vertical bars indicate TbORC1/CDC6 binding sites with FDR score = 0.05.

A TREU927 GUTat10.1-containing BAC-26P8 genomic DNA insert expressing *VSG 10.1* was sequenced by Lacount *et al* (2001). The 35 kb contig, which is part of chromosome 10, spans from 50 bp repeats at one end to telomeric repeats at the other end (LaCount *et al*. 2001). Downstream of the 50 bp repeats are typical BES features: a *VSG* promoter followed by seven *ESAG* genes (*ESAGs* 7, 6, 5, 8a, 3, 4 and 8b), five pseudogenes, 15 kb of 70 bp repeats, a single *VSG* ORF (*10.1*), followed by telomeric DNA repeats (LaCount *et al*. 2001). No TbORC1/CDC6 binding sites were present close to the promoter region, although there are probes spanning this region. Six TbORC1/CDC6 binding sites were found downstream of the promoter (between 5 and 32 kb away; Figure 5-10), but only one out of six has an FDR score of 0.05, the rest being 0.2. The TbORC1/CDC6 binding site with the highest confidence of binding (FDR score <0.05) in this ES is at a pseudo *VSG* sequence located ~12 kb downstream from the promoter.

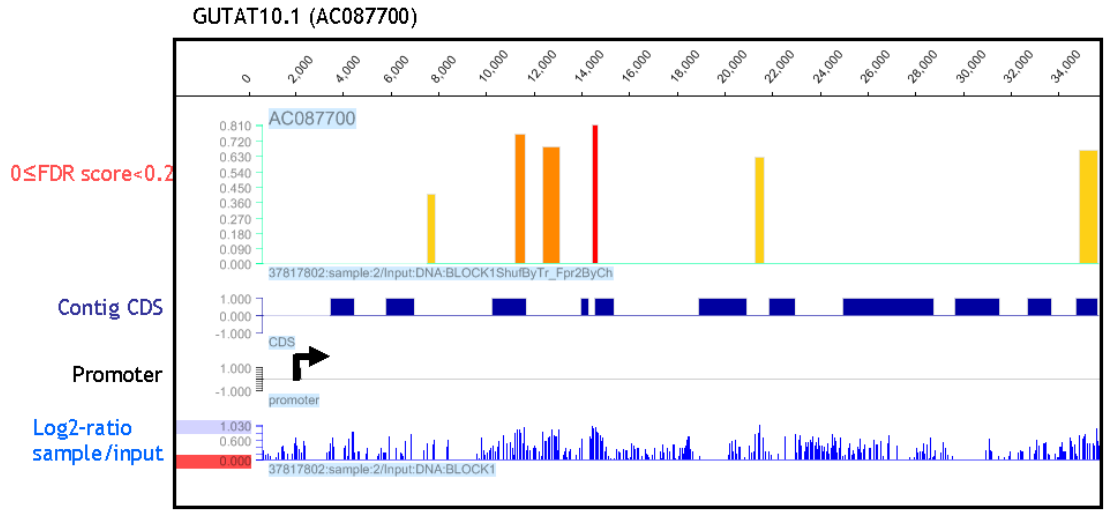


Figure 5-10 – TbORC1/CDC6 binding within the GUTat10.1 BES contig at the end of Chromosome 10

Signalmap out profile for a 35 kb contig DNA sequence of chromosome 10; All features are same as described above (Figure 5-5) except that in this case the location of the promoter for the BES is indicated by an arrow, with the direction of transcription shown by the arrow head; red vertical bar indicate TbORC1/CDC6 binding sites with FDR score = 0.05 while orange and yellow vertical bars indicate TbORC1/CDC6 binding sites with FDR score > 0.05.

A chromosome 8 TAR clone contig was also analysed, which consists of a degenerate BES in which a promoter is conserved with downstream *ESAGs* (*ESAG6*, *5*, *4*), but lacking a *VSG* ORF or further *ESAG*s. The *ESAG6* gene present in this sequence is a pseudogene, and there is no *ESAG7* (L. Marcello, personal communication). Examination of this contig revealed two TbORC1/CDC6 binding sites, each with a FDR score of 0.2; Figure 5-11). Both binding sites were located downstream of promoter, one around 3 kb downstream and the other overlapping with *ESAG5*, which is 6 kb downstream of the promoter (interestingly, one of the binding sites in the GUTat 10.1 BES also overlaps with *ESAG5*)

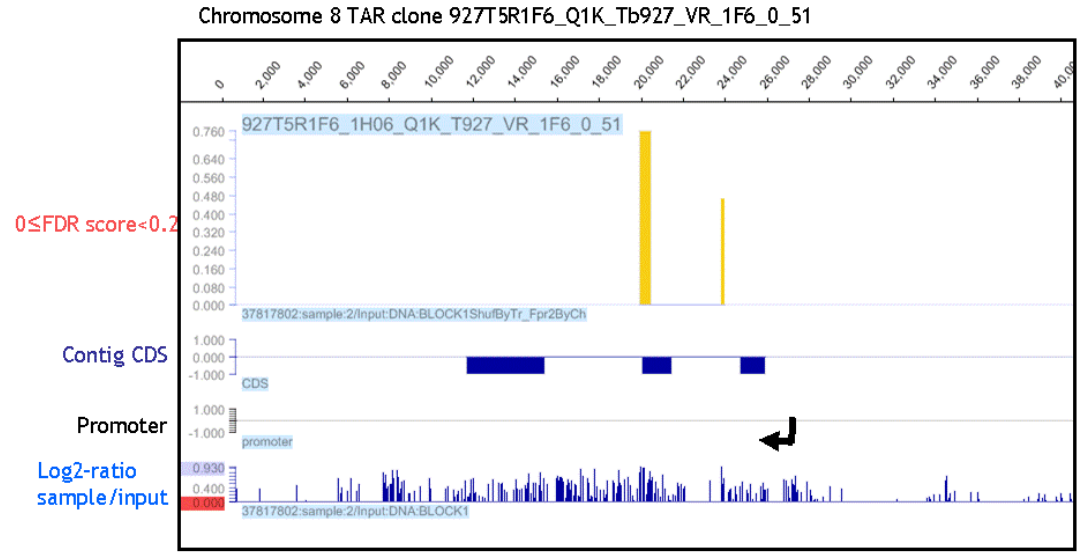


Figure 5-11 – TbORC1/CDC6 binding within a BES of Chromosome 8 TAR clone

Signalmap out profile for a chromosome 8 TAR clone; All features are same as described above (Figure 5-5) except that in this case the location of the promoter for the BES is indicated by an arrow, with the direction of transcription shown by the arrow head; yellow vertical bars indicate TbORC1/CDC6 binding sites with FDR score > 0.05.

5.5.5.2 TbORC1/CDC6 binding to metacyclic expression sites (MES)

For localising TbORC1/CDC6 binding sites in MESs, we analysed two MES on Chr10 and Chr11, and two pseudo MES on Chrs. 5 and 6. Again, each of these are outlined in turn.

The left end of Chr10 contains a fully assembled MES, with a promoter and a downstream *VSG*. As in the case for BES promoters, this MES promoter does not seem to be bound by TbORC1/CDC6. However, in contrast to the BES, TbORC1/CDC6 binding sites appeared much closer to the promoter, the nearest only 200 bp downstream and 100 bp upstream. The TbORC1/CDC6 binding sites (three in the 6 kb downstream of the promoter and ten within 25 kb upstream) all have low FDR scores (<0.05) and in total cover more than half of the 31 kb region considered (Figure 5-12). Within this region the TbORC1/CDC6 binding sites overlap with *ESAG1*, the functional *VSG* and an upstream pseudo-*VSG*.

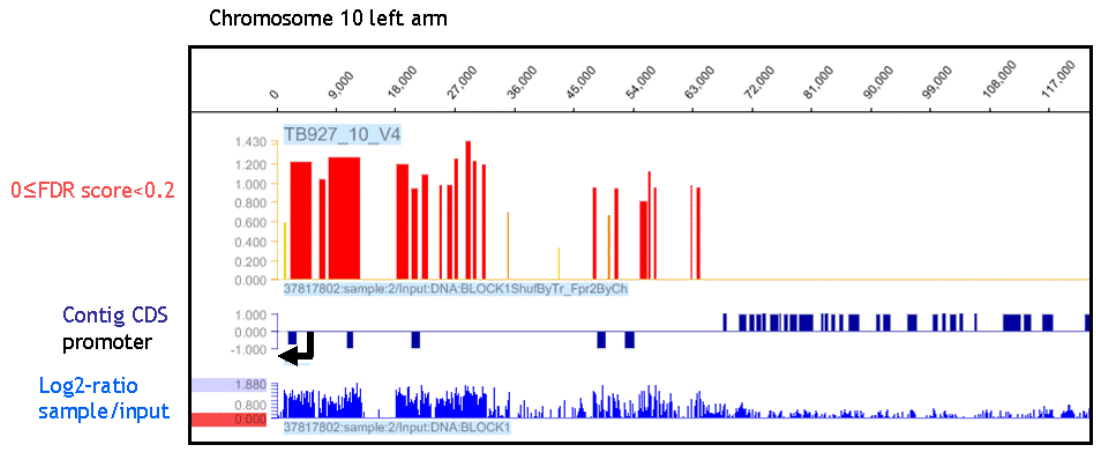


Figure 5-12 – TbORC1/CDC6 binding within a MES of Chromosome 10 left arm contig

Signalmap out profile for a contig DNA sequence of the left arm of chromosome 10 containing a MES is shown; All features are same as described above (Figure 5-5) except that in this case the location of the promoter for the MES is indicated by an arrow, with the direction of transcription shown by the arrow head; red vertical bars indicate TbORC1/CDC6 binding sites with FDR score = 0.05

The left end of Chr11 also contains a fully assembled MES, although it did not display the same level of TbORC1/CDC6 binding as the MES on Chr10: two binding sites were observed downstream (2.5 and 3.6 kb from promoter) and four upstream (between 5 and 12 kb upstream of promoter), and all exhibited a higher FDR score value (0.2), with the exception of the most telomere-proximal (overlapping the *MVSG*; FDR = 0.2).

Similar patterns were exhibited by two other potential MES (contigs for which are incomplete and do not include any potential downstream *MVSG*) on the left ends of both Chr5 and Chr6 (927T6L1A4-11g05.p1k, 6.7 kb; and 927T5L5E3-1f09.p1k, 13.5 kb). On the Chr5 MES, TbORC1/CDC6 binding sites surrounded the promoter in a similar manner to the Chr10 MES (Figure 5-12), ranging from 600 bp downstream and from 100 bp upstream, although the FDR score values were higher (0.2) and the binding site coverage of the contig is not as dense (less than 20%). Only a single TbORC1/CDC6 binding site, with an FDR score of 0.2, was present 3 kb upstream of the Chr.6 MES promoter. In both Chr6 and Chr10 the binding sites overlap with *ESAG1*, whereas in both Chr11 and Chr5 the upstream binding sites overlap with *ESAG9*.

In summary, not all MES regions appeared to be bound by TbORC1/CDC6 in the same way, although similarities between the four putative MES can be drawn. A

repeat of the experiment should clarify whether the variability seen is purely experimental, or whether it is a stable trait of each locus.

5.6 Validation of TbORC1/CDC6 binding by quantitative real-time PCR

Although the distinctive pattern of TbORC1/CDC6 binding, localised to SSRs in the chromosome cores, suggests that the microarray output is not artefactual or random, it is necessary to independently validate the results, and confirm that the observed peaks correspond with increased DNA abundance in the ChIP DNA relative to input DNA. In order to do this, a qPCR absolute quantification method was used for specific loci to determine the amount of ChIP DNA present in the samples sent for microarray relative to input DNA. For validation of a divergent SSR in the core of Chr1, a predicted TbORC1/CDC6 binding site at position 281401-284218 was chosen. This had a high confidence value of binding, with an FDR score below 0.05. Primers were designed using the Primer Express software (Applied Biosystems) for four loci around the binding site: primer1 (CT_OL71/CT_OL72) was located upstream of the binding at position 278690-279050, primer2 and primer3 (CT_OL73/CT_OL74 and CT_OL75/CT_OL76) were located within the breadth of the binding site at positions 281498-281930 and 282686-283145, respectively, and primer4 (CT_OL77/CT_OL78) was located downstream of the binding site at position 292478-293045 (Figure 5-13). For each primer pair, a qPCR master mix was made according to the following recipe: 12.5 µl of SYBR mix (provided in kit), 1.0 µl of each primer (300 nM stock), 9.5 µl of dH_2O, and 1.0 µl template DNA (concentrations are described below) or dH_2O for a non-template control. The input and chIP DNA samples used for the microarray were used as template DNA for the qPCR. An ABI Prism 7500 thermocycler was used for the qPCR, as described in manufacturer's manual. For each primer pair, a standard curve was generated at five DNA concentrations (2 ng, 1 ng, 0.5 ng, 0.25 ng, 0.125 ng, 0 ng) using input DNA as the template. The chIP DNA was used as the test sample at two concentrations (1 ng and 0.5 ng) and assayed in the same run as the standards.

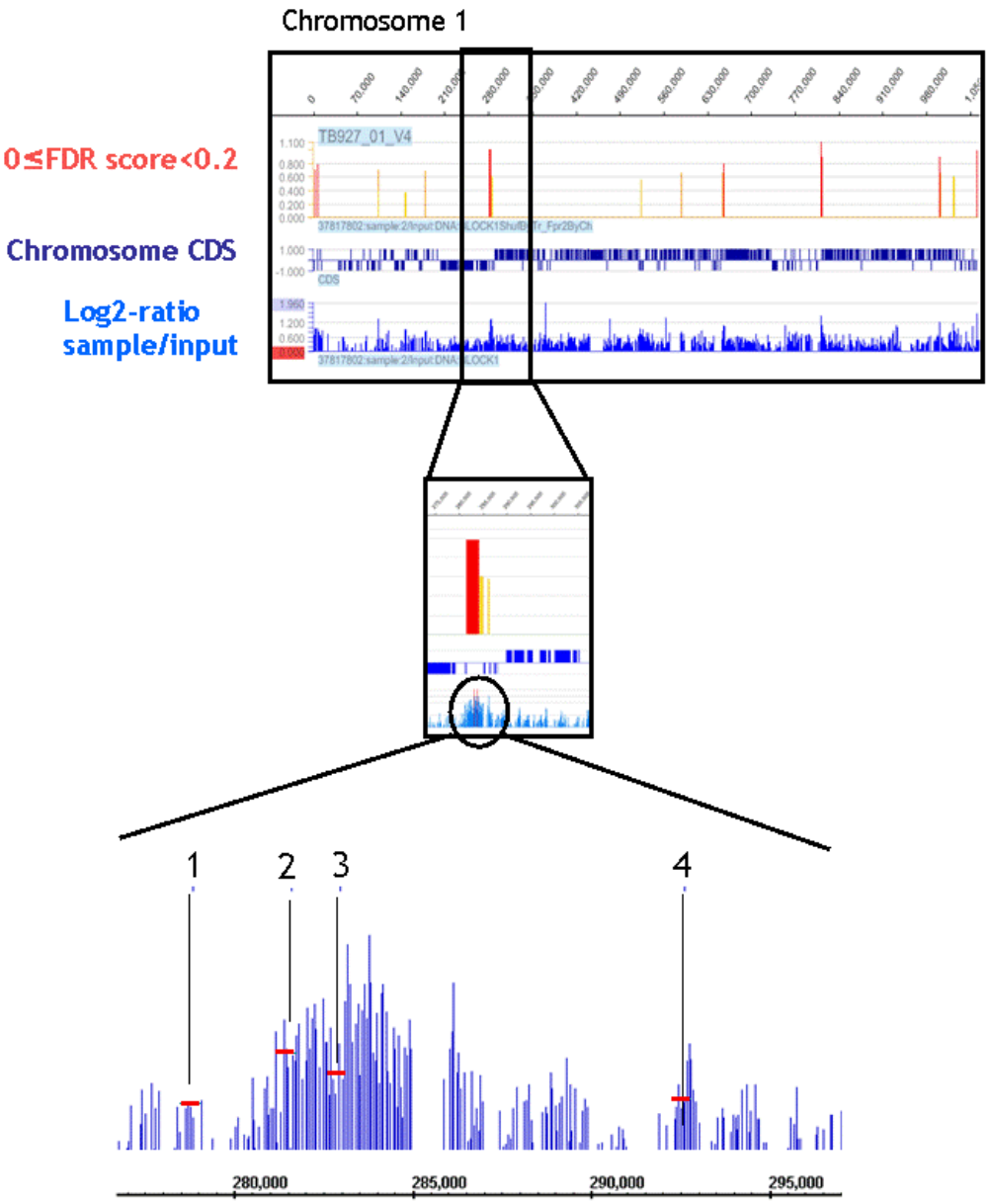


Figure 5-13 – Primer positions for validation of a TbORC1/CDC6 binding site on chromosome 1 by qPCR

Top panel: Signalmap output profile showing TbORC1/CDC6 binding sites on chromosome 1 as has been previously shown in Figure 5-5; Middle panel: an enlargement of a selected binding site on chromosome 1; Bottom panel: enlargement of the representative raw log2-ratio data for the selected binding site, primers used for quantitative PCR are indicated by horizontal red bars labelled 1, 2, 3, and 4 to indicated the respective positions of primer1, primer2, primer3, and primer4 within the region.

During the qPCR set up, a dissociation step was included at the end of the reaction to generate a dissociation curve for each primer pair in order to assess the homogeneity of the PCR products and the presence or absence of primer dimers, and hence the specificity of the reaction. The dissociation curve showed a single uniform peak at 72.5 ˚C for all four primers, confirming that all primers

were specific for the chosen loci and there were no secondary products present (Figure 5-14).

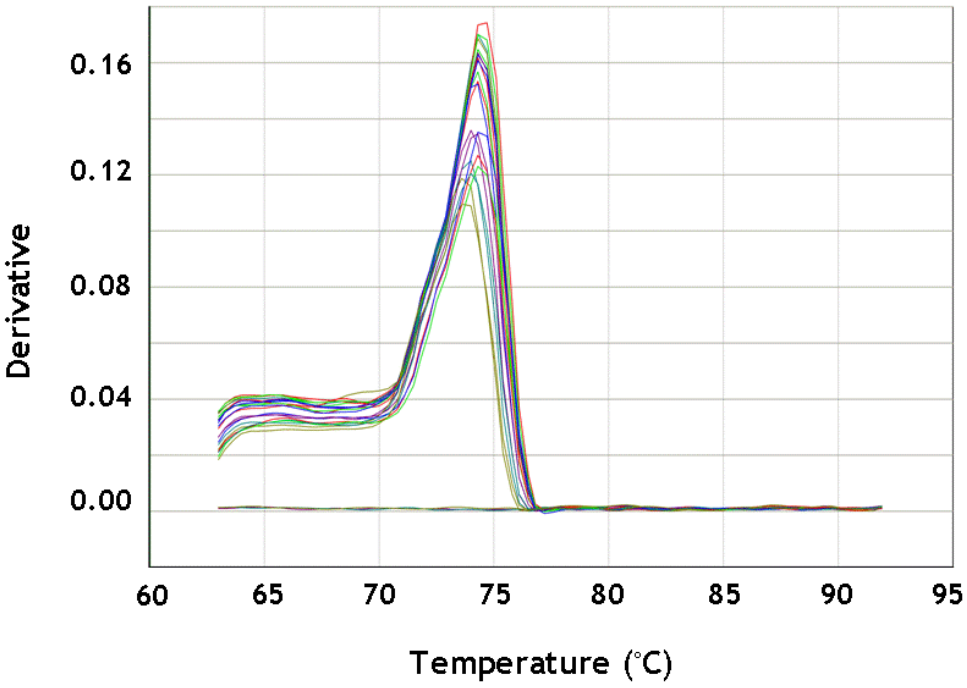


Figure 5-14 – Dissociation curves for the amplification product of qPCR primers

The y-axis (derivative) denotes the change in fluorescence, that is, the rate at which amplification products are denaturing at a given temperature, shown on the x-axis. Colours of lines represent different detectors for each of the four primers used (explained in text); the horizontal line at y = o represents a derivative for distilled water used as a non-template control (lacking any DNA substrate).

After the run, the PCR threshold values (C_T values) generated for input DNA (standard) were used to generate a standard curve. All four standard curves for each primer pair had a correlation coefficient (R^2 value) of 0.99. Using the standard curve for each primer pair, the C_T values of both the 1 ng and 0.5 ng ChIP DNA were interpolated to determine the absolute amount of DNA present at the four loci.

The quantification of ChIP DNA relative to input DNA showed that for primer2, located within the breadth of the predicted TbORC1/CDC6 binding site, qPCR of the immunoprecipitated material was five times more efficient for both the 1 ng and 0.5 ng samples compared with primer1 and primer4, both of which were located outside of the putative binding site (Figure 5-13). Although primer3 is also located within the breadth of the chIP-chip filtered peak for TbORC1/CDC6,

the unfiltered, scaled log2 ratio data show that there is no enrichment of the ChIP DNA relative to input DNA at the specific 60 bp sequence PCR-amplified in the qPCR. This result verifies the location of a TbORC1/CDC6 binding site at this locus.

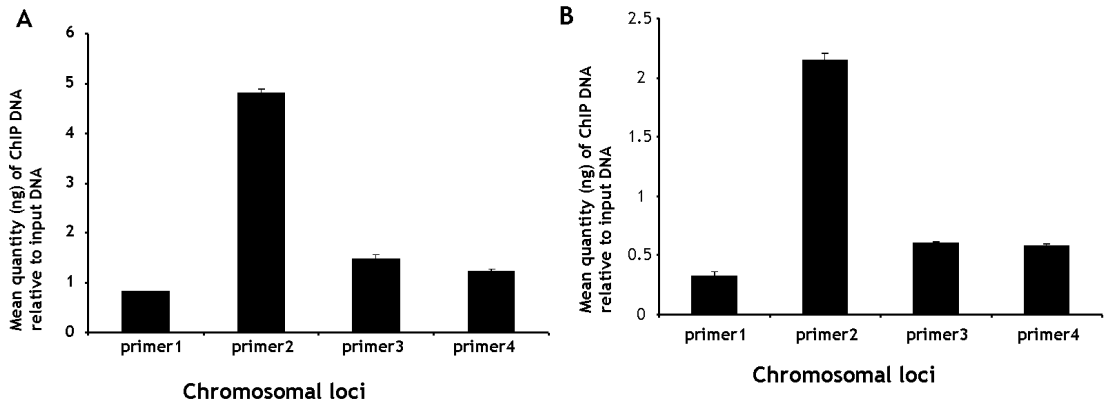


Figure 5-15 – ChIP-qPCR validation of TbORC1/CDC6 binding on a strand switch region of Chromosome 1

The mean absolute amount of 1 ng (A) or 0.5 ng (B) of ChIP DNA was compared to total input DNA by qPCR for specific loci indicated by primer1, primer2, primer3 and primer4 (see text for details of location of primers) on a divergent strand switch region of Chr1. Error bars indicate standard deviation of three independent experiments.

5.7 Discussion

5.7.1 TbORC1/CDC6 binding sites and replication origins in T. brucei

In the last decade, the identification of replication origins in eukaryotes (for example in *S. pombe* and *S. cerevisiae*) by genome-wide approaches, such as using chIP coupled with DNA microarrays, has become more popular (Hayashi *et al*. 2007; Heichinger *et al*. 2006; Xu *et al*. 2006; Wyrick *et al*. 2001). Many of these studies, though not all [e.g. (Heichinger *et al*. 2006)], have exploited the concept that the localisation of binding sites of pre-RC components such as ORC and MCM proteins would facilitate the identification of origins of DNA replication. Using ChIP-chip of ORC and MCM proteins in *S. cerevisiae* cells, Wyrick and colleagues (2001) identified 531 and 443 sites in the genome that bound ORC and MCM proteins, respectively, with a confidence level for binding of $p \leq 0.025$ (Wyrick *et al*. 2001). Of these loci ~80 % were demonstrated to be

origins of replication, by comparing the ORC- and MCM- binding sites with positions which had previously been identified to possess ARS activity. In similar work by Hayashi *et al*. (2006), Orc1-chIP and Mcm6-chIP in G_1–arrested *S. pombe* cells yielded a total of 460 sites that were occupied by both proteins (Hayashi *et al*. 2007). Using BrdU incorporation coupled with ChIP-chip, the authors showed that 307 (67 %) of these binding sites fired early in S-phase while 153 (33 %) of them fired late or were inefficient origins, although these loci co-localised with pre-RC proteins (Hayashi *et al*. 2007). The studies described above, as well as others (Heichinger *et al*. 2006), have concluded that the spatial distribution of pre-RC proteins correlates with origin usage. However, their eventual selection for firing DNA replication is variable. Nevertheless, binding of the pre-RC complex is necessary for any sequence to have the potential to function as an origin of replication.

Using the same reasoning, we have used ChIP-chip to map the binding sites for the single ORC-related protein that could be identified bioinformatically in *T. brucei*: TbORC1/CDC6. At a stringent confidence interval for binding ($p \leq 0.05$) of TbORC1/CDC6 with each DNA probe on our array, we identified 242 binding sites spanning Chrs 1-10, and at a modest confidence interval ($p < 0.2$) we identified 685 binding sites. A direct question that arises is what percentage of these binding sites are DNA replication origins in the *T. brucei* megabase chromosomes? At present, we are not able to answer this. A major obstacle has been providing experimental validation of these putative binding sites as origins of DNA replication in *T. brucei*. Two-dimensional agarose gel electrophoresis (Brewer & Fangman 1987), Short-Nascent strand DNA purification (Cadoret *et al*. 2008) coupled with microarray or Solexa Sequencing and marker frequency analysis (Lundgren *et al*. 2004) have all been attempted, without success to date (L. Marcello, Glasgow, and S.D. Bell Lab, Oxford). We are therefore limited in this discussion to the confines of available chIP-chip data, and therefore will simply to refer to the TbORC1/CDC6-associated DNA sequence loci identified here as binding sites, but will consider their potential to act as origins of DNA replication.

In *S. pombe*, Orc1 and MCM6 binding sites are observed with an average separation of 26.7 kb within the central core of the chromosomes, and are enriched at subtelomeres (Hayashi *et al*. 2007), except in Chr. 3. Using BrdU

incorporation coupled with ChIP-chip suggests that replication firing occurs differentially, with pre-RC sites in the central core functioning as early firing origins in S-phase compared with the subtelomere pre-RC sites, which are late-firing (Hayashi *et al*. 2007). The higher density of the subtelomere sites may then be needed to ensure efficient replication at the ends of chromosomes. This distribution of pre-RC sites throughout most of the *S. pombe* chromosomes appears similar to what we observed here for the distribution of TbORC1/CDC6 binding sites in the megabase chromosomes of *T. brucei*. Thus, all TbORC1/CDC6 binding sites might be replication origins, although we have to demonstrate this assertion experimentally. However, it is not clear whether the definition of the chromosome core and subtelomere in *S. pombe* is comparable with the clear distinction between the diploid 'housekeeping' core and aneuploid, *VSG*-rich subtelomere in *T. brucei*. Indeed, it is not clear if the relative enrichment we see at the *T. brucei* subtelomeres is comparable, as it is not quantified in *S. pombe* (Hayashi *et al*. 2007). TbORC1/CDC6 binding sites observed in the *T. brucei* megabase chromosome "core" shower greater dispersion (~110 kb between sites) compared with the average separation of 26.7 kb in *S. pombe* chromosome cores (Hayashi *et al*. 2007). Any such distinctions might reflect differences in TbORC1/CDC6 function relative to Orc1 in *S. pombe*, or they may simply reflect differences in the structural organisation of protein coding genes in the two organisms (polycistronic for *T. brucei* and monocistronic for *S. pombe*). If the latter is true, this may indicate that *T. brucei* nuclear DNA replication initiation is consistent with findings in other organisms that suggest DNA replication origin distribution and organisation of gene transcription are inherently linked (see below).

As mentioned above, no consensus sequence elements have emerged in the characterisation of *T. brucei* sequences that are able to confer extrachromosomal stability to circular plasmids or artificial linear molecules (see Section 1.8). Attempts to identify consensus elements for TbORC1/CDC6 binding, through comparing the sequences of chIP-chip-derived interaction loci, have not to date been successful (L. Marcello, pers. comm.). Though AT-rich regions in SSRs that separate directional gene clusters have been shown to confer mitotic stability in *T. brucei*, these have been implicated as having centromeric functions (Obado *et al*. 2007), are GC-rich in *T. cruzi* (Obado *et al*. 2005) and,

puzzlingly, appear not to confer chromosome stability in a close relative, *Leishmania* (Dubessay *et al*. 2002). For *T. brucei*, more importantly, it appears that each chromosome possesses a single centromere at an SSR (Obado *et al*. 2007), which contrasts with the binding of TbORC1/CDC6 we see to most SSRs (~50 % of sites within the chromosome cores). Thus, though centromeres and TbORC1/CDC6 binding sites may co-localise in a small number of loci, this is a minority of TbORC1/CDC6 sites. Furthermore, finer mapping of both TbORC1/CDC6 binding and the elements that dictate centromere activity is needed to determine if they share sequence features. Nevertheless, it remains possible that the unique architecture of kinetoplastid genomes has constrained DNA replication, centromeres and transcription initiation (see below) to localise to SSRs and, perhaps, to share functional determinants.

TbORC1/CDC6 binding sites and transcriptional regulation

First reported in human cells for the *β-globin* locus, the link between DNA replication and gene transcription is becoming a widespread observation, having being reported for a number of other organisms including *Drosophila* (Sequeira-Mendes *et al*. 2009), *Xenopus* (Danis *et al*. 2004) and CHO cells (Saha *et al*. 2004). Upstream of the *β-globin* origin of replication, a *β-globin* Locus Control Region (LCR) located proximal to the *β-globin* promoter was discovered and deletion of the LCR resulted in loss of promoter activity, as well as ablation of origin activity, suggesting that this origin locus was linked with both replication and transcription (Aladjem *et al*. 1995). In *Xenopus* and *Drosophila* embryos, prior to the onset of gene transcription, replication origins are more randomly dispersed in the genome, but at the onset of transcription become localised to intergenic regions (Sasaki *et al*. 1999; Hyrien *et al*. 1995). In HeLa and mouse embryonic stem cells, the co-localisation of replication origins with CpG islands that are in close proximity with promoters is also an emerging paradigm, which suggests a global coordination of gene transcription and replication timing (Sequeira-Mendes *et al*. 2009; Cadoret *et al*. 2008).

In *T. brucei,* a similar phenomenon had been hinted at, with the identification of a PMS that confers stability to a plasmid or an artificial linear molecule and depends on the presence of the PARP promoter in close proximity (Patnaik *et al*. 1996; Lee *et al*. 1995; Patnaik *et al*. 1994). However, gene transcription in *T.*

brucei and related kinetoplastids is regulated differently from other eukaryotes (Clayton 2002). In *T. brucei* RNA PolII transcribes most protein coding genes (with the exception of *VSGs*) as large, polycistronic units, or directional gene clusters (Clayton 2002). Furthermore, the available evidence suggests that transcription initiation is largely unregulated and localises upstream of the first gene in the transcription unit, though promoters have not been identified. Although our microarray-based approach shows no TbORC1/CDC6 binding sites at the PARP promoter (as this was masked because it is present in four copies), the predominant co-localisation of TbORC1/CDC6 binding sites with divergent SSRs, rather than with convergent SSRs, is potentially very interesting. Divergent SSRs have been reported to be RNA PolII transcription start sites in *T. brucei* (Siegel *et al.* 2009), whereas convergent SSRs would be regions where two transcripts meet. It is therefore possible that TbORC1/CDC6 binds to RNA PolII transcription initiation sites. What aspect of this might be responsible for the binding is unclear; it may be that previously unrecognised DNA sequences are shared as promoters and as replication origins, or it might be that TbORC1/CDC6 is recruited by RNA PolII-associated factors. Alternatively, TbORC1/CDC6 binding might reflect chromatin architecture in these SSRs: it may simply be that the putative open chromatin allows TbORC1/CDC6 access, or specific chromatin modifications might recruit the protein. To date, acetylated histone H4 (residue K10), trimethylated histone H3 (residue K4), variants of histones H2A and H2B (named H2AZ and H2BV), and the hypermodified base J have been localised, predominantly if not exclusively in some cases, at divergent SSRs (Cliffe *et al*. 2010; Wright, Siegel, & Cross 2010; Siegel *et al.* 2009). Consistent with any of these suggestions, the localisation of TbORC1/CDC6 binding sites within, and equally distributed between, forward and reverse coding strands of polycistronic units appeared to colocalise with these chromatin modifications (fine mapping of this overlap is awaiting analysis). These sites appear then to be transcription start sites with these units, and would therefore share features with the divergent SSRs, and do not undermine the putative link between transcription initiation and origin specification.

Verification of the suggestion that TbORC1/CDC6 binds RNA PolIII transcription start sites will require further work. Most notably, it will be important to fine map TbORC1/CDC6 binding sites within divergent SSRs and ask if these overlap,

or are neighbours, with features around the RNA polII transcription start sites, since some of the SSRs are very large. It will also be important to experimentally test if the TbORC1/CDC6 localisation we observed to convergent SSRs is real. We also do not know if TbORC1/CDC6 localisation is specific to RNA PolII transcription, or may be related to transcription in general: because the rDNA arrays are repeated in *T. brucei*, they are not represented in the microarray, and we cannot therefore evaluate if TbORC1/CDC6 binds these RNA Pol I units. Though MES and BES are also RNA PolI transcribed, these are inactive in the procyclic form cells analysed here. Finally, it is important to consider the implications of high density localisation of TbORC1/CDC to the *VSG* arrays, which are thought to be constitutively transcriptionally silent.

5.7.2 TbORC1/CDC6 binding sites and VSG architecture

As mentioned earlier, in megabase chromosomes in *T. brucei VSGs* are located either in *VSG* arrays, which are located within the telomere-distal portion of the chromosome (the subtelomere), or in ESs, where they are located in close proximity to telomeric repeats (McCulloch 2004). Unlike array *VSGs,* telomeric *VSGs* can be transcribed if they are present within an active BES or MES (Barry & McCulloch 2001). It is thought that the 70 bp repeats in the arrays enhance recombination by gene conversion to an active ES. This event is crucial for trypanosome survival in the bloodstream of the mammalian host, where the parasite is exposed to immune attack (Barry & McCulloch 2001). We have shown that TbORC1/CDC6 binding sites are more densely clustered in the *VSG* arrays than in the chromosome cores and that within these arrays TbORC1/CDC6 preferentially localises to the 5' *VSG* component. In addition, we have attempted to map TbORC1/CDC6 binding in the ES that are available from the sequencing of the TREU 927 genome to date. Here, we see evidence for binding, though not with a clear pattern. We did not find evidence that TbORC1/CDC6 binds to either promoter [MES and BES promoters are diverged in sequence; (Ginger *et al.* 2002)], but did find localisation upstream and downstream of the promoters. Where *VSG*s were found in the ESs, a definite conclusion about preferential localisation of TbORC1/CDC6 with the 5' component cannot be reached as most of the telomeric ends of chromosomes are being assembled.

What is the significance of TbORC1/CDC6 subtelomeric localisation? Given the lack of transcription in these loci, it is possible that subtelomere localisation shares features with the SSRs, which are untranscribed regions between transcription units. However, there is no evidence for any transcription initiation in the *VSG* arrays, unlike the SSRs. If TbORC1/CDC6 recruitment to the SSRs was due to transcription initiation, a distinct mode of recruitment to the *VSG* arrays must be invoked, perhaps implying that these sites do not represent replication origins. This requires experimental testing. It also poses a problem: if replication origins are defined by transcription start sites, is replication from origins adjacent to the subtelomeres sufficient to duplicate these loci, which can represent as much as 75 % of a chromosome (Callejas *et al*. 2006)? One other hypothesis, consistent with the activation of *MVSG* expression after TbORC1/CDC6 RNAi, is that TbORC1/CDC6 binding in the subtelomeres contributes to transcriptional repression of *VSG* expression, via the formation of heterochromatin (see Section 3.7 for experimental data to support this). This would be consistent with observed roles for ORC components in establishing silent chromatin in other organisms. In *S. cerevisiae*, ORC contributes to silencing of the mating type loci (Sasaki & Gilbert 2007) and stress-induced repression of further genes (Burhans *et al*. 2006; Ramachandran *et al*. 2006). In *P. falciparum*, Orc1 has been shown to associate with Sir2 and to bind telomeres, and may have a role in telomeric silencing, including of *var* genes (Mancio-Silva *et al*. 2008). If this is correct, it remains to be determined how TbORC1/CDC6 acts in this way, and how it associates with chromatin. A final hypothesis is that TbORC1/CDC6 binding in the *VSG* arrays might regulate the recombination of *VSGs* into the ES. This remains to be tested, but could be related to chromatin-mediated silencing in suppressing such recombination, or it may promote recombination via replication.

5.8 Highlight of major findings

Here, using chIP coupled with a high resolution tiling array we have determined the genomic locations of TbORC1/CDC6 binding on ten of the eleven *T. brucei* megabase chromosomes. In total, up to 685 sites were identified; 278 loci were located within the core of chromosomes, 330 in *VSG* arrays and 81 in subtelomeres. We postulate that at least some of these binding sites are likely

to function as origins of replication in *T. brucei* megabase chromosomes. Within the chromosome cores, TbORC1/CDC6 binding sites show a significant overlap with boundaries of polycistronic transcription units, at putative RNA PolII transcription start sites; this may be consistent with a link between DNA replication and gene transcription in *T. brucei*. A significant proportion of TbORC1/CDC6 binding sites are present in *VSG* arrays, consistent with a previous observation of the involvement of TbORC1/CDC6 in *VSG* expression control, and thus antigenic variation, in *T. brucei*.

6 Perspectives

The raison d'être for studying the process of nuclear DNA replication in *T. brucei* is the fundamental importance of the process. Beyond this, two parasite-specific features also prompted the investigations described in this thesis: (1) the mechanistic architecture of the process of initiation of nuclear DNA replication in *T. brucei* appeared to be simplified; and (2) the central and essential nature of the process of DNA replication has allowed the development of new drugs in other organisms (such as Ganciclovir and Ciprofloxacin), which might suggest that the same could be true for this parasite.

Bioinformatic, biochemical and genetic studies of yeasts (*S. cerevisiae* and *S. pombe*) and metazoan cells have implicated at least fourteen proteins (Orc1-Orc6, Cdc6, Cdt1, Mcm2-Mcm7) as constituting the core machinery required to define an initiation event that is necessary to replicate their genetic material; DNA (Dutta & Bell 2006). Apart from these protein initiators, much effort over the years has also been dedicated to the identification of *cis*-acting elements that the main initiator protein complex, ORC, binds to, otherwise known as origins of replication. From this picture, two fundamental questions can be asked with respect to nuclear DNA replication initiation in *T. brucei*: (a) how conserved are the parasite initiator factors that define origin of replication sequences relative to other eukaryotes; and (b) what defines the sequence elements that constitute origins of DNA replication? In this work, we have begun to shed light on both of these questions in *T. brucei*.

6.1 Interaction of TbORC1/CDC6 with other proteins

Based on analyses of small subunit ribosomal RNA genes the African trypanosome *T. brucei* has previously been described as a “primitive organism” or “an early divergent”, single-celled flagellated parasitic protozoa (Sogin *et al.* 1989). This description is highly compatible with several lines of evidence, which have revealed numerous striking and fundamentally distinct biological phenomena from yeasts, humans and plants. Examples of these include trans-splicing (Lee & Van der Ploeg 1997) and RNA editing (Horton & Landweber 2002). However, recent phylogenetic analyses based on morphological and comparative molecular

studies, the latter emerging from available whole genome information, suggest that *T. brucei* and associated members of the order Kinetoplastida may not be primitive after all (Dacks *et al.* 2008). Failure of BLAST searches of the *T. brucei* genome to identify clear homologues of Orcs2-6 and Cdt1 have been interpreted by at least three publications (Fritz-Laylin *et al.* 2010; Cavalier-Smith 2010; Godoy *et al.* 2009) as indicating that these genes are absent, leading to the hypothesis that DNA replication initiation in *T. brucei* might be another aspect of biology that is divergent. Furthermore, Cavalier-Smith has used this hypothesis as a premise, among other considerations, to re-classify the phylum Euglenozoa by segregating it from infrakingdom Excavata (Cavalier-Smith 2010). The reasoning behind this is that Cavalier-Smith suggests these data indicate that the root of the eukaryotic kingdom is between Euglenozoa and all other eukaryotes. Prior to the publication of the genome of *T. brucei* the process of nuclear DNA replication initiation in this parasite, and indeed in all kinetoplastids, remained largely unexplored, in contrast to the advanced description of mitochondrial genome replication (Liu *et al.* 2005).

In this work, we have shown by RNAi that TbORC1/CDC6 is an essential protein in both BSF and PCF *T. brucei*. Furthermore, we show that it is involved in cell cycle regulation in BSF; it is involved in nuclear DNA replication in PCF; it regulates the expression of *MVSGs* (and perhaps *BVSG*s); and it interacts with at least one other ORC-like trypanosome protein. We provide evidence that this protein, which was previously uncharacterised and unidentifiable by bioinformatic searches alone, is the *T. brucei* orthologue of Orc4, and therefore name it TbORC4 (Figure 6-1).

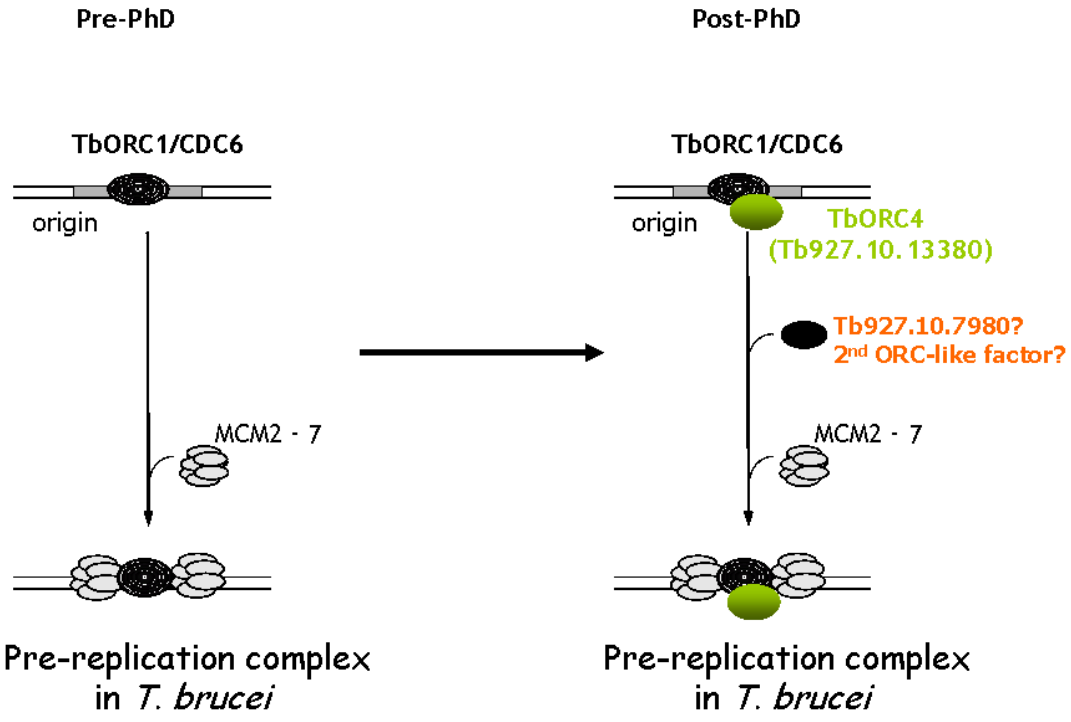


Figure 6-1 – Model for *T. brucei* pre-replication complex

The MCM2-7 complex is loaded onto origin by TbORC1/CDC6 alone (Pre-PhD); and two evolutionarily ORC related proteins (TbORC1/CDC6, TbORC4) and possibly Tb927.10.7980 load the MCM2-7 complex (Post PhD).

Godoy *et al* (2009) showed that TbORC1/CDC6 is able to complement *S. cerevisiae* Cdc6 temperature-sensitive mutants, but not Orc1 mutants (Godoy *et al*. 2009). Having identified two more proteins (TbORC4 and Tb927.10.7980) that putatively interact with TbORC1/CDC6, and have the same RNAi phenotype in BSF as TbORC1/CDC6, the question now is: are any of these proteins able to complement yeast Orc1? In other words, is it conceivable that the TbORC1/CDC6 protein that was considered the sole *T. brucei* ORC component is actually functionally analogous (even orthologous) with Cdc6, and might one of the other factors, despite being highly diverged in sequence, provide Orc1 functions?. Complementation assays using yeast temperature-sensitive mutants might allow this to be addressed. Follow-up experiments will be needed to test if the predicted AAA^+ folds present in these proteins confer DNA-binding characteristics *in vitro*, or whether their role is in determining the structure and function of *T.brucei* ORC. Expression and purification of all ORC-like proteins and DNA-binding assays, such as electrophoretic mobility shift assays (EMSA), could test this *in vitro*. Attempts were made early in this PhD to express and purify

TbORC1/CDC6 from *E. coli*, but the protein proved to be very insoluble (data not shown). Perhaps, then, its stability is dependent on co-expression with the further putative ORC-like factors identified here.

To understand the architecture of the putative *T. brucei* ORC, coIP will be needed to ask if TbORC4 and Tb927.10.7980 interact with each other, and it will be necessary to ask if IP of these proteins reveals yet further ORC components. One very distantly related ORC-like factor (Tb09.160.3120) was not examined here, and this should be analysed in the same way. Co-localisation experiments *in vivo* to confirm nuclear localisation of the proteins, and to test how they interact with TbORC1/CDC6 throughout the cell cycle, should be done to support the co-IP experiments already performed. Ultimately, to test if these factors do act in origin designation, chromatin-IP experiments coupled with either qPCR or hybridisation to the *T. brucei* microarray needs to be done to analyse genome localisation. It remains possible, of course, that these factors do not act together as an ORC, but that they associate with TbORC1/CDC6 at specific times in the cell cycle, or at specific genomic locations, to modulate the function of TbORC1/CDC6. If this were the case, they may provide other *T. brucei* functions (e.g. they might act in *VSG* regulation), and TbORC1/CDC6 may then be an archaeal-like origin designator.

One drawback in the co-IP experiments performed here has been to distinguish between interactions that occurred as a result of binding with the Myc-tag on TbORC1/CDC6 and interactions that occurred as a result of binding with TbORC1/CDC6 itself. With hindsight, we would have done a Myc-tag only pull-down experiment to filter any proteins that are common in both IP experiments. Since we did not do a Myc-IP mock experiment, a solution will be to exchange the tags in the proteins in co-IP experiments. For example, in the co-IP of “TbORC4”-HA and TbORC1/CDC6-Myc, an exchange of tags to have “TbORC4”-myc and TbORC1/CDC6-HA will confirm that indeed the interaction we have shown does not occur via the myc tag but via TbORC1/CDC6 and “TbORC4”. This procedure should be repeated for all interactors found in the IP-mass spectrometry approach.

6.2 Interaction of TbORC1/CDC6 with origins

In *S. cerevisiae* (Wyrick *et al*. 2001), *S. pombe* (Hayashi *et al*. 2007) and higher eukaryotes Orc1 binds to origins of DNA replication and has been used as bait to map genome-wide locations of origins. Here, we have used ChIP-chip to create a near genome-wide map of TbORC1/CDC6 binding sites. The localisation of these binding sites fit with the nuclear architecture of *T. brucei* chromosomes and appears reminiscent of a typical eukaryotic organisation (Hayashi *et al*. 2007; Heichinger *et al*. 2006). In particular, the localisation of the protein to transcription start sites (promoters have not been defined) within the core appears to fit with patterns described elsewhere. Higher density localisation to to subtelomeric and telomeric regions has also been observed in other eukaryotes.

To complement our ChIP-chip TbORC1/CDC6 localisation data at telomeric and subtelomeric loci it would have been desirable to perform mobility shift and supershift assays, as well chIP-qPCR and chIP-blot experiments. Interactions of TbORC1/CDC6 and TbRAP1 (Yang *et al*. 2009), TbISW1 (Hughes *et al*. 2007) and TbTRF2 would strongly support a role of TbORC1/CDC6 in transcriptional activation at BES and MES in trypanosomes. Co-localisation experiments of tagged TbORC1/CDC6 with histone H4k10ac, histone variants H2BV and H2AZ will strongly corroborate our hypothesis that DNA replication in *T. brucei* is linked to gene transcription.

At the moment, we and our collaborators (H. Farr and S.D. Bell; Oxford) have not been able to identify origins of DNA replication by a variety of approaches: two-dimensional agarose gel electrophoresis, Marker Frequency analysis, and Short Nascent Strand synthesis coupled with Solexa sequencing. Currently, attempts to use S-phase synchronised cells are on-going and might enrich for origins. We are also in the process of characterising TbORC1/CDC6 binding sites to see if specific DNA sequence motifs are identifiable. Once this is done, if we are able to identify conserved motifs, variable lengths of DNA including these motifs could be cloned and targeted to specific locations on chromosomes, or to articifial plasmids. Using ChIP, we could then test if TbORC1/CDC6 recognises and binds to these sequences. Using 2D gel electrophoresis, we could also ask if

origins are fired from these loci. Although these might represent an artificial chromosomal context, as other factors such as chromatin organisation are likely to affect origin preference, it will confirm that the identified TbORC1/CDC6 binding sites are even more likely to function as origins in their natural context and those specific motifs are likely to define origins in this organism.

6.3 Closing Remarks

T. brucei remains a pathogen of major medical, veterinary and economic impact. In common with other kinetoplastid parasites, safe and effective drugs against *T. brucei* are still needed, and resistance to the existing drugs is increasing (Barrett 2006). DNA replication studies have been instrumental in our understanding of the basis of many diseases; for example, cancer can result from non-regulated DNA replication. In fact, several drugs (such as Ganciclovir and Ciprofloxacin) have been developed targeting aspects of the DNA replication mechanism in a number of organisms, including bacteria and viruses. This thesis has provided a mechanistic insight into nuclear DNA replication in *T. brucei*, an area in which very little work had been done and it sheds a novel perspective on eukaryotic DNA replication strategies. The analysis we have undertaken has allowed the identification of potentially diverged factors of the trypanosomatid DNA replication apparatus (for example, TbORC4 and, perhaps, Tb927.10.7980), which if further characterised may differ sufficiently from that of the host (where significant work has been done) to represent targets for the future generation of novel therapeutic agents.

7 Appendices

Appendix 1 MCM protein IDs used for BLAST of the TriTryp database

MCM2

>gi|3912|emb|CAA37615.1| MCM2 [Saccharomyces cerevisiae]
>gi|602896|emb|CAA54503.1| MCM2 [Saccharomyces cerevisiae]
>gi|536021|emb|CAA84842.1| MCM2 [Saccharomyces cerevisiae]
>gi|453237|emb|CAA52635.1| MCM2 [Saccharomyces cerevisiae]
>gi|20977583|gb|AAM28219.1| DNA replication licensing factor [Danio rerio]
>gi|157437450|gb|EDO81660.1| MCM2 [Giardia lamblia ATCC 50803]
>gi|253741552|gb|EES98420.1| MCM2 [Giardia intestinalis ATCC 50581]
>gi|159118230|ref|XP_001709334.1| MCM2 [Giardia lamblia ATCC 50803]
>gi|308158142|gb|EFO60951.1| MCM2 [Giardia lamblia P15]
>gi|3642638|gb|AAC36510.1| MCM2 [Mus musculus]
>gi|159765762|gb|ABW97990.1| mcm2 [Hemiselmis andersenii]
>gi|160331215|ref|XP_001712315.1| mcm2 [Hemiselmis andersenii]
>gi|7299005|gb|AAF54207.1| minichromosome maintenance 2 [Drosophila melanogaster]
>gi|17137132|ref|NP_477121.1| minichromosome maintenance 2 [Drosophila melanogaster]
>gi|296474646|gb|DAA16761.1| KIAA0030-like [Bos taurus]
>gi|296491164|gb|DAA33237.1| cyclin L1 [Bos taurus]
>gi|33356547|ref|NP_004517.2| DNA replication licensing factor MCM2 [Homo sapiens]
>gi|147898891|ref|NP_001080759.1| DNA replication licensing factor mcm2 [Xenopus laevis]
>gi|585465|sp|P29469.2|MCM2_YEAST RecName: Full=DNA replication licensing factor MCM2; AltName: Full=Minichromosome maintenance protein 2
>gi|172088119|ref|NP_032590.2| DNA replication licensing factor MCM2 [Mus musculus]

MCM3

>gi|21105417|gb|AAM34652.1|AF506208_1 DNA replication licensing factor [Danio rerio]
>gi|829621|gb|AAA80227.1| MCM3 [Xenopus laevis]
>gi|3953607|dbj|BAA34731.1| MCM3 [Drosophila melanogaster]
>gi|308162155|gb|EFO64566.1| MCM3 [Giardia lamblia P15]
>gi|157434045|gb|EDO78280.1| MCM3 [Giardia lamblia ATCC 50803]
>gi|253742964|gb|EES99567.1| MCM3 [Giardia intestinalis ATCC 50581]
>gi|159111445|ref|XP_001705954.1| MCM3 [Giardia lamblia ATCC 50803]
>gi|159766078|gb|ABW98304.1| mcm3 [Hemiselmis andersenii]
>gi|160331845|ref|XP_001712629.1| mcm3 [Hemiselmis andersenii]
>gi|7290573|gb|AAF46023.1| minichromosome maintenance 3 [Drosophila melanogaster]
>gi|24639835|ref|NP_511048.2| minichromosome maintenance 3 [Drosophila melanogaster]
>gi|296474394|gb|DAA16509.1| DNA replication licensing factor MCM3 [Bos taurus]
>gi|6631095|ref|NP_002379.2| DNA replication licensing factor MCM3 [Homo sapiens]
>gi|33859484|ref|NP_032589.1| DNA replication licensing factor MCM3 [Mus musculus]
>gi|19857543|sp|P25205.3|MCM3_HUMAN RecName: Full=DNA replication licensing factor MCM3; AltName: Full=DNA polymerase alpha holoenzyme-associated protein P1; AltName: Full=P1-MCM3; AltName: Full=RLF subunit beta; AltName: Full=p102

>gi|2506834|sp|P25206.2|MCM3_MOUSE RecName: Full=DNA replication licensing factor MCM3; AltName: Full=DNA polymerase alpha holoenzyme-associated protein P1; AltName: Full=P1-MCM3
>gi|126822|sp|P24279.1|MCM3_YEAST RecName: Full=DNA replication licensing factor MCM3; AltName: Full=Minichromosome maintenance protein 3
>gi|75334009|sp|Q9FL33.1|MCM3_ARATH RecName: Full=DNA replication licensing factor MCM3 homolog; AltName: Full=Minichromosome maintenance protein 3 homolog
>gi|164448574|ref|NP_001013604.2| DNA replication licensing factor MCM3 [Bos taurus]
>gi|57530231|ref|NP_001006421.1| DNA replication licensing factor MCM3 [Gallus gallus]

MCM4

>gi|259130476|gb|ACV95639.1| MCM4 [Leishmania donovani]
>gi|2754697|gb|AAC52018.1| MCM4 [Homo sapiens]
>gi|11761139|dbj|BAA73964.1| MCM4 [Gallus gallus]
>gi|11862812|dbj|BAB19264.1| MCM4 [Xenopus laevis]
>gi|308159481|gb|EFO62010.1| MCM4 [Giardia lamblia P15]
>gi|157436985|gb|EDO81197.1| MCM4 [Giardia lamblia ATCC 50803]
>gi|253745243|gb|EET01291.1| MCM4 [Giardia intestinalis ATCC 50581]
>gi|159117302|ref|XP_001708871.1| MCM4 [Giardia lamblia ATCC 50803]
>gi|159766065|gb|ABW98291.1| mcm4 [Hemiselmis andersenii]
>gi|160331819|ref|XP_001712616.1| mcm4 [Hemiselmis andersenii]
>gi|296480668|gb|DAA22783.1| minichromosome maintenance complex component 4 [Bos taurus]
>gi|7304207|gb|AAF59242.1| disc proliferation abnormal [Drosophila melanogaster]
>gi|17137242|ref|NP_477185.1| disc proliferation abnormal [Drosophila melanogaster]
>gi|33469919|ref|NP_005905.2| DNA replication licensing factor MCM4 [Homo sapiens]
>gi|33469917|ref|NP_877423.1| DNA replication licensing factor MCM4 [Homo sapiens]
>gi|255918149|ref|NP_032591.3| DNA replication licensing factor MCM4 [Mus musculus]
>gi|68571766|sp|P33991.5|MCM4_HUMAN RecName: Full=DNA replication licensing factor MCM4; AltName: Full=CDC21 homolog; AltName: Full=P1-CDC21
>gi|11559506|gb|AAG37988.1|AF083323_1 DNA replication licensing factor MCM4 [Plasmodium falciparum]
>gi|115495629|ref|NP_001068626.1| DNA replication licensing factor MCM4 [Bos taurus]
>gi|54020819|ref|NP_001005655.1| DNA replication licensing factor mcm4 [Xenopus (Silurana) tropicalis]

MCM5

>gi|47678565|emb|CAG30403.1| MCM5 [Homo sapiens]
>gi|109451950|emb|CAK54847.1| MCM5 [synthetic construct]
>gi|109451374|emb|CAK54548.1| MCM5 [synthetic construct]
>gi|308158211|gb|EFO60998.1| MCM5 [Giardia lamblia P15]
>gi|157434121|gb|EDO78355.1| MCM5 [Giardia lamblia ATCC 50803]
>gi|253748087|gb|EET02444.1| MCM5 [Giardia intestinalis ATCC 50581]
>gi|159111596|ref|XP_001706029.1| MCM5 [Giardia lamblia ATCC 50803]
>gi|159765671|gb|ABW97899.1| mcm5 [Hemiselmis andersenii]
>gi|160331033|ref|XP_001712224.1| mcm5 [Hemiselmis andersenii]
>gi|7299365|gb|AAF54557.1| minichromosome maintenance 5 [Drosophila melanogaster]
>gi|24645774|ref|NP_524308.2| minichromosome maintenance 5 [Drosophila melanogaster]
>gi|296487399|gb|DAA29512.1| DNA replication licensing factor MCM5 [Bos taurus]

>gi|169144979|emb|CAQ08830.1| minichromosome maintenance complex component 5 [Homo sapiens]
>gi|169144978|emb|CAQ08829.1| minichromosome maintenance complex component 5 [Homo sapiens]
>gi|23510448|ref|NP_006730.2| DNA replication licensing factor MCM5 [Homo sapiens]
>gi|46358340|ref|NP_848523.2| DNA replication licensing factor MCM5 [Danio rerio]
>gi|57525409|ref|NP_001006243.1| DNA replication licensing factor MCM5 [Gallus gallus]
>gi|284413774|ref|NP_001017327.2| DNA replication licensing factor mcm5 [Xenopus (Silurana) tropicalis]
>gi|22347793|gb|AAM95977.1| DNA replication licensing factor Mcm5 [Danio rerio]
>gi|116248539|sp|Q0V8B7.1|MCM5_BOVIN RecName: Full=DNA replication licensing factor MCM5

MCM6

>gi|29725889|gb|AAO89010.1| MCM6 [Saccharomyces cerevisiae]
>gi|161728849|dbj|BAF94254.1| Mcm6 [Rattus norvegicus]
>gi|161728828|dbj|BAF94234.1| Mcm6 [Rattus norvegicus]
>gi|3953609|dbj|BAA34732.1| MCM6 [Drosophila melanogaster]
>gi|308162402|gb|EFO64801.1| MCM6 [Giardia lamblia P15]
>gi|157434295|gb|EDO78527.1| MCM6 [Giardia lamblia ATCC 50803]
>gi|253744910|gb|EET01045.1| MCM6 [Giardia intestinalis ATCC 50581]
>gi|159111942|ref|XP_001706201.1| MCM6 [Giardia lamblia ATCC 50803]
>gi|159765901|gb|ABW98128.1| mcm6 [Hemiselmis andersenii]
>gi|160331492|ref|XP_001712453.1| mcm6 [Hemiselmis andersenii]
>gi|7290738|gb|AAF46184.1| minichromosome maintenance 6 [Drosophila melanogaster]
>gi|17530827|ref|NP_511065.1| minichromosome maintenance 6 [Drosophila melanogaster]
>gi|296490539|gb|DAA32652.1| DNA replication licensing factor MCM6 [Bos taurus]
>gi|2497824|sp|Q14566.1|MCM6_HUMAN RecName: Full=DNA replication licensing factor MCM6; AltName: Full=p105MCM
>gi|7427519|ref|NP_005906.2| DNA replication licensing factor MCM6 [Homo sapiens]
>gi|6678832|ref|NP_032593.1| DNA replication licensing factor MCM6 [Mus musculus]
>gi|75026271|sp|Q9V461.1|MCM6_DROME RecName: Full=DNA replication licensing factor Mcm6; Short=DmMCM6
>gi|114052981|ref|NP_001039699.1| DNA replication licensing factor MCM6 [Bos taurus]
>gi|57529699|ref|NP_001006527.1| DNA replication licensing factor MCM6 [Gallus gallus]
>gi|108860789|sp|Q2KIZ8.1|MCM6_BOVIN RecName: Full=DNA replication licensing factor MCM6

MCM7

>gi|54290089|dbj|BAD61056.1| MCM7 [Bombyx mori]
>gi|20977575|gb|AAM28215.1| DNA replication licensing factor [Danio rerio]
>gi|3953611|dbj|BAA34733.1| MCM7 [Drosophila melanogaster]
>gi|308160888|gb|EFO63355.1| MCM7 [Giardia lamblia P15]
>gi|157434691|gb|EDO78919.1| MCM7 [Giardia lamblia ATCC 50803]
>gi|253744222|gb|EET00456.1| MCM7 [Giardia intestinalis ATCC 50581]
>gi|159112730|ref|XP_001706593.1| MCM7 [Giardia lamblia ATCC 50803]
>gi|159765708|gb|ABW97936.1| mcm7 [Hemiselmis andersenii]
>gi|160331107|ref|XP_001712261.1| mcm7 [Hemiselmis andersenii]

>gi|115908135|ref|XP_001200231.1| PREDICTED: similar to DNA replication licensing factor; MCM7, partial [Strongylocentrotus purpuratus]
>gi|115740038|ref|XP_001180038.1| PREDICTED: similar to DNA replication licensing factor; MCM7, partial [Strongylocentrotus purpuratus]
>gi|7295030|gb|AAF50357.1| minichromosome maintenance 7 [Drosophila melanogaster]
>gi|17647617|ref|NP_523984.1| minichromosome maintenance 7 [Drosophila melanogaster]
>gi|296472979|gb|DAA15094.1| DNA replication licensing factor MCM7 [Bos taurus]
>gi|33469968|ref|NP_005907.3| DNA replication licensing factor MCM7 isoform 1 [Homo sapiens]
>gi|33469922|ref|NP_877577.1| DNA replication licensing factor MCM7 isoform 2 [Homo sapiens]
>gi|10242373|ref|NP_032594.1| DNA replication licensing factor MCM7 [Mus musculus]
>gi|20981696|sp|P33993.4|MCM7_HUMAN RecName: Full=DNA replication licensing factor MCM7; AltName: Full=CDC47 homolog; AltName: Full=P1.1-MCM3
>gi|12230233|sp|O75001.1|MCM7_SCHPO RecName: Full=DNA replication licensing factor mcm7; AltName: Full=Minichromosome maintenance protein 7
>gi|115529274|ref|NP_001020516.2| DNA replication licensing factor MCM7 [Bos taurus]

Appendix 2 Gene IDs for Orc4 subunits used to generate ORC4 alignment

>SceOrc4 (P54791)
>EcuOrc5? (NP_597618)
>HsaOrc4 (O43929)
>AthOrc4 (CAE01428)
>TanOrc4 TA12985 |||hypothetical protein|Theileria annulata|chr 02|||Manual
>CpaOrc4 cgd2_1550
>PfaOrc4 PF13_0189 |||hypothetical protein, conserved|Plasmodium falciparum|chr 13|SANGER||Manual
>DdiOrc4 DDB0168430 ||||Dictyostelium discoideum|chr 2|||Auto
>GlaOrc4 ctg02_3
>TthOrc4 51.m00235
>TvaOrc4 TVAG_365960
>DmeOrc4 gi|7291857|gb|AAF47276.1| origin recognition complex subunit 4 [Drosophila melanogaster]

Appendix 3 Gene IDs of ORC subunits used for generation of phylogenetic tree

>SceOrc1 (P54784)
>SceOrc2 (P32833)
>SceOrc3 (P54790)
>SceOrc4 (P54791)
>SceOrc5 (P50874)
>SceOrc6 (P38826)
>SceCDC6 (NP_012341)
>EcuOrc1 (NP_597612)
>EcuOrc2 (NP_586228)
>EcuOrc5? (NP_597618)
>EcuCdc6 (XP_955574)
>HsaOrc1 (Q13415)
>HsaOrc2 (Q13416)
>HsaOrc3 (Q9UBD5)
>HsaOrc4 (O43929)
>HsaOrc5 (O43913)
>HsaOrc6 (Q9Y5N6)
>HsaCDC6 (NP_001245)
>AthOrc1a (CAD13174)
>AthOrc2 (Q38899)
>AthOrc3 (AAT37463)
>AthOrc4 (CAE01428)
>AthOrc5 (NP_194720)
>AthOrc6 (Q9ZVH3)
>AthCdc6 (CAC81074)
>TanOrc1 TA04740 |||origin recognition complex protein 1, putative|Theileria annulata|chr 03|||Manual
>TanOrc2 TA19765 |||origin recognition complex -like protein, putative|Theileria annulata|chr 01|||Manual
>TanOrc4 TA12985 |||hypothetical protein|Theileria annulata|chr 02|||Manual
>TanCdc6 TA14555 |||CDC6-like ATPase, putative|Theileria annulata|chr 02|||Manual
>CpaOrc1 cgd4_4320
>CpaOrc2 cgd4_1930
>CpaOrc4 cgd2_1550
>CpaOrc5 cgd4_430
>CpaCdc6 cgd7_2310
>PfaOrc1 PFL0150w |||origin recognition complex 1 protein|Plasmodium falciparum|chr 12|STANFORD||Manual
>PfaOrc2 MAL7P1.21 |||origin recognition complex subunit, putative|Plasmodium falciparum|chr 7|SANGER||Manual
>PfaOrc4 PF13_0189 |||hypothetical protein, conserved|Plasmodium falciparum|chr 13|SANGER||Manual
>PfaOrc5 PFB0720c |||hypothetical protein, conserved|Plasmodium falciparum|chr 2|TIGR||Manual
>PfaCdc6 PFE0155w
>DdiOrc1 DDB0218435 ||||Dictyostelium discoideum|chr 4|||Auto
>DdiOrc2 DDB0190652 ||||Dictyostelium discoideum|chr 1|||Auto
>DdiOrc3 DDB0216767 ||||Dictyostelium discoideum|chr 1|||Auto

>DdiOrc4 DDB0168430 ||||Dictyostelium discoideum|chr 2|||Auto
>DdiOrc5 DDB0191826 |||hypothetical protein|Dictyostelium discoideum|chr 6|||Auto
>DdiOrc6 DDB0186183 ||||Dictyostelium discoideum|chr 4|||Auto
>DdiCdc6 DDB0203551 ||||Dictyostelium discoideum|chr 2|||Auto
>EhiOrc1/Cdc6 224.m00087 hypothetical protein 224.t00013 AAFB01000646
>EhiOrc2 (XP_651724)
>GlaOrc1/Cdc6 ctg02_20
>GlaOrc4 ctg02_3
>TpsOrc1a |Thaps3|263879|thaps1_ua_kg.chr_12000030
>TpsOrc1b fgenesh1_pg.C_chr_6000022 [Thaps3:5867]
>TpsOrc2 jgi|Thaps3|260866|thaps1_ua_kg.chr_1000126
>TpsCdc6 jgi|Thaps3|8536|fgenesh1_pg.C_chr_10000306
>TthOrc1 154.m00093
>TthOrc2 104.m00101
>TthOrc4 51.m00235
>TthOrc5 142.m00114
>TbrORC1/CDC6 (Tb11.02.5110)(EAN79686)
>TcrORC1/CDC6 (Tc00.1047053511159.20)(XP_806474)
>LmaORC1/CDC6 (F28.0030)(CAJ04600)
>TvaOrc1 TVAG_340580
>TvaCdc6 TVAG_376490
>TvaOrc2 TVAG_185790
>TvaOrc4 TVAG_365960
>Tb09.160.3120 | *Trypanosoma brucei* TREU927 | hypothetical protein, conserved | protein | length=1018
>Tb10.389.0050 tbrucei_v5:Tb10.389.0050/Tb927.10.13380 hypothetical protein, conserved
>Tb927.10.7980 hypothetical protein, conserved (Tb10.6k15.2570)
>DmeOrc1 gi|7304200|gb|AAF59236.1| origin recognition complex subunit 1 [Drosophila melanogaster]
>DmeOrc2 gi|7299828|gb|AAF55006.1| origin recognition complex subunit 2 [Drosophila melanogaster]
>DmeOrc3 gi|5081626|gb|AAD39472.1|AF139062_1 origin recognition complex subunit 3 [Drosophila melanogaster]
>DmeOrc4 gi|7291857|gb|AAF47276.1| origin recognition complex subunit 4 [Drosophila melanogaster]
>DmeOrc5 gi|7298101|gb|AAF53340.1| origin recognition complex subunit 5 [Drosophila melanogaster]
>DmeOrc6 gi|7303844|gb|AAF58890.1| origin recognition complex subunit 6 [Drosophila melanogaster]

Appendix 4 Genome files for microarray

core	files	length	No. contigs	content
Tb927_01_May08_v4.embl.gz	1	1064672	1	
Tb927_02_May08_v4.embl.gz	1	1193648	1	
Tb927_03_May08_v4.embl.gz	1	1653225	1	
Tb927_04_May08_v4.embl.gz	1	1590432	1	
Tb927_05_May08_v4.embl.gz	1	1608198	1	
Tb927_06_May08_v4.embl.gz	1	1618915	1	
Tb927_07_May08_v4.embl.gz	1	2205233	1	
Tb927_08_May08_v4.embl.gz	1	2481190	1	
TRYP9.cons27OCT08.gz	1	3542885	2	
Tb927_10_May08_v4.embl.gz	1	4054025	1	
Tb927_11_chain1_contigs_v5.fas.gz	1	4489688	14	
Tb927_11_chain2_contigs_v5.fas.gz	1	272620	3	
Tb927_11_chain3_contigs_v5.fas.gz	1	131299	3	
		25906030		
Subtelomeric sequences				
Chr5h_5K5.fas	1	158740	1	Chr5
BAC26D11_v3.embl.gz	1	142770	1	Chr5 TbAT locus
chr8h_27P2.fas	1	146393	1	Chr8
Tb927_11_Xhomologue_contigs_v5.fas.gz	1	580726	2	
Tb927_11_Yhomologue_contigs_v5.fas.gz	1	429803	2	
Tb927_11_chain1_LHEhomologue_contigs_v5.fas.gz	1	14898	1	
Tb927_09_remaining_6Kb_v5.fas.gz	1	522210	30	
Tb927_10_remaining_v4.embl.gz	1	865254	32	
		2860794		
Telomeric sequences				
Tb927_2L_I5_contigs1-6.embl	1	139208	6	BES-like
Tb927_2R_3E10_contigs2-4.embl	1	72378	3	
Tb927_4R_I8_Mar08.fas.gz	1	62719	1	
927T5L5E3-1f09.p1k	1	13474	1	Contains MES
Tb927_5R_1F2_contigsv1.fas	1	187762	8	Contains VSGs
Tb927_6L_1A4_contigsv2_2-16.embl	1	157939	15	Contains MES
Tb927_8R_1F6_contigsv3.fas.gz	1	57306	3	Contains part of BES
AC087700_GUTat telo35k.fas	1	35726	1	BES pt2
AC087701_GUTat telo75k.fas	1	75311	1	BES pt1
		801823		
total	30	29568647	140	

Remaining core regions from table above

TRYP9.cons27OCT08.gz

```
fasta_record          1 3472654    tryp_IXb-180h05.q1ka TRYP9.0.3648
fasta_record    3472655 3542885    tryp_IXb-293e04.p1c TRYP9.0.85070
```

Tb927_11_chain1_contigs_v5.fas.gz. . . . . Oct 28 17:51   1348K

```
fasta_record          1  163188    J21197Cg06.q1k TRYP11P7.0.20100
fasta_record     163189  176390    GTRYP_6322_6323-1a02.w2k6322
TRYP11P6.0.9973
fasta_record     176391  296585    tryp_XI-874e08.p1k
TRYP11P4.0.5290
fasta_record     296586  433652    tryp_XI-890e12.p1k
TRYP11P4.0.3259
```

fasta_record 433653 568587 TP7d2Lh12.p1k TRYP11P4.0.8510
fasta_record 568588 1417114 tryp_XI-491d08.p1ka TRYP11P3.0.16673
fasta_record 1417115 1712115 tryp_XI-463c04.p1k TRYP11P5.0.2016
fasta_record 1712116 1835124 TP25A10Ea02.p1k TRYP11P6.0.6601
fasta_record 1835125 2215864 TP27I12Cf01.p1k TRYP11P7.0.8359
fasta_record 2215865 2321151 tryp_XI-924a01.p1k TRYP11P2.0.1650
fasta_record 2321152 2586020 tryp_XI-976a09.q2k95 TRYP11P9.0.4907
fasta_record 2586021 3036733 tryp_XI-330a07.q1k TRYP11P1.0.10229
fasta_record 3036734 3252777 TP2H24Gb03.q1k TRYP11P1.0.16681
fasta_record 3252778 **4489688** tryp_XI-419d08.q2kA161 TRYP11P8.0.24064

Tb927_11_chain2_contigs_v5.fas.gz. . . . . Oct 28 17:51 84K
fasta_record 1 69920 tryp_XI-428a01.p1ka TRYP11P6.0.1
fasta_record 69921 233603 tryp_XI-1074f04.q1k TRYP11P8.0.642
fasta_record 233604 **272620** TP48c6Fh07.q1k TRYP11P6.0.11240

Tb927_11_chain3_contigs_v5.fas.gz. . .. . Oct 28 17:51 40K
fasta_record 1 59030 tryp_XI-496g08.q1k TRYP11P7.0.16684
fasta_record 59031 80079 TP46e1Ca07.p2k2166 TRYP11P2.0.624
fasta_record 80080 **131299** tryp_XI-371e06.p2k5133 TRYP11P7.0.20739

Subtelomeric sequences

Tb927_11_Xhomologue_contigs_v5.fas.gz. . . . . Oct 28 17:51 176K
fasta_record 1 494177 tryp_XI-322b06.p1k TRYP11P8.0.18497
fasta_record 494178 580726 TP27G7Ah06.q1k TRYP11P7.0.11531

Tb927_11_Yhomologue_contigs_v5.fas.gz. . . . . Oct 28 17:51 130K
fasta_record 1 353602 TP3g7Hh01.q1k TRYP11P5.0.17854
fasta_record 353603 429803 J22763Jb08.q1k TRYP11P7.0.20742

Tb927_11_chain1_LHEhomologue_contigs_v5.fas.gz . Oct 28 17:51 5041
>tryp_XI-157e02.p1c TRYP11P7.0.22752 **14898**bp

Tb927_09_remaining_6Kb_v5.fas.gz
fasta_record 1 105076 TP26P4-10b07.q1k TRYP9.0.81996
fasta_record 105077 180962 tryp_IXb-68c09.p1c TRYP9.0.21459
fasta_record 180963 223982 tryp_IXb-314f06.q1c TRYP9.0.83811
fasta_record 223983 266768 TP26P4-2d02.q1k TRYP9.0.79194
fasta_record 266769 292726 TP3F6-6h08.p1k TRYP9.0.87204
fasta_record 292727 310765 tryp_IXb-337a08.q1c TRYP9.0.84971
fasta_record 310766 328359 tryp_IXa-9g07.q1c TRYP9.0.1357
fasta_record 328360 342142 tryp_IXb-90d08.q1c TRYP9.0.78174
fasta_record 342143 355876 Tp_ends-40b06.q1k TRYP9.0.79877
fasta_record 355877 367812 tryp_IXb-385f12.p1c TRYP9.0.2975
fasta_record 367813 379071 tryp_IXb-186g06.p1c TRYP9.0.87714
fasta_record 379072 390112 tryp_IXb-354h01.q1c TRYP9.0.78353
fasta_record 390113 399888 tryp_IXb-292h05.q1c TRYP9.0.85916
fasta_record 399889 409641 tryp_IXa-2f12.q1c TRYP9.0.86440
fasta_record 409642 418193 tryp_IXa-17f02.q1c TRYP9.0.23707
fasta_record 418194 426201 tryp_IXa-23c06.p1c TRYP9.0.88680
fasta_record 426202 434149 tryp_IXb-360b06.p1c TRYP9.0.80492
fasta_record 434150 441974 TP23e5-1g01.q1k TRYP9.0.80678
fasta_record 441975 449812 TP28J11-2b06.p1k TRYP9.0.87973
fasta_record 449813 457591 tryp_IXa-3h12.p1c TRYP9.0.23660
fasta_record 457592 464617 TP30EO1-1e08.p1k TRYP9.0.77746
fasta_record 464618 471380 tryp_IXb-75h06.q1c TRYP9.0.22760

```
fasta_record    471381 478026    tryp_IXa-2c04.p1ka TRYP9.0.84936
fasta_record    478027 484639    tryp_IXb-211e12.p1c TRYP9.0.587
fasta_record    484640 491215    tryp_IXb-215c04.p1c TRYP9.0.1028
fasta_record    491216 497664    TP8h5-He11.p1k TRYP9.0.20543
fasta_record    497665 503873    TP22C17-5h06.p1c TRYP9.0.23356
fasta_record    503874 510074    tryp_IXb-380g06.q1c TRYP9.0.197
fasta_record    510075 516231    tryp_IXa-33d02.p1c TRYP9.0.887
fasta_record    516232 522210    TP22C17-3a09.p1c TRYP9.0.81626

Tb927_10_remaining_v4.embl.gz
source               1  26266    /origid="TP3F6-6f06.p1k"
source           26267  40256    /origid="TP3F6-6h01.p1k"
source           40257  51590    /origid="TPB1C9-1a08.q1k"
source           51591  66283    /origid="Tp_ends-39g01.p1k"  no
probes (repeats)
source           66284  79243    /origid="tryp_X-106b08.p1c"  no
probes (repeats)
source           79244  89400    /origid="tryp_X-149c07.p1c"
source           89401 101769    /origid="tryp_X-149e07.q1c"  no
probes (repeats)
source          101770 119545    /origid="tryp_X-155h05.q1c"
source          119546 142206    /origid="tryp_X-174f04.q1c"
source          142207 161636    /origid="tryp_X-179b07.p1c"
source          161637 187665    /origid="tryp_X-188b09.p2kB601"
source          187666 212557    /origid="tryp_X-206a03.p1c"
source          212558 255431    /origid="tryp_X-220a01.q1c"
source          255432 267898    /origid="tryp_X-232f09.q1c"
source          267899 329822    /origid="tryp_X-254c10.q1c"
source          329823 341021    /origid="tryp_X-264e11.p1c"
source          341022 351266    /origid="tryp_X-275g09.p1c"  no
probes (repeats)
source          351267 382388    /origid="tryp_X-276b11.q1c"
source          382389 424917    /origid="tryp_X-284f09.p1c"  no
probes (repeats)
source          424918 489696    /origid="tryp_X-302f11.q1ca"  first
part no probes (repeats)
source          489697 506672    /origid="tryp_X-313e04.q2kB520"
source          506673 596326    /origid="tryp_X-324h11.p1k"
source          596327 628954    /origid="tryp_X-333c11.p1c"  no
probes (repeats)
source          628955 644712    /origid="tryp_X-36a02.p1c"
source          644713 661218    /origid="tryp_X-423d07.p2kB114"
source          661219 676570    /origid="tryp_X-432f03.p2kB42"
source          676571 723138    /origid="tryp_X-448g06.q1c"
source          723139 738015    /origid="tryp_X-46d09.q1c"  no
probes [4 VSG pseudogenes)
source          738016 750267    /origid="tryp_X-48f04.q1c"
source          750268 784033    /origid="tryp_X-50f09.p2kB193"
source          784034 799211    /origid="tryp_X-54d09.q1c"
source          799212 815787    /origid="tryp_X-68d04.q1c"
source          815788 865254    /origid="tryp_X-76f06.p1c"


Telomeric sequences
Tb927_2L_I5_contigs1-6.embl
fasta_record         1  54822    T927_IIL_1-5Id10.p1k T927_IIL_1-
5.0.2758
fasta_record     54823  81466    T927_IIL_1-5R2c11.p1k T927_IIL_1-
5.0.2268
fasta_record     81467 101891    T927_IIL_1-5X2b11.p1k T927_IIL_1-
5.0.421
fasta_record    101892 116892    T927_IIL_1-5Q2d08.p1k T927_IIL_1-
5.0.1360
fasta_record    116893 128227    T927_IIL_1-5Rf06.q1k T927_IIL_1-
5.0.3699
```

```
fasta_record    128228 139208    T927_IIL_1-5J2f08.p1k T927_IIL_1-
5.0.567


Tb927_2R_3E10_contigs2-4.embl
fasta_record        1 10409    T927_IIR_3E10-21b10.p1k
T927_IIR_3E10.0.1545
fasta_record    10410 66030    T927_IIR_3E10-17e04.p1k
T927_IIR_3E10.0.193
fasta_record    66031 72378    T927_IIR_3E10-17b02.p1k
T927_IIR_3E10.0.52

Tb927_4R_I8_Mar08.fas.gz
T927_teloIVR-33d07.q1k    VSG

927T5L5E3-1f09.p1k

Tb927_5R_1F2_contigsv1.fas
fasta_record         1  44726    927T8R1F2-42e01.q1k
T927_VIIIR_1F2.0.2493    VSG
fasta_record     44727  85990    927T8R1F2-20c07.p1k
T927_VIIIR_1F2.0.3343    VSG
fasta_record     85991 108895    927T8R1F2-36f11.q1k
T927_VIIIR_1F2.0.44    VSG
fasta_record    108896 125319    927T8R1F2-28d06.p1k
T927_VIIIR_1F2.0.2049    VSG
fasta_record    125320 139229    927T8R1F2-3b02.q1k
T927_VIIIR_1F2.0.773    VSG
fasta_record    139230 151329    927T8R1F2-15c03.p1k
T927_VIIIR_1F2.0.1225    VSG
fasta_record    151330 161698    927T8R1F2-24g09.q1k
T927_VIIIR_1F2.0.1616    VSG
fasta_record    161699 171605    927T8R1F2-4g03.p1k
T927_VIIIR_1F2.0.1534    VSG
fasta_record    171606 180650    927T8R1F2-21g02.p1k
T927_VIIIR_1F2.0.1101    VSG
fasta_record    180651 187762    927T8R1F2-26c08.q1k
T927_VIIIR_1F2.0.1910    VSG


fasta_record         1  28228    927T6L1A4-32a11.q1k
T927_VIL_1A4.0.2730
fasta_record     28229  44806    927T6L1A4-5f05.p1k
T927_VIL_1A4.0.2278
fasta_record     44807  58655    927T6L1A4-18g07.p1k
T927_VIL_1A4.0.2688
fasta_record     58656  69790    927T6L1A4-41b09.p1k
T927_VIL_1A4.0.1138
fasta_record     69791  80379    927T6L1A4-27e05.p1k
T927_VIL_1A4.0.1863
fasta_record     80380  90443    927T6L1A4-1b06.q1k
T927_VIL_1A4.0.112
fasta_record     90444 100171    927T6L1A4-34b05.p1k
T927_VIL_1A4.0.3798
fasta_record    100172 108329    927T6L1A4-19a02.p1k
T927_VIL_1A4.0.2545
fasta_record    108330 116251    927T6L1A4-14c03.q1k
T927_VIL_1A4.0.4023
fasta_record    116252 123690    927T6L1A4-4e05.q1k
T927_VIL_1A4.0.2369
fasta_record    123691 130886    927T6L1A4-8f11.q1k
T927_VIL_1A4.0.1651
fasta_record    130887 138006    927T6L1A4-6e08.p1k
T927_VIL_1A4.0.1500
```

```
fasta_record    138007 145046    927T6L1A4-18f03.p1k
T927_VIL_1A4.0.3117
fasta_record    145047 151666    927T6L1A4-16h06.p1k
T927_VIL_1A4.0.805
fasta_record    151667 157939    927T6L1A4-11g05.p1k
T927_VIL_1A4.0.2549    #
```

Tb927_8R_1F6_contigsv3.fas.gz

```
fasta_record        1 40608    927T5R1F6-1h06.q1k T927_VR_1F6.0.51
```

BES hit VSG (pseudo?)

```
fasta_record    40609 52165    927T5R1F6-11f02.p1k
T927_VR_1F6.0.1000    VSG
fasta_record    52166 57306    927T5R1F6-2a08.q1k T927_VR_1F6.0.746
VSG
```

8 References

Adl, S. M., Simpson, A. G. B., 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. J. R. 2005, "The new higher level classification of eukaryotes with emphasis on the taxonomy of protists", *Journal of Eukaryotic Microbiology*, vol. 52, no. 5, pp. 399-451.

Aladjem, M. I., Falaschi, A., & Kowalski, D. 2006, "Eukaryotic DNA Replication Origins," in *DNA Replication and Human Disease*, Second edn, M. L. DePamphilis, ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 31-61.

Aladjem, M. I. 2007, "Replication in context: dynamic regulation of DNA replication patterns in metazoans", *Nature Reviews Genetics*, vol. 8, no. 8, pp. 588-600.

Aladjem, M. I., Groudine, M., Brody, L. L., Dieken, E. S., Fournier, R. E. K., Wahl, G. M., & Epner, E. M. 1995, "Participation of the Human Beta-Globin Locus-Control Region in Initiation of DNA-Replication", *Science*, vol. 270, no. 5237, pp. 815-819.

Alarcon, C. M., Son, H. J., Hall, T., & Donelson, J. E. 1994, "A Monocistronic Transcript for A Trypanosome Variant Surface Glycoprotein", *Molecular and Cellular Biology*, vol. 14, no. 8, pp. 5579-5591.

Alsford, S. & Horn, D. 2008, "Single-locus targeting constructs for reliable regulated RNAi and transgene expression in *Trypanosoma brucei*", *Molecular and Biochemical Parasitology*, vol. 161, no. 1, pp. 76-79.

Andrade, M. A., Petosa, C., O'Donoghue, S. I., Muller, C. W., & Bork, P. 2001, "Comparison of ARM and HEAT protein repeats", *Journal of Molecular Biology*, vol. 309, no. 1, pp. 1-18.

Aparicio, O. M., Weinstein, D. M., & Bell, S. P. 1997, "Components and dynamics of DNA replication complexes in S-cerevisiae: Redistribution of MCM proteins and Cdc45p during S phase", *Cell*, vol. 91, no. 1, pp. 59-69.

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., Subramanian, S., Thibodeau, R., Tivey, A., Treatman, C., Velarde, G., & Wang, H. M. 2010, "TriTrypDB: a functional genomic resource for the Trypanosomatidae", *Nucleic Acids Research*, vol. 38, p. D457-D462.

Baldinger, T. & Gossen, M. 2009, "Binding of Drosophila Orc Proteins to Anaphase Chromosomes Requires Cessation of Mitotic Cyclin-Dependent Kinase Activity", *Molecular and Cellular Biology*, vol. 29, no. 1, pp. 140-149.

Barrett, M. P. 2006, "The rise and fall of sleeping sickness", *Lancet*, vol. 367, no. 9520, pp. 1377-1378.

Barrett, M. P., Boykin, D. W., Brun, R., & Tidwell, R. R. 2007, "Human African trypanosomiasis: pharmacological re-engagement with a neglected disease", *Br J Pharmacol*, vol. 152, no. 8, pp. 1155-1171.

Barrett, M. P., Burchmore, R. J. S., Stich, A., Lazzari, J. O., Frasch, A. C., Cazzulo, J. J., & Krishna, S. 2003, "The trypanosomiases", *Lancet*, vol. 362, no. 9394, pp. 1469-1480.

Barrett, M. P. 1999, "The fall and rise of sleeping sickness", *The Lancet*, vol. 353, no. 9159, pp. 1113-1114.

Barry, E. R. & Bell, S. D. 2006, "DNA replication in the archaea", *Microbiology and Molecular Biology Reviews*, vol. 70, no. 4, p. 876-+.

Barry, J. D. & McCulloch, R. 2001, "Antigenic variation in trypanosomes: Enhanced phenotypic variation in a eukaryotic parasite", *Advances in Parasitology, Vol 49*, vol. 49, pp. 1-70.

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*", *Molecular and Biochemical Parasitology*, vol. 91, no. 1, pp. 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 Research*, vol. 14, no. 11, pp. 2319-2329.

Bell, S. P., Kobayashi, R., & Stillman, B. 1993, "Yeast origin recognition complex functions in transcription silencing and DNA replication", *Science*, vol. 262, no. 5141, pp. 1844-1849.

Bell, S. P., Mitchell, J., Leber, J., Kobayashi, R., & Stillman, B. 1995, "The Multidomain Structure of Orc1P Reveals Similarity to Regulators of Dna-Replication and Transcriptional Silencing", *Cell*, vol. 83, no. 4, pp. 563-568.

Bell, S. P. & Stillman, B. 1992, "ATP - Dependent Recognition of Eukaryotic Origins of DNA-Replication by A Multiprotein Complex", *Nature*, vol. 357, no. 6374, pp. 128-134.

Bell, S. P. & Dutta, A. 2002, "DNA replication in eukaryotic cells", *Annual Review of Biochemistry*, vol. 71, no. 1, pp. 333-374.

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. A., 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., Melville, S. E., & El Sayed, N. M. 2005, "The Genome of the African Trypanosome *Trypanosoma brucei*", *Science*, vol. 309, no. 5733, pp. 416-422.

Bird, R. E., Caro, L., MARTUSCE.J, & Louarn, J. 1972, "Origin and Sequence of Chromosome Replication in Escherichia-Coli", *Journal of Molecular Biology*, vol. 70, no. 3, p. 549-&.

Bochman, M. L. & Schwacha, A. 2009, "The Mcm Complex: Unwinding the Mechanism of a Replicative Helicase", *Microbiology and Molecular Biology Reviews*, vol. 73, no. 4, pp. 652-683.

Borst, P. & Sabatini, R. 2008, "Base J: Discovery, Biosynthesis, and Possible Functions", *Annual Review of Microbiology*, vol. 62, pp. 235-251.

Bowers, J. L., Randell, J. C. W., Chen, S. Y., & Bell, S. P. 2004, "ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication", *Molecular Cell*, vol. 16, no. 6, pp. 967-978.

Bramhill, D. & Kornberg, A. 1988, "Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the E. coli chromosome", *Cell*, vol. 52, no. 5, pp. 743-755.

Brewer, B. J. & Fangman, W. L. 1987, "The Localization of Replication Origins on Ars Plasmids in Saccharomyces-Cerevisiae", *Cell*, vol. 51, no. 3, pp. 463-471.

Bruhn, D. F., Mozeleski, B., Falkin, L., & Klingbeil, M. M. 2010, "Mitochondrial DNA polymerase POLIB is essential for minicircle DNA replication in African trypanosomes", *Molecular Microbiology*, vol. 75, no. 6, pp. 1414-1425.

Brun, R. 1992, "Cultivation of Human Pathogenic Blood Protozoa - A Citation-Classic Commentary on Cultivation and Invitro Cloning of Procyclic Culture Forms of Trypanosoma-Brucei in A Semi-Defined Medium by Brun,R., and Schonenberger,M", *Current Contents/Agriculture Biology & Environmental Sciences* no. 20, p. 8.

Brun, R., Blum, J., Chappuis, F., & Burri, C. 2010, "Human African trypanosomiasis", *Lancet*, vol. 375, no. 9709, pp. 148-159.

Burgess, R. J., Guy, M. P., & Zhang, Z. G. 2009, "Fueling transcriptional silencing with Gas1", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 27, pp. 10879-10880.

Burhans, T., Ramachandran, L., Wang, J. X., Liang, P., Patterton, H. G., Breitenbach, M., & Burhans, W. C. 2006, "Non-random clustering of stress-related genes during evolution of the S. cerevisiae genome", *Bmc Evolutionary Biology*, vol. 6.

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", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 41, pp. 15837-15842.

Callejas, S., Leech, V., Reitter, C., & Melville, S. 2006, "Hemizygous subtelomeres of an African trypanosome chromosome may account for over 75% of chromosome length", *Genome Research*, vol. 16, no. 9, pp. 1109-1118.

Carpentieri, F., De Felice, M., De Falco, M., Rossi, M., & Pisani, F. M. 2002, "Physical and functional interaction between the mini-chromosome maintenance-like DNA helicase and the single-stranded DNA binding protein from the crenarchaeon Sulfolobus solfataricus", *Journal of Biological Chemistry*, vol. 277, no. 14, pp. 12118-12127.

Carr, K. M. & Kaguni, J. M. 2001, "Stoichiometry of DnaA and DnaB protein in initiation at the Escherichia coli chromosomal origin", *Journal of Biological Chemistry*, vol. 276, no. 48, pp. 44919-44925.

Cavalier-Smith, T. 2010, "Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree", *Biology Letters*, vol. 6, no. 3, pp. 342-345.

Chaves, I., Rudenko, G., Dirks-Mulder, A., Cross, M., & Borst, P. 1999, "Control of variant surface glycoprotein gene-expression sites in *Trypanosoma brucei*", *Embo Journal*, vol. 18, no. 17, pp. 4846-4855.

Chen, S., de Vries, M. A., & Bell, S. P. 2007, "Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading", *Genes & Development*, vol. 21, no. 22, pp. 2897-2907.

Chen, X., Jorgenson, E., & Cheung, S. T. 2009, "New tools for functional genomic analysis", *Drug Discovery Today*, vol. 14, no. 15-16, pp. 754-760.

Chen, Z. Q., Speck, C., Wendel, P., Tang, C. Y., Stillman, B., & Lill, H. L. 2008, "The architecture of the DNA replication origin recognition complex in Saccharomyces cerevisiae", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 30, pp. 10326-10331.

Chesnokov, I., Remus, D., & Botchan, M. 2001, "Functional analysis of mutant and wild-type Drosophila origin recognition complex", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 21, pp. 11997-12002.

Chong, J. P. J., Hayashi, M. K., Simon, M. N., Xu, R. M., & Stillman, B. 2000, "A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 4, pp. 1530-1535.

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", *Nature Structural & Molecular Biology*, vol. 13, no. 8, pp. 684-690.

Clayton, C. E. 2002, "Life without transcriptional control? From fly to man and back again", *Embo Journal*, vol. 21, no. 8, pp. 1881-1888.

Cliffe, L. J., Siegel, T. N., Marshall, M., Cross, G. A. M., & Sabatini, R. 2010, "Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of *Trypanosoma brucei*", *Nucleic Acids Research*, vol. 38, no. 12, pp. 3923-3935.

Clyne, R. K. & Kelly, T. J. 1995, "Genetic analysis of an ARS element from the fission yeast Schizosaccharomyces pombe", *Embo Journal*, vol. 14, no. 24, pp. 6348-6357.

Collins, N., Poot, R. A., Kukimoto, I., Garcia-Jimenez, C., Dellaire, G., & Varga-Weisz, P. D. 2002, "An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin", *Nature Genetics*, vol. 32, no. 4, pp. 627-632.

Cook, J. G., Chasse, D. A. D., & Nevins, J. R. 2004, "The regulated association of Cdt1 with minichromosome maintenance proteins and Cdc6 in mammalian cells", *Journal of Biological Chemistry*, vol. 279, no. 10, pp. 9625-9633.

Cotterill, S. & Kearsey, S. E. 2009, "DNAReplication: a database of information and resources for the eukaryotic DNA replication community", *Nucleic Acids Research*, vol. 37, p. D837-D839.

Crooke, E., Hwang, D. S., Skarstad, K., Thony, B., & Kornberg, A. 1991, "Escherichia-Coli Minichromosome Replication - Regulation of Initiation at Oric", *Research in Microbiology*, vol. 142, no. 2-3, pp. 127-130.

Dacks, J. B. & Doolittle, W. F. 2001, "Reconstructing/deconstructing the earliest eukaryotes: How comparative genomics can help", *Cell*, vol. 107, no. 4, pp. 419-425.

Dacks, J. B., Walker, G., & Field, M. C. 2008, "Implications of the new eukaryotic systematics for parasitologists", *Parasitology International*, vol. 57, no. 2, pp. 97-104.

Danis, E., Brodolin, K., Menut, S., Maiorano, D., Girard-Reydet, C., & Mechali, M. 2004, "Specification of a DNA replication origin by a transcription complex", *Nature Cell Biology*, vol. 6, no. 8, pp. 721-730.

Das, A. K., Cohen, P. T. W., & Barford, D. 1998, "The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-

mediated protein-protein interactions", *Embo Journal*, vol. 17, no. 5, pp. 1192-1199.

De Felice, M., Esposito, L., Pucci, B., Carpentieri, F., De Falco, M., Rossi, M., & Pisani, F. M. 2003, "Biochemical characterization of a CDC6-like protein from the crenarchaeon Sulfolobus solfataricus", *Journal of Biological Chemistry*, vol. 278, no. 47, pp. 46424-46431.

De Felice, M., Esposito, L., Rossi, M., & Pisani, F. M. 2006, "Biochemical characterization of two Cdc6/ORC1-like proteins from the crenarchaeon Sulfolobus solfataricus", *Extremophiles*, vol. 10, no. 1, pp. 61-70.

Delgado, S., Gomez, M., Bird, A., & Antequera, F. 1998, "Initiation of DNA replication at CpG islands in mammalian chromosomes", *Embo Journal*, vol. 17, no. 8, pp. 2426-2435.

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", *Molecular Cell*, vol. 35, no. 4, pp. 403-413.

DePamphilis, M. L. 2005, "Cell cycle dependent regulation of the origin recognition complex", *Cell Cycle*, vol. 4, no. 1, pp. 70-79.

Dhar, S. K., Delmolino, L., & Dutta, A. 2001, "Architecture of the human origin recognition complex", *Journal of Biological Chemistry*, vol. 276, no. 31, pp. 29067-29071.

Diffley, J. F. X., Cocker, J. H., Dowell, S. J., & Rowley, A. 1994, "2 Steps in the Assembly of Complexes at Yeast Replication Origins In-Vivo", *Cell*, vol. 78, no. 2, pp. 303-316.

Dionne, I., Robinson, N. P., McGeoch, A. T., Marsh, V. L., Reddish, A., & Bell, S. D. 2003, "DNA replication in the hyperthermophilic archaeon Sulfolobus solfataricus", *Biochemical Society Transactions*, vol. 31, pp. 674-676.

Donovan, S., Harwood, J., Drury, L. S., & Diffley, J. F. X. 1997, "Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 11, pp. 5611-5616.

Donti, T. R., Datta, S., Sandoval, P. Y., & Kapler, G. M. 2009, "Differential targeting of Tetrahymena ORC to ribosomal DNA and non-rDNA replication origins", *Embo Journal*, vol. 28, no. 3, pp. 223-233.

Dreesen, O., Li, B. B., & Cross, G. A. M. 2007, "Telomere structure and function in trypanosomes: a proposal", *Nature Reviews Microbiology*, vol. 5, no. 1, pp. 70-75.

Drury, L. S., Perkins, G., & Diffley, J. F. X. 1997, "The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast", *Embo Journal*, vol. 16, no. 19, pp. 5966-5976.

Dubessay, P., Ravel, C., Bastien, P., Crobu, L., Dedet, J. P., Pages, M., & Blaineau, C. 2002, "The switch region on Leishmania major chromosome 1 is not required for mitotic stability or gene expression, but appears to be essential", *Nucleic Acids Research*, vol. 30, no. 17, pp. 3692-3697.

Dubey, D. D., Zhu, J. G., Carlson, D. L., Sharma, K., & Huberman, J. A. 1994, "3 Ars Elements Contribute to the Ura4-Replication Origin Region in the Fission Yeast, Schizosaccharomyces-Pombe", *Embo Journal*, vol. 13, no. 15, pp. 3638-3647.

Dueber, E. L. C., Corn, J. E., Bell, S. D., & Berger, J. M. 2007, "Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex", *Science*, vol. 317, no. 5842, pp. 1210-1213.

Duggin, I. G. & Bell, S. D. 2006, "The chromosome replication machinery of the archaeon Sulfolobus solfataricus", *Journal of Biological Chemistry*, vol. 281, no. 22, pp. 15029-15032.

Duggin, I. G., McCallum, S. A., & Bell, S. D. 2008, "Chromosome replication dynamics in the archaeon Sulfolobus acidocaldarius", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 43, pp. 16737-16742.

Duncker, B. P., Chesnokov, I. N., & McConkey, B. J. 2009, "The origin recognition complex protein family", *Genome Biology*, vol. 10, no. 3.

Dutta, A. & Bell, S. P. 2006, "Assembly of Pre-replication Complexes," in *DNA Replication and Human Disease*, Second edn, M. L. DePamphilis, ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 63-88.

Ehrenberg, J. & Ault, S. 2005, "Neglected diseases of neglected populations: Thinking to reshape the determinants of health in Latin America and the Caribbean", *BMC Public Health*, vol. 5, no. 1, p. 119.

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. H., Huang, Y. T., Kindlund, E., Ktingbeil, 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., Tarteton, 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*, vol. 309, no. 5733, pp. 409-415.

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. M., 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., Metville, S. E., Donelson, J. E., Andersson, B., Stuart, K. D., & Hall, N. 2005b, "Comparative genomics of trypanosomatid parasitic protozoa", *Science*, vol. 309, no. 5733, pp. 404-409.

El Sayed, N. M. A., Ghedin, E., Song, J. M., MacLeod, A., Bringaud, F., Larkin, C., Wanless, D., Peterson, J., Hou, L. H., Taylor, S., Tweedie, A., Biteau, N., Khalak, H. G., Lin, X. Y., Mason, T., Hannick, L., Caler, E., Blandin, G., Bartholomeu, D., Simpson, A. J., Kaul, S., Zhao, H., Pai, G., Van Aken, S., Utterback, T., Haas, B., Koo, H. L., Umayam, L., Suh, B., Gerrard, C., Leech, V., Qi, R., Zhou, S. G., Schwartz, D., Feldblyum, T., Salzberg, S., Tait, A., Turner, C. M. R., Ullu, E., White, O., Melville, S., Adams, M. D., Fraser, C. M., & Donelson, J. E. 2003, "The sequence and analysis of *Trypanosoma brucei* chromosome II", *Nucleic Acids Research*, vol. 31, no. 16, pp. 4856-4863.

El Sayed, N. M., Hegde, P., Quackenbush, J., Melville, S. E., & Donelson, J. E. 2000, "The African trypanosome genome", *International Journal for Parasitology*, vol. 30, no. 4, pp. 329-345.

Elias, M. C. & Faria, M. 2009, "Are There Epigenetic Controls in Trypanosoma cruzi?", *Natural Genetic Engineering and Natural Genome Editing*, vol. 1178, pp. 285-290.

Erzberger, J. P. & Berger, J. M. 2006, "Evolutionary relationships and structural mechanisms of AAA plus proteins", *Annual Review of Biophysics and Biomolecular Structure*, vol. 35, pp. 93-114.

Erzberger, J. P., Pirruccello, M. M., & Berger, J. M. 2002a, "The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation", *Embo Journal*, vol. 21, no. 18, pp. 4763-4773.

Erzberger, J. P., Pirruccello, M. M., & Berger, J. M. 2002b, "The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation", *Embo Journal*, vol. 21, no. 18, pp. 4763-4773.

Fang, L. H., Davey, M. J., & O'Donnell, M. 1999, "Replisome assembly at oriC, the replication origin of E-coli, reveals an explanation for initiation sites outside an origin", *Molecular Cell*, vol. 4, no. 4, pp. 541-553.

Feng, L., Wang, B., Driscoll, B., & Jong, A. 2000, "Identification and characterization of Saccharomyces cerevisiae Cdc6 DNA-binding properties", *Molecular Biology of the Cell*, vol. 11, no. 5, pp. 1673-1685.

Fevre, E. M., Picozzi, K., Jannin, J., Welburn, S. C., & Maudlin, I. 2006, "Human African Trypanosomiasis: Epidemiology and Control," in *Advances in Parasitology*

Control of Human Parasitic Diseases, Volume 61 edn, H. M. David, ed., Academic Press, pp. 167-221.

Figueiredo, L. M., Janzen, C. J., & Cross, G. A. M. 2008, "A histone methyltransferase modulates antigenic variation in African trypanosomes", *Plos Biology*, vol. 6, no. 7, pp. 1539-1548.

Fire, A., Xu, S. Q., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. 1998, "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans", *Nature*, vol. 391, no. 6669, pp. 806-811.

Forsburg, S. L. 2004, "Eukaryotic MCM proteins: Beyond replication initiation", *Microbiology and Molecular Biology Reviews*, vol. 68, no. 1, p. 109-+.

Foss, M., McNally, F. J., Laurenson, P., & Rine, J. 1993, "Origin recognition complex (ORC) in transcriptional silencing and DNA replication in S. cerevisiae", *Science*, vol. 262, no. 5141, pp. 1838-1844.

Francon, P., Maiorano, D., & Mechali, M. 1999, "Initiation of DNA replication in eukaryotes: questioning the origin", *Febs Letters*, vol. 452, no. 1-2, pp. 87-91.

Fritz-Laylin, L. K., Prochnik, S. E., Ginger, M. L., Dacks, J. B., Carpenter, M. L., Field, M. C., Kuo, A., Paredez, A., Chapman, J., Pham, J., Shu, S. Q., Neupane, R., Cipriano, M., Mancuso, J., Tu, H., Salamov, A., Lindquist, E., Shapiro, H., Lucas, S., Grigoriev, I. V., Cande, W. Z., Fulton, C., Rokhsar, D. S., & Dawson, S. C. 2010, "The Genome of Naegleria gruberi Illuminates Early Eukaryotic Versatility", *Cell*, vol. 140, no. 5, pp. 631-642.

Frolova, N. S., Schek, N., Tikhmyanova, N., & Coleman, T. R. 2002, "Xenopus Cdc6 performs separate functions in initiating DNA replication", *Molecular Biology of the Cell*, vol. 13, no. 4, pp. 1298-1312.

Fuller, R. S., Funnell, B. E., & Kornberg, A. 1984, "The dnaA protein complex with the E. coli chromosomal replication origin (oriC) and other DNA sites", *Cell*, vol. 38, no. 3, pp. 889-900.

Garcia, A., Courtin, D., Solano, P., Koffi, M., & Jamonneau, V. 2006, "Human African trypanosomiasis: connecting parasite and host genetics", *Trends in Parasitology*, vol. 22, no. 9, pp. 405-409.

Gerbi, S. A., Strezoska, Z., & Waggener, J. M. 2002, "Initiation of DNA replication in multicellular eukaryotes", *Journal of Structural Biology*, vol. 140, no. 1-3, pp. 17-30.

Ghedin, E., Bringaud, F., Peterson, J., Myler, P., Berriman, M., Ivens, A., Andersson, B., Bontempi, E., Eisen, J., Angiuoli, S., Wanless, D., Von Arx, A., Murphy, L., Lennard, N., Salzberg, S., Adams, M. D., White, O., Hall, N., Stuart, K., Fraser, C. M., & El Sayed, N. M. A. 2004, "Gene synteny and evolution of genome architecture in trypanosomatids", *Molecular and Biochemical Parasitology*, vol. 134, no. 2, pp. 183-191.

Gilbert, D. M. 2010, "Evaluating genome-scale approaches to eukaryotic DNA replication", *Nat Rev Genet*.

Gill, E. E. & Fast, N. M. 2007, "Stripped-down DNA repair in a highly reduced parasite", *Bmc Molecular Biology*, vol. 8.

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", *Eukaryotic Cell*, vol. 1, no. 6, pp. 1000-1009.

Gluenz, E., Sharma, R., Carrington, M., & Gull, K. 2008, "Functional characterization of cohesin subunit SCC1 in *Trypanosoma brucei* and dissection of mutant phenotypes in two life cycle stages", *Molecular Microbiology*, vol. 69, no. 3, pp. 666-680.

Godoy, P. D. D., Nogueira, 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", *Eukaryotic Cell*, vol. 8, no. 10, pp. 1592-1603.

Gottesdiener, K., Garciaanoveros, J., Lee, M. G., & Vanderploeg, L. H. T. 1990, "Chromosome Organization of the Protozoan Trypanosoma-Brucei", *Molecular and Cellular Biology*, vol. 10, no. 11, pp. 6079-6083.

Grabowski, B. & Kelman, Z. 2003, "Archaeal DNA replication: Eukaryal proteins in a bacterial context", *Annual Review of Microbiology*, vol. 57, pp. 487-516.

Gratzner, H. G. 1982, "Monoclonal-Antibody to 5-Bromodeoxyuridine and 5-Iododeoxyuridine - A New Reagent for Detection of Dna-Replication", *Science*, vol. 218, no. 4571, pp. 474-475.

Grunfelder, C. G., Engstler, M., Weise, F., Schwarz, H., Stierhof, Y. D., Morgan, G. W., Field, M. C., & Overath, P. 2003, "Endocytosis of a glycosylphosphatidylinositol-anchored protein via clathrin-coated vesicles, sorting by default in endosomes, and exocytosis via RAB11-positive carriers", *Molecular Biology of the Cell*, vol. 14, no. 5, pp. 2029-2040.

Gull, K. 2002, "The cell biology of parasitism in *Trypanosoma brucei*: Insights and drug targets from genomic approaches?", *Current Pharmaceutical Design*, vol. 8, no. 4, pp. 241-256.

Gunzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L. C., Chung, H. M., Lee, P. T., & Lee, M. G. S. 2003, "RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*", *Eukaryotic Cell*, vol. 2, no. 3, pp. 542-551.

Hall, N., Berriman, M., Lennard, N. J., Harris, B. R., Hertz-Fowler, C., Bart-Delabesse, E. N., Gerrard, C. S., Atkin, R. J., Barron, A. J., Bowman, S., Bray-Allen, S. P., Bringaud, F., Clark, L. N., Corton, C. H., Cronin, A., Davies, R., Doggett, J., Fraser, A., Gruter, E., Hall, S., Harper, A. D., Kay, M. P., Leech, V., Mayes, R., Price, C., Quail, M. A., Rabbinowitsch, E., Reitter, C., Rutherford, K., Sasse, J., Sharp, S., Shownkeen, R., MacLeod, A., Taylor, S., Tweedie, A., Turner, C. M. R., Tait, A., Gull, K., Barrell, B., & Melville, S. E. 2003, "The DNA sequence of chromosome I of an African

trypanosome: gene content, chromosome organisation, recombination and polymorphism", *Nucleic Acids Research*, vol. 31, no. 16, pp. 4864-4873.

Hammarton, T. C. 2007, "Cell cycle regulation in *Trypanosoma brucei*", *Molecular and Biochemical Parasitology*, vol. 153, no. 1, pp. 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", *Journal of Biological Chemistry*, vol. 278, no. 25, pp. 22877-22886.

Hammarton, T. C., Engstler, M., & Mottram, J. C. 2004, "The *Trypanosoma brucei* cyclin, CYC2, is required for cell cycle progression through G(1) phase and for maintenance of procyclic form cell morphology", *Journal of Biological Chemistry*, vol. 279, no. 23, pp. 24757-24764.

Hammarton, T. C., Lillico, S. G., Welburn, S. C., & Mottram, J. C. 2005, "*Trypanosoma brucei* MOB1 is required for accurate and efficient cytokinesis but not for exit from mitosis", *Molecular Microbiology*, vol. 56, no. 1, pp. 104-116.

Hammarton, T. C., Wickstead, B., & Mckean, P. G. 2007, "Cell Structure, Cell Division and Cell Cycle," in , J. D. Barry et al., eds., Horizon Bioscience, Wymondham, UK, pp. 91-131.

Hartwell, L. H. 1976, "Sequential Function of Gene Products Relative to Dna-Synthesis in Yeast-Cell Cycle", *Journal of Molecular Biology*, vol. 104, no. 4, pp. 803-817.

Harvey, K. J. & Newport, J. 2003, "Metazoan origin selection - Origin recognition complex chromatin binding is regulated by Cdc6 recruitment and ATP hydrolysis", *Journal of Biological Chemistry*, vol. 278, no. 49, pp. 48524-48528.

Hayashi, M., Katou, Y., Itoh, T., Tazumi, M., Yamada, Y., Takahashi, T., Nakagawa, T., Shirahige, K., & Masukata, H. 2007, "Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast", *Embo Journal*, vol. 26, no. 5, pp. 1327-1339.

Heichinger, C., Penkett, C. J., Bahler, J., & Nurse, P. 2006, "Genome-wide characterization of fission yeast DNA replication origins", *Embo Journal*, vol. 25, no. 21, pp. 5171-5179.

Helfert, S., Estevez, A. M., Bakker, B., Michels, P., & Clayton, C. 2001, "Roles of triosephosphate isomerase and aerobic metabolism in *Trypanosoma brucei*", *Biochemical Journal*, vol. 357, pp. 117-125.

Hemerly, A. S., Prasanth, S. G., Siddiqui, K., & Stillman, B. 2009, "Orc1 Controls Centriole and Centrosome Copy Number in Human Cells", *Science*, vol. 323, no. 5915, pp. 789-793.

Herbig, U., Marlar, C. A., & Fanning, E. 1999, "The Cdc6 nucleotide-binding site regulates its activity in DNA replication in human cells", *Molecular Biology of the Cell*, vol. 10, no. 8, pp. 2631-2645.

Hergovich, A., Stegert, M. R., Schmitz, D., & Hemmings, B. A. 2006, "NDR kinases regulate essential cell processes from yeast to humans", *Nature Reviews Molecular Cell Biology*, vol. 7, no. 4, pp. 253-264.

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. M., Rudenko, G., Barry, J. D., Louis, E. J., & Berriman, M. 2008, "Telomeric Expression Sites Are Highly Conserved in *Trypanosoma brucei*", *Plos One*, vol. 3, no. 10.

Hertz-Fowler, C., Renauld, H., & Berriman, M. 2007, "The Genome of *Trypanosoma brucei*," in , J. D. Barry et al., eds., Horizon Bioscience, Wymondham, UK, pp. 91-131.

Hiratani, L., Takebayashi, S., Lu, J. J., & Gilbert, D. M. 2009, "Replication timing and transcriptional control: beyond cause and effect-part II", *Current Opinion in Genetics & Development*, vol. 19, no. 2, pp. 142-149.

Hirumi, H. & Hirumi, K. 1994, "Axenic Culture of African Trypanosome Blood-Stream Forms", *Parasitology Today*, vol. 10, no. 2, pp. 80-84.

Hofmann, J. F. X. & Beach, D. 1994, "Cdt1 Is An Essential Target of the Cdc10/Sct1 Transcription Factor - Requirement for Dna-Replication and Inhibition of Mitosis", *Embo Journal*, vol. 13, no. 2, pp. 425-434.

Horn, D. & Cross, G. A. M. 1995, "A Developmentally-Regulated Position Effect at A Telomeric Locus in Trypanosoma-Brucei", *Cell*, vol. 83, no. 4, pp. 555-561.

Horton, T. L. & Landweber, L. F. 2002, "Rewriting the information in DNA: RNA editing in kinetoplastids and myxomycetes", *Curr.Opin.Microbiol.*, vol. 5, no. 6, pp. 620-626.

Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D., & Savioli, L. 2007, "Control of Neglected Tropical Diseases", *The New England Journal of Medicine*, vol. 357, no. 10, pp. 1018-1027.

Hsiao, C. L. & Carbon, J. 1979, "High-Frequency Transformation of Yeast by Plasmids Containing the Cloned Yeast Arg4 Gene", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, no. 8, pp. 3829-3833.

Hughes, K., Wand, M., Foulston, L., Young, R., Harley, K., Terry, S., Ersfeld, K., & Rudenko, G. 2007, "A novel ISWI is involved in VSG expression site downregulation in African trypanosomes", *Embo Journal*, vol. 26, no. 9, pp. 2400-2410.

Hyrien, O., Maric, C., & Mechali, M. 1995, "Transition in Specification of Embryonic Metazoan Dna-Replication Origins", *Science*, vol. 270, no. 5238, pp. 994-997.

Ishimi, Y. 1997, "A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex", *Journal of Biological Chemistry*, vol. 272, no. 39, pp. 24508-24513.

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. R., 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. T., 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. G., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B., & Myler, P. J. 2005, "The genome of the kinetoplastid parasite, Leishmania major", *Science*, vol. 309, no. 5733, pp. 436-442.

Iyer, L. M., Leipe, D. D., Koonin, E. V., & Aravind, L. 2004, "Evolutionary history and higher order classification of AAA+ ATPases", *Journal of Structural Biology*, vol. 146, no. 1-2, pp. 11-31.

Jacob, F., Brenner, S., & Cuzin, F. 1963, "On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon", *Cold Spring Harbor Symp Quant Biol*, vol. **256**, pp. 298-300.

Joiner, K., Brown, E., Hammer, C., Warren, K., & Frank, M. 1983, "Studies on the Mechanism of Bacterial-Resistance to Complement-Mediated Killing .3. C5B-9 Deposits Stably on Rough and Type-7 S-Pneumoniae Without Causing Bacterial Killing", *Journal of Immunology*, vol. 130, no. 2, pp. 845-849.

Kanter, D. M., Bruck, I., & Kaplan, D. L. 2008, "Mcm Subunits Can Assemble into Two Different Active Unwinding Complexes", *Journal of Biological Chemistry*, vol. 283, no. 45, pp. 31172-31182.

Kasiviswanathan, R., Shin, J. H., & Kelman, Z. 2005, "Interactions between the archaeal Cdc6 and MCM proteins modulate their biochemical properties", *Nucleic Acids Research*, vol. 33, no. 15, pp. 4940-4950.

Kasiviswanathan, R., Shin, J. H., Melamud, E., & Kelman, Z. 2004, "Biochemical characterization of the Methanothermobacter thermautotrophicus minichromosome maintenance (MCM) helicase N-terminal domains", *Journal of Biological Chemistry*, vol. 279, no. 27, pp. 28358-28366.

Katinka, M. D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J., & Vivares, C.

P. 2001, "Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi", *Nature*, vol. 414, no. 6862, pp. 450-453.

Kawakami, H. & Katayama, T. 2010, "DnaA, ORC, and Cdc6: similarity beyond the domains of life and diversity", *Biochemistry and Cell Biology-Biochimie et Biologie Cellulaire*, vol. 88, no. 1, pp. 49-62.

Kelly, J. M. 1995, "Trypanosomatid Shuttle Vectors - New Tools for the Functional Dissection of Parasite Genomes", *Parasitology Today*, vol. 11, no. 12, pp. 447-451.

Kelly, T. J. & Stillman, B. 2006, "Duplication of DNA in Eukaryotic Cells," in *DNA Replication and Human Disease*, Second edn, M. L. DePamphilis, ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 1-29.

Kelman, L. M. & Kelman, Z. 2003, "Archaea: an archetype for replication initiation studies?", *Molecular Microbiology*, vol. 48, no. 3, pp. 605-615.

Kim, K. S. & Donelson, J. E. 1997, "Co-duplication of a variant surface glycoprotein gene and its promoter to an expression site in African trypanosomes", *Journal of Biological Chemistry*, vol. 272, no. 39, pp. 24637-24645.

Kim, T. H. & Ren, B. 2006, "Genome-Wide Analysis of Protein-DNA Interactions", *Annual Review of Genomics and Human Genetics*, vol. 7, no. 1, pp. 81-102.

Kimura, H., Takizawa, N., Nozaki, N., & Sugimoto, K. 1995, "Molecular-Cloning of Cdna-Encoding Mouse Cdc21 and Cdc46 Homologs and Characterization of the Products - Physical Interaction Between P1(Mcm3) and Cdc46 Proteins", *Nucleic Acids Research*, vol. 23, no. 12, pp. 2097-2104.

Kitagawa, R., Ozaki, T., Moriya, S., & Ogawa, T. 1998, "Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for Escherichia coli DnaA protein", *Genes & Development*, vol. 12, no. 19, pp. 3032-3043.

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*, vol. 88, no. 4, pp. 493-502.

Klemm, R. D. & Bell, S. P. 2001, "ATP bound to the origin recognition complex is important for preRC formation", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 15, pp. 8361-8367.

Klingbeil, M. M., Drew, M. E., Liu, Y. N., Morris, J. C., Motyka, S. A., Saxowsky, T. T., Wang, Z. F., & Englund, P. T. 2001, "Unlocking the secrets of trypanosome kinetoplast DNA network replication", *Protist*, vol. 152, no. 4, pp. 255-262.

Klingbeil, M. M. & Englund, P. T. 2004, "Closing the gaps in kinetoplast DNA network replication", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 13, pp. 4333-4334.

Kong, D. C. & DePamphilis, M. L. 2001a, "Site-specific DNA binding of the Schizosaccharomyces pombe origin recognition complex is determined by the Orc4 subunit", *Molecular and Cellular Biology*, vol. 21, no. 23, pp. 8095-8103.

Kong, D. C. & DePamphilis, M. L. 2001b, "Site-specific DNA binding of the Schizosaccharomyces pombe origin recognition complex is determined by the Orc4 subunit", *Molecular and Cellular Biology*, vol. 21, no. 23, pp. 8095-8103.

Kreitz, S., Ritzi, M., Baack, M., & Knippers, R. 2001, "The human origin recognition complex protein 1 dissociates from chromatin during S phase in HeLa cells", *Journal of Biological Chemistry*, vol. 276, no. 9, pp. 6337-6342.

Kumar, P. & Wang, C. C. 2006, "Dissociation of cytokinesis initiation from mitotic control in a eukaryote", *Eukaryotic Cell*, vol. 5, no. 1, pp. 92-102.

LaCount, D. J., El Sayed, N. M., Kaul, S., Wanless, D., Turner, C. M. R., & Donelson, J. E. 2001, "Analysis of a donor gene region for a variant surface glycoprotein and its expression site in African trypanosomes", *Nucleic Acids Research*, vol. 29, no. 10, pp. 2012-2019.

Lamb, J. R., Tugendreich, S., & Hieter, P. 1995, "Tetratrico Peptide Repeat Interactions - to Tpr Or Not to Tpr", *Trends in Biochemical Sciences*, vol. 20, no. 7, pp. 257-259.

Landeira, D., Bart, J. M., Van Tyne, D., & Navarro, M. 2009, "Cohesin regulates VSG monoallelic expression in trypanosomes", *Journal of Cell Biology*, vol. 186, no. 2, pp. 243-254.

Lebofsky, R. & Bensimon, A. 2005, "DNA replication origin plasticity and perturbed fork progression in human inverted repeats", *Molecular and Cellular Biology*, vol. 25, no. 15, pp. 6789-6797.

Lee, D. G. & Bell, S. P. 1997, "Architecture of the yeast origin recognition complex bound to origins of DNA replication", *Molecular and Cellular Biology*, vol. 17, no. 12, pp. 7159-7168.

Lee, J. K. & Hurwitz, J. 2000, "Isolation and characterization of various complexes of the minichromosome maintenance proteins of Schizosaccharomyces pombe", *Journal of Biological Chemistry*, vol. 275, no. 25, pp. 18871-18878.

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", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 24, pp. 13589-13594.

Lee, M. G. & Van der Ploeg, L. H. 1997, "Transcription of protein-coding genes in trypanosomes by RNA polymerase I", *Annu.Rev.Microbiol.*, vol. 51, pp. 463-489.

Lee, M. G. S., E, Y. P., & Axelrod, N. 1995, "Construction of trypanosome artificial minichromosomes", *Nucleic Acids Research*, vol. 23, no. 23, pp. 4893-4899.

Li, B., Espinal, A., & Cross, G. A. M. 2005, "Trypanosome telomeres are protected by a homologue of mammalian TRF2", *Molecular and Cellular Biology*, vol. 25, no. 12, pp. 5011-5021.

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", *Molecular and Cellular Biology*, vol. 22, no. 1, pp. 105-116.

Li, Z. Y. & Wang, C. C. 2003, "A PHO80-like cyclin and a B-type cyclin control the cell cycle of the procyclic form of *Trypanosoma brucei*", *Journal of Biological Chemistry*, vol. 278, no. 23, pp. 20652-20658.

Lidonnici, M. R., Rossi, R., Paixao, S., Mendoza-Maidonado, R., Paolinelli, R., Arcangeli, C., Giacca, M., Biamonti, G., & Montecucco, A. 2004, "Subnuclear distribution of the largest subunit of the human origin recognition complex during the cell cycle", *Journal of Cell Science*, vol. 117, no. 22, pp. 5221-5231.

Liu, B. Y., Liu, Y. N., Motyka, S. A., Agbo, E. E. C., & Englund, P. T. 2005, "Fellowship of the rings: the replication of kinetoplast DNA", *Trends in Parasitology*, vol. 21, no. 8, pp. 363-369.

Liu, B. Y., Molina, H., Kalume, D., Pandey, A., Griffith, J. D., & Englund, P. T. 2006, "Role of p38 in replication of *Trypanosoma brucei* kinetoplast DNA", *Molecular and Cellular Biology*, vol. 26, no. 14, pp. 5382-5393.

Liu, B. Y., Wang, J. Y., Yildirir, G., & Englund, P. T. 2009a, "TbPIF5 Is a *Trypanosoma brucei* Mitochondrial DNA Helicase Involved in Processing of Minicircle Okazaki Fragments", *Plos Pathogens*, vol. 5, no. 9.

Liu, B. Y., Yildirir, G., Wang, J. Y., Tolun, G., Griffith, J. D., & Englund, P. T. 2010, "TbPIF1, a *Trypanosoma brucei* Mitochondrial DNA Helicase, Is Essential for Kinetoplast Minicircle Replication", *Journal of Biological Chemistry*, vol. 285, no. 10, pp. 7056-7066.

Liu, J. Y., 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", *Molecular Cell*, vol. 6, no. 3, pp. 637-648.

Liu, Y., Richards, T. A., & Aves, S. J. 2009b, "Ancient diversification of eukaryotic MCM DNA replication proteins", *Bmc Evolutionary Biology*, vol. 9.

Loupart, M. L., Krause, S. A., & Heck, M. M. S. 2000, "Aberrant replication timing induces defective chromosome condensation in Drosophila ORC2 mutants", *Current Biology*, vol. 10, no. 24, pp. 1547-1556.

Lukes, J., Guilbride, D. L., Votypka, J., Zikova, A., Benne, R., & Englund, P. T. 2002, "Kinetoplast DNA network: Evolution of an improbable structure", *Eukaryotic Cell*, vol. 1, no. 4, pp. 495-502.

Lukes, J., Hashimi, H., & Zikova, A. 2005, "Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates", *Current Genetics*, vol. 48, no. 5, pp. 277-299.

Lundgren, M., Andersson, A., Chen, L. M., Nilsson, P., & Bernander, R. 2004, "Three replication origins in Sulfolobus species: Synchronous initiation of chromosome replication and asynchronous termination", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 18, pp. 7046-7051.

MacAlpine, D. & Bell, S. 2005, "A genomic view of eukaryotic DNA replication", *Chromosome Research*, vol. 13, no. 3, pp. 309-326.

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 Research*, vol. 20, no. 2, pp. 201-211.

Machida, Y. J., Hamlin, J. L., & Dutta, A. 2005, "Right Place, Right Time, and Only Once: Replication Initiation in Metazoans", *Cell*, vol. 123, no. 1, pp. 13-24.

Mah, A. S., Jang, J., & Deshaies, R. J. 2001, "Protein kinase Cdc15 activates the Dbf2-Mob1 kinase complex", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 13, pp. 7325-7330.

Maine, G. T., Sinha, P., & Tye, B. K. 1984, "Mutants of S-Cerevisiae Defective in the Maintenance of Minichromosomes", *Genetics*, vol. 106, no. 3, pp. 365-385.

Maiorano, D., Moreau, J., & Mechali, M. 2000, "XCDT1 is required for the assembly of pre-replicative complexes in Xenopus laevis", *Nature*, vol. 404, no. 6778, pp. 622-625.

Maiti, A. K. & Sinha, P. 1992, "The Mcm2 Mutation of Yeast Affects Replication, Rather Than Segregation Or Amplification of the 2 Micron Plasmid", *Journal of Molecular Biology*, vol. 224, no. 3, pp. 545-558.

Mancio-Silva, L., Rojas-Meza, A. P., Vargas, M., Scherf, A., & Hernandez-Rivas, R. 2008, "Differential association of Orc1 and Sir2 proteins to telomeric domains in Plasmodium falciparum", *Journal of Cell Science*, vol. 121, no. 12, pp. 2046-2053.

Marahrens, Y. & Stillman, B. 1992, "A Yeast Chromosomal Origin of DNA-Replication Defined by Multiple Functional Elements", *Science*, vol. 255, no. 5046, pp. 817-823.

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 Research*, vol. 17, no. 9, pp. 1344-1352.

Maric, C. & Prioleau, M. N. 2010, "Interplay between DNA replication and gene expression: a harmonious coexistence", *Current Opinion in Cell Biology*, vol. 22, no. 3, pp. 277-283.

Martin, K. L. & Smith, T. K. 2005, "The myo-inositol-1-phosphate synthase gene is essential in Tryponosoma brucei", *Biochemical Society Transactions*, vol. 33, pp. 983-985.

Martin, K. L. & Smith, T. K. 2006, "Phosphatidylinositol synthesis is essential in bloodstream form *Trypanosoma brucei*", *Biochemical Journal*, vol. 396, pp. 287-295.

Martinez-Calvillo, S., Vizuet-de-Rueda, J. C., Florencio-Martinez, L. E., Manning-Cela, R. G., & Figueroa-Angulo, E. E. 2010, "Gene Expression in Trypanosomatid Parasites", *Journal of Biomedicine and Biotechnology*.

Masai, H., Matsumoto, S., You, Z. Y., Yoshizawa-Sugata, N., & Oda, M. 2010, "Eukaryotic Chromosome DNA Replication: Where, When and How?", *Annual Review of Biochemistry, Vol 79*, vol. 79, pp. 89-130.

Maslov, D. A., Podtipaev, S. A., & Lukes, J. 2001, "Phylogeny of the kinetoplastida: Taxonomic problems and insights into the evolution of parasitism", *Memorias do Instituto Oswaldo Cruz*, vol. 96, no. 3, pp. 397-402.

Masterson, W. J., Doering, T. L., Hart, G. W., & Englund, P. T. 1989, "A Novel Pathway for Glycan Assembly - Biosynthesis of the Glycosyl-Phosphatidylinositol Anchor of the Trypanosome Variant Surface Glycoprotein", *Cell*, vol. 56, no. 5, pp. 793-800.

Matsunaga, F., Forterre, P., Ishino, Y., & Myllykallio, H. 2001, "In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 20, pp. 11152-11157.

Matthews, K. R. 2005, "The developmental cell biology of *Trypanosoma brucei*", *Journal of Cell Science*, vol. 118, no. 2, pp. 283-290.

Maudlin, I. 2006, "African trypanosomiasis", *Annals of Tropical Medicine and Parasitology*, vol. 100, no. 8, pp. 679-701.

McCulloch, R. 2004, "Antigenic variation in African trypanosomes: monitoring progress", *Trends in Parasitology*, vol. 20, no. 3, pp. 117-121.

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", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 9, pp. 2811-2816.

McGarry, T. J. & Kirschner, M. W. 1998, "Geminin, an inhibitor of DNA replication, is degraded during mitosis", *Cell*, vol. 93, no. 6, pp. 1043-1053.

Mckean, P. G. 2003, "Coordination of cell cycle and cytokinesis in *Trypanosoma brucei*", *Current Opinion in Microbiology*, vol. 6, no. 6, pp. 600-607.

McNairn, A. J., Okuno, Y., Misteli, T., & Gilbert, D. M. 2005, "Chinese hamster ORC subunits dynamically associate with chromatin throughout the cell-cycle", *Experimental Cell Research*, vol. 308, no. 2, pp. 345-356.

Mechali, M. 2010, "Eukaryotic DNA replication origins: many choices for appropriate answers", *Nat Rev Mol Cell Biol*, vol. 11, no. 10, pp. 728-738.

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", *Molecular and Biochemical Parasitology*, vol. 94, no. 2, pp. 155-173.

Messer, W. 2002, "The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication", *Fems Microbiology Reviews*, vol. 26, no. 4, pp. 355-374.

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*, vol. 81, no. 8-9, pp. 819-825.

Metzenberg, S. & Agabian, N. 1994, "Mitochondrial Minicircle Dna Supports Plasmid Replication and Maintenance in Nuclei of Trypanosoma-Brucei", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 13, pp. 5962-5966.

Metzker, M. L. 2010, "Sequencing technologies [mdash] the next generation", *Nat Rev Genet*, vol. 11, no. 1, pp. 31-46.

Milman, N., Motyka, S. A., Englund, P. T., Robinson, D., & Shlomai, J. 2007, "Mitochondrial origin-binding protein UMSBP mediates DNA replication and segregation in trypanosomes", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 49, pp. 19250-19255.

Mohammad, M. M., Donti, T. R., Yakisich, J. S., Smith, A. G., & Kapler, G. M. 2007, "Tetrahymena ORC contains a ribosomal RNA fragment that participates in rDNA origin recognition", *Embo Journal*, vol. 26, no. 24, pp. 5048-5060.

Molloy, S. 2010, "PARASITOLOGY Epigenetic regulation of VSG expression", *Nature Reviews Microbiology*, vol. 8, no. 3.

Moon, K. Y., Kong, D., Lee, J. K., Raychaudhuri, S., & Hurwitz, J. 1999, "Identification and reconstitution of the origin recognition complex from Schizosaccharomyces pombe", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 22, pp. 12367-12372.

Morris, J. C., Drew, M. E., Klingbeil, M. M., Motyka, S. A., Saxowsky, T. T., Wang, Z. F., & Englund, P. T. 2001, "Replication of kinetoplast DNA: an

update for the new millennium", *International Journal for Parasitology*, vol. 31, no. 5-6, pp. 453-458.

Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., Best, A. A., Cande, W. Z., Chen, F., Cipriano, M. J., Davids, B. J., Dawson, S. C., Elmendorf, H. G., Hehl, A. B., Holder, M. E., Huse, S. M., Kim, U. U., Lasek-Nesselquist, E., Manning, G., Nigam, A., Nixon, J. E. J., Palm, D., Passamaneck, N. E., Prabhu, A., Reich, C. I., Reiner, D. S., Samuelson, J., Svard, S. G., & Sogin, M. L. 2007, "Genomic minimalism in the early diverging intestinal parasite Giardia lamblia", *Science*, vol. 317, no. 5846, pp. 1921-1926.

Morrison, L. J., Marcello, L., & McCulloch, R. 2009, "Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity", *Cellular Microbiology*, vol. 11, no. 12, pp. 1724-1734.

Mott, M. L. & Berger, J. M. 2007, "DNA replication initiation: mechanisms and regulation in bacteria", *Nat Rev Micro*, vol. 5, no. 5, pp. 343-354.

Motyka, S. A. & Englund, P. T. 2004, "RNA interference for analysis of gene function in trypanosomatids", *Current Opinion in Microbiology*, vol. 7, no. 4, pp. 362-368.

Musahl, C., Schulte, D., Burkhart, R., & Knippers, R. 1995, "A Human Homolog of the Yeast Replication Protein Cdc21 - Interactions with Other Mcm-Proteins", *European Journal of Biochemistry*, vol. 230, no. 3, pp. 1096-1101.

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. L., Ivens, A., & Stuart, K. 1999, "Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-coding genes", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 6, pp. 2902-2906.

Navarro, M. & Gull, K. 2001, "A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*", *Nature*, vol. 414, no. 6865, pp. 759-763.

Neuwald, A. F., Aravind, L., Spouge, J. L., & Koonin, E. V. 1999, "AAA(+): A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes", *Genome Research*, vol. 9, no. 1, pp. 27-43.

Newlon, C. S. 1996, "DNA replication in yeast," in *DNA Replication in Eukaryotic Cells*, M. L. DePamphilis, ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 873-914.

Ngo, H., Tschudi, C., Gull, K., & Ullu, E. 1998, "Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 25, pp. 14687-14692.

Nievera, C., Torgue, J. J. C., Grimwade, J. E., & Leonard, A. C. 2006, "SeqA blocking of DnaA-oriC interactions ensures staged assembly of the E-coli Pre-RC", *Molecular Cell*, vol. 24, no. 4, pp. 581-592.

Nishitani, H., Lygerou, Z., Nishimoto, T., & Nurse, P. 2000, "The Cdt1 protein is required to license DNA for replication in fission yeast", *Nature*, vol. 404, no. 6778, pp. 625-628.

Norseen, J., Thomae, A., Sridharan, V., Aiyar, A., Schepers, A., & Lieberman, P. M. 2008, "RNA-dependent recruitment of the origin recognition complex", *Embo Journal*, vol. 27, no. 22, pp. 3024-3035.

Nowakowski, R. S., Lewin, S. B., & Miller, M. W. 1989, "Bromodeoxyuridine Immunohistochemical Determination of the Lengths of the Cell-Cycle and the Dna-Synthetic Phase for An Anatomically Defined Population", *Journal of Neurocytology*, vol. 18, no. 3, pp. 311-318.

Ntambi, J. M., Shapiro, T. A., Ryan, K. A., & Englund, P. T. 1986, "Ribonucleotides Associated with A Gap in Newly Replicated Kinetoplast Dna Minicircles from Trypanosoma-Equiperdum", *Journal of Biological Chemistry*, vol. 261, no. 25, pp. 1890-1895.

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 Biology*, vol. 8, no. 3.

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 Research*, vol. 15, no. 1, pp. 36-43.

Ochatt, C. M., Butikofer, P., Navarro, M., Wirtz, E., Boschung, M., Armah, D., & Cross, G. A. M. 1999, "Conditional expression of glycosylphosphatidylinositol phospholipase C in *Trypanosoma brucei*", *Molecular and Biochemical Parasitology*, vol. 103, no. 1, pp. 35-48.

Ogawa, T., Yamada, Y., Kuroda, T., Kishi, T., & Moriya, S. 2002, "The datA locus predominantly contributes to the initiator titration mechanism in the control of replication initiation in Escherichia coli", *Molecular Microbiology*, vol. 44, no. 5, pp. 1367-1375.

Ogawa, Y., Takahashi, T., & Masukata, H. 1999, "Association of fission yeast Orp1 and Mcm6 proteins with chromosomal replication origins", *Molecular and Cellular Biology*, vol. 19, no. 10, pp. 7228-7236.

Okuno, Y., Satoh, H., Sekiguchi, M., & Masukata, H. 1999, "Clustered adenine/thymine stretches are essential for function of a fission yeast replication origin", *Molecular and Cellular Biology*, vol. 19, no. 10, pp. 6699-6709.

Pak, D. T. S., 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*, vol. 91, no. 3, pp. 311-323.

Papadopoulou, B., Roy, G., & Ouellette, M. 1994, "Autonomous Replication of Bacterial-Dna Plasmid Oligomers in Leishmania", *Molecular and Biochemical Parasitology*, vol. 65, no. 1, pp. 39-49.

Pasero, P., Duncker, B. P., Schwob, E., & Gasser, S. M. 1999, "A role for the Cdc7 kinase regulatory subunit Dbf4p in the formation of initiation-competent origins of replication", *Genes & Development*, vol. 13, no. 16, pp. 2159-2176.

Patnaik, P. K. 1997, "Studies with artificial extrachromosomal elements in trypanosomatids: Could specificity in the initiation of DNA replication be linked to that in transcription?", *Parasitology Today*, vol. 13, no. 12, pp. 468-471.

Patnaik, P. K., Axelrod, N., Vanderploeg, L. H. T., & Cross, G. A. M. 1996, "Artificial linear mini-chromosomes for *Trypanosoma brucei*", *Nucleic Acids Research*, vol. 24, no. 4, pp. 668-675.

Patnaik, P. K., Fang, X. D., & Cross, G. A. M. 1994, "The Region Encompassing the Procyclic Acidic Repetitive Protein (Parp) Gene Promoter Plays A Role in Plasmid Dna-Replication in Trypanosoma-Brucei", *Nucleic Acids Research*, vol. 22, no. 20, pp. 4111-4118.

Patnaik, P. K., Kulkarni, S. K., & Cross, G. A. M. 1993, "Autonomously Replicating Single-Copy Episomes in Trypanosoma-Brucei Show Unusual Stability", *Embo Journal*, vol. 12, no. 6, pp. 2529-2538.

Pays, E., Vanhamme, L., & Perez-Morga, D. 2004, "Antigenic variation in *Trypanosoma brucei*: facts, challenges and mysteries", *Current Opinion in Microbiology*, vol. 7, no. 4, pp. 369-374.

Pham, V. P., Rothman, P. B., & Gottesdiener, K. M. 1997, "Binding of trans-acting factors to the double-stranded variant surface glycoprotein (VSG) expression site promoter of *Trypanosoma brucei*", *Molecular and Biochemical Parasitology*, vol. 89, no. 1, pp. 11-23.

Piatti, S., Lengauer, C., & Nasmyth, K. 1995, "Cdc6 Is An Unstable Protein Whose De-Novo Synthesis in G(1) Is Important for the Onset of S-Phase and for Preventing A Reductional Anaphase in the Budding Yeast Saccharomyces-Cerevisiae", *Embo Journal*, vol. 14, no. 15, pp. 3788-3799.

Pines, J. 1995, "Cyclins and Cyclin-Dependent Kinases - A Biochemical View", *Biochemical Journal*, vol. 308, pp. 697-711.

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", *Journal of Cell Science*, vol. 112, no. 24, pp. 4641-4650.

Polaczek, P., Kwan, K., Liberles, D. A., & Campbell, J. L. 1997, "Role of architectural elements in combinatorial regulation of initiation of DNA replication in Escherichia coli", *Molecular Microbiology*, vol. 26, no. 2, pp. 261-275.

Pollack, J. R. & Iyer, V. R. 2002, "Characterizing the physical genome", *Nature Genetics*, vol. 32, pp. 515-521.

Prasanth, S. G., Mendez, J., Prasanth, K. V., & Stillman, B. 2004, "Dynamics of pre-replication complex proteins during the cell division cycle", *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, vol. 359, no. 1441, pp. 7-16.

Prasanth, S. G., Prasanth, K. V., & Stillman, B. 2002, "Orc6 involved in DNA replication, chromosome segregation, and cytokinesis", *Science*, vol. 297, no. 5583, pp. 1026-1031.

Prokhorova, T. A. & Blow, J. J. 2000, "Sequential MCM/P1 subcomplex assembly is required to form a heterohexamer with replication licensing activity", *Journal of Biological Chemistry*, vol. 275, no. 4, pp. 2491-2498.

Ramachandran, L., Burhans, D. T., Laun, P., Wang, J. X., Liang, P., Weinberger, M., Wissing, S., Jarolim, S., Suter, B., Madeo, F., Breitenbach, M., & Burhans, W. C. 2006, "Evidence for ORC-dependent repression of budding yeast genes induced by starvation and other stresses", *Fems Yeast Research*, vol. 6, no. 5, pp. 763-776.

Randell, J. C. W., Bowers, J. L., Rodriguez, H. K., & Bell, S. P. 2006, "Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase", *Molecular Cell*, vol. 21, no. 1, pp. 29-39.

Rao, H. & Stillman, B. 1995, "The Origin Recognition Complex Interacts with A Bipartite Dna-Binding Site Within Yeast Replicators", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 6, pp. 2224-2228.

Ray, D. S. 1989, "Conserved Sequence Blocks in Kinetoplast Minicircles from Diverse Species of Trypanosomes", *Molecular and Cellular Biology*, vol. 9, no. 3, pp. 1365-1367.

Redmond, S., Vadivelu, J., & Field, M. C. 2003, "RNAit: an automated web-based tool for the selection of RNAi targets in *Trypanosoma brucei*", *Molecular and Biochemical Parasitology*, vol. 128, no. 1, pp. 115-118.

Remus, D., Beall, E. L., & Botchan, M. R. 2004, "DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding", *Embo Journal*, vol. 23, no. 4, pp. 897-907.

Renauld, H., Kelly, J. M., & Horn, D. 2007, "Chromosome Structure and Dynamics," in , J. D. Barry et al., eds., Horizon Bioscience, Wymondham, UK, pp. 91-131.

Robinson, N. P. & Bell, S. D. 2007, "Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 14, pp. 5806-5811.

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*, vol. 116, no. 1, pp. 25-38.

Robinson, N. P. & Bell, S. D. 2005, "Origins of DNA replication in the three domains of life", *FEBS Journal*, vol. 272, no. 15, pp. 3757-3766.

Roditi, I., Furger, A., Ruepp, S., Schurch, N., & Butikofer, P. 1998, "Unravelling the procyclin coat of *Trypanosoma brucei*", *Molecular and Biochemical Parasitology*, vol. 91, no. 1, pp. 117-130.

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", *Current Biology*, vol. 6, no. 11, pp. 1416-1425.

Roth, A. & Messer, W. 1995, "The DNA-Binding Domain of the Initiator Protein DnaA", *Embo Journal*, vol. 14, no. 9, pp. 2106-2111.

Rudenko, G., Blundell, P. A., Dirksmulder, A., Kieft, R., & Borst, P. 1995, "A Ribosomal Dna Promoter Replacing the Promoter of A Telomeric Vsg Gene-Expression Site Can be Efficiently Switched on and Off in Trypanosoma-Brucei", *Cell*, vol. 83, no. 4, pp. 547-553.

Rudenko, G., Blundell, P. A., Taylor, M. C., Kieft, R., & Borst, P. 1994, "Vsg Gene-Expression Site Control in Insect Form Trypanosoma-Brucei", *Embo Journal*, vol. 13, no. 22, pp. 5470-5482.

Rudenko, G. 2010, "Epigenetics and transcriptional control in African trypanosomes", *Essays in Biochemistry*, vol. 48, no. 1, pp. 201-219.

Saha, S., Shan, Y. J., 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 & Development*, vol. 18, no. 4, pp. 397-410.

Santocanale, C. & Diffley, J. F. X. 1997, "Genomic footprinting of budding yeast replication origins during the cell cycle", *Cell Cycle Control*, vol. 283, pp. 377-390.

Sasaki, T. & Gilbert, D. M. 2007, "The many faces of the origin recognition complex", *Current Opinion in Cell Biology*, vol. 19, no. 3, pp. 337-343.

Sasaki, T., Sawado, T., Yamaguchi, M., & Shinomiya, T. 1999, "Specification of regions of DNA replication initiation during embryogenesis in the 65-kilobase DNApol alpha-dE2F locus of Drosophila melanogaster", *Molecular and Cellular Biology*, vol. 19, no. 1, pp. 547-555.

Sato, M., Gotow, T., You, Z. Y., Komamura-Kohno, Y., Uchiyama, Y., Yabuta, N., Nojima, H., & Ishimi, Y. 2000, "Electron microscopic observation and single-stranded DNA binding activity of the Mcm4,6,7 complex", *Journal of Molecular Biology*, vol. 300, no. 3, pp. 421-431.

Schaper, S. & Messer, W. 1995, "Interaction of the Initiator Protein Dnaa of Escherichia-Coli with Its Dna Target", *Journal of Biological Chemistry*, vol. 270, no. 29, pp. 17622-17626.

Schimanski, B., Nguyen, T. N., & Gunzl, A. 2005, "Highly efficient tandem affinity purification of trypanosome protein complexes based on a novel epitope combination", *Eukaryotic Cell*, vol. 4, no. 11, pp. 1942-1950.

Schubert, V. 2009, "SMC Proteins and Their Multiple Functions in Higher Plants", *Cytogenetic and Genome Research*, vol. 124, no. 3-4, pp. 202-214.

Schulte, D., Richter, A., Burkhart, R., Musahl, C., & Knippers, R. 1996, "Properties of the human nuclear protein p85Mcm - Expression, nuclear localization and interaction with other Mcm proteins", *European Journal of Biochemistry*, vol. 235, no. 1-2, pp. 144-151.

Schwacha, A. & Bell, S. P. 2001, "Interactions between two catalytically distinct MCM subgroups are essential for coordinated ATP hydrolysis and DNA replication", *Molecular Cell*, vol. 8, no. 5, pp. 1093-1104.

Sclafani, R. A. & Holzen, T. M. 2007, "Cell cycle regulation of DNA replication", *Annual Review of Genetics*, vol. 41, pp. 237-280.

Segurado, M., de Luis, A., & Antequera, F. 2003, "Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe", *Embo Reports*, vol. 4, no. 11, pp. 1048-1053.

Sekimizu, K., Bramhill, D., & Kornberg, A. 1987, "ATP Activates DnaA Protein in Initiating Replication of Plasmids Bearing the Origin of the Escherichia-Coli Chromosome", *Cell*, vol. 50, no. 2, pp. 259-265.

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 Journal*, vol. 25, no. 21, pp. 5150-5158.

Senga, T., Sivaprasad, U., Zhu, W., Park, J. H., Arias, E. E., Walter, J. C., & Dutta, A. 2006, "PCNA is a co-factor for Cdt1 degradation by CUL4/DDB1 mediated N-terminal ubiquitination", *Journal of Biological Chemistry* p. M512705200.

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 Genetics*, vol. 5, no. 4.

Shapiro, T. A. & Englund, P. T. 1995, "The Structure and Replication of Kinetoplast DNA", *Annual Review of Microbiology*, vol. 49, pp. 117-143.

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", *Molecular and Biochemical Parasitology*, vol. 113, no. 1, pp. 171-173.

Sheu, Y. J. & Stillman, B. 2010, "The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4", *Nature*, vol. 463, no. 7277, p. 113-U127.

Shin, J. H., Grabowski, B., Kasiviswanathan, R., Bell, S. D., & Kelman, Z. 2003, "Regulation of Minichromosome Maintenance Helicase Activity by Cdc6", *Journal of Biological Chemistry*, vol. 278, no. 39, pp. 38059-38067.

Shlomai, J. 2004, "The structure and replication of kinetoplast DNA", *Current Molecular Medicine*, vol. 4, no. 6, pp. 623-647.

Sibley, L. D. 2004, "Intracellular parasite invasion strategies", *Science*, vol. 304, no. 5668, pp. 248-253.

Siegel, T. N., Hekstra, D. R., Kemp, L. E., Figueiredo, L. M., Lowell, J. E., Fenyo, D., Wang, X. N., Dewell, S., & Cross, G. A. M. 2009, "Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*", *Genes & Development*, vol. 23, no. 9, pp. 1063-1076.

Siegel, T. N., Hekstra, D. R., Wang, X., Dewell, S., & Cross, G. A. M. 2010, "Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites", *Nucleic Acids Research* p. gkq237.

Simpson, A. G. B. & Roger, A. J. 2004, "The real 'kingdoms' of eukaryotes", *Current Biology*, vol. 14, no. 17, p. R693-R696.

Simpson, L. & Dasilva, A. 1971, "Isolation and Characterization of Kinetoplast Dna from Leishmania-Tarentolae", *Journal of Molecular Biology*, vol. 56, no. 3, p. 443-&.

Simpson, L., Thiemann, O. H., Savill, N. J., Alfonzo, J. D., & Maslov, D. A. 2000, "Evolution of RNA editing in trypanosome mitochondria", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 13, pp. 6986-6993.

Singleton, M. R., Morales, R., Grainge, I., Cook, N., Isupov, M. N., & Wigley, D. B. 2004, "Conformational changes induced by nucleotide binding in Cdc6/ORC from Aeropyrum pernix", *Journal of Molecular Biology*, vol. 343, no. 3, pp. 547-557.

Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A., & Peattie, D. A. 1989, "Phylogenetic Meaning of the Kingdom Concept - An Unusual Ribosomal-Rna from Giardia-Lamblia", *Science*, vol. 243, no. 4887, pp. 75-77.

Son, H. J., Cook, G. A., Hall, T., & Donelson, J. E. 1989, "Expression Site Associated Genes of Trypanosoma-Brucei-Rhodesiense", *Molecular and Biochemical Parasitology*, vol. 33, no. 1, pp. 59-66.

Speck, C., Chen, Z. Q., Li, H. L., & Stillman, B. 2005, "ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA", *Nature Structural & Molecular Biology*, vol. 12, no. 11, pp. 965-971.

Speck, C. & Messer, W. 2001, "Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA", *Embo Journal*, vol. 20, no. 6, pp. 1469-1476.

Speck, C. & Stillman, B. 2007, "Cdc6 ATPase activity regulates ORC center x Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication", *Journal of Biological Chemistry*, vol. 282, no. 16, pp. 11705-11714.

Speck, C., Weigel, C., & Messer, W. 1999, "ATP- and ADP-DnaA protein, a molecular switch in gene regulation", *Embo Journal*, vol. 18, no. 21, pp. 6169-6176.

Stanojcic, S., Lemaitre, J. M., Brodolin, K., Danis, E., & Mechali, M. 2008, "In Xenopus egg extracts, DNA replication initiates preferentially at or near asymmetric AT sequences", *Molecular and Cellular Biology*, vol. 28, no. 17, pp. 5265-5274.

Stockdale, C., Swiderski, M. R., Barry, J. D., & McCulloch, R. 2008, "Antigenic variation in *Trypanosoma brucei*: Joining the DOTs", *Plos Biology*, vol. 6, no. 7, pp. 1386-1391.

Sugimoto, N., Tatsumi, Y., Tsurumi, T., Matsukage, A., Kiyono, T., Nishitani, H., & Fujita, M. 2004, "Cdt1 Phosphorylation by Cyclin A-dependent Kinases Negatively Regulates Its Function without Affecting Geminin Binding", *Journal of Biological Chemistry*, vol. 279, no. 19, pp. 19691-19697.

Sun, J. Y. & Kong, D. C. 2010, "DNA replication origins, ORC/DNA interaction, and assembly of pre-replication complex in eukaryotes", *Acta Biochimica et Biophysica Sinica*, vol. 42, no. 7, pp. 433-439.

Tada, S. 2007, "Cdt1 and geminin: role during cell cycle progression and DNA damage in higher eukaryotes", *Frontiers in Bioscience*, vol. 12, pp. 1629-1641.

Tada, S., Li, A., Maiorano, D., Mechali, M., & Blow, J. J. 2001, "Repression of origin assembly in metaphase depends on inhibition of RLF-B/Cdt1 by geminin", *Nature Cell Biology*, vol. 3, no. 2, pp. 107-113.

Takahashi, N., Yamaguchi, Y., Yamairi, F., Makise, M., Takenaka, H., Tsuchiya, T., & Mizushima, T. 2004, "Analysis on origin recognition complex containing Orc5p with defective Walker A motif", *Journal of Biological Chemistry*, vol. 279, no. 9, pp. 8469-8477.

Tamura, K., Sakazaki, R., Kuramochi, S., & Nakamura, A. 1988, "Occurrence of H-Antigen-Z66 of R-Phase in Cultures of Salmonella Serovar Typhi Originated from Indonesia", *Epidemiology and Infection*, vol. 101, no. 2, pp. 311-314.

Tanaka, S. & Diffley, J. F. X. 2002, "Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase", *Nature Cell Biology*, vol. 4, no. 3, pp. 198-207.

Taylor, J. E. & Rudenko, G. 2006, "Switching trypanosome coats: what's in the wardrobe?", *Trends in Genetics*, vol. 22, no. 11, pp. 614-620.

Tetley, L., Turner, C. M. R., 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", *Journal of Cell Science*, vol. 87, pp. 363-372.

Tetley, L. & Vickerman, K. 1985, "Differentiation in Trypanosoma-Brucei - Host-Parasite Cell-Junctions and Their Persistence During Acquisition of the Variable Antigen Coat", *Journal of Cell Science*, vol. 74, no. MAR, pp. 1-19.

Theis, J. F. & Newlon, C. S. 1997, "The ARS309 chromosomal replicator of Saccharomyces cerevisiae depends on an exceptional ARS consensus sequence", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 20, pp. 10786-10791.

Triolo, T. & Sternglanz, R. 1996, "Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing", *Nature*, vol. 381, no. 6579, pp. 251-253.

Tsakraklides, V. & Bell, S. P. 2010, "Dynamics of Pre-replicative Complex Assembly", *Journal of Biological Chemistry*, vol. 285, no. 13, pp. 9437-9443.

Tsuyama, T., Tada, S., Watanabe, S., Seki, M., & Enomoto, T. 2005, "Licensing for DNA replication requires a strict sequential assembly of Cdc6 and Cdt1 onto chromatin in Xenopus egg extracts", *Nucleic Acids Research*, vol. 33, no. 2, pp. 765-775.

Tu, X. M. & 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", *Journal of Biological Chemistry*, vol. 279, no. 19, pp. 20519-20528.

Tye, B. K. 1999, "MCM proteins in DNA replication", *Annual Review of Biochemistry*, vol. 68, pp. 649-686.

Ueda, J., Saito, H., Watanabe, H., & Evers, B. M. 2005, "Novel and quantitative DNA dot-blotting method for assessment of in vivo proliferation", *American Journal of Physiology-Gastrointestinal and Liver Physiology*, vol. 288, no. 4, p. G842-G847.

Ullu, E., Djikeng, A., Shi, H. F., & Tschudi, C. 2002, "RNA interference: advances and questions", *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, vol. 357, no. 1417, pp. 65-70.

Ullu, E., Tschudi, C., & Chakraborty, T. 2004, "RNA interference in protozoan parasites", *Cellular Microbiology*, vol. 6, no. 6, pp. 509-519.

Waltz, S. E., Trivedi, A. A., & Leffak, M. 1996, "DNA replication initiates non-randomly at multiple sites near the c-myc gene in HeLa cells", *Nucleic Acids Research*, vol. 24, no. 10, pp. 1887-1894.

Wang, Z. F., 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", *Journal of Biological Chemistry*, vol. 275, no. 51, pp. 40174-40179.

Weiden, M., Osheim, Y. N., Beyer, A. L., & Van der Ploeg, L. H. 1991, "Chromosome structure: DNA nucleotide sequence elements of a subset of the minichromosomes of the protozoan *Trypanosoma brucei*", *Molecular and Cellular Biology*, vol. 11, no. 8, pp. 3823-3834.

Weigel, C., Schmidt, A., Seitz, H., Tungler, D., Welzeck, M., & Messer, W. 1999, "The N-terminus promotes oligomerization of the Escherichia coli initiator protein DnaA", *Molecular Microbiology*, vol. 34, no. 1, pp. 53-66.

Weinreich, M., Debeer, M. A. P., & Fox, C. A. 2004, "The activities of eukaryotic replication origins in chromatin", *Biochimica et Biophysica Acta-Gene Structure and Expression*, vol. 1677, no. 1-3, pp. 142-157.

Weinreich, M., Liang, C., & Stillman, B. 1999, "The Cdc6p nucleotide-binding motif is required for loading Mcm proteins onto chromatin", *Proceedings of the National Academy of Sciences*, vol. 96, no. 2, pp. 441-446.

Whittaker, A. J., Royzman, I., & Orr-Weaver, T. L. 2000, "Drosophila Double parked: a conserved, essential replication protein that colocalizes with the origin recognition complex and links DNA replication with mitosis and the down-regulation of S phase transcripts", *Genes & Development*, vol. 14, no. 14, pp. 1765-1776.

Wickstead, B., Ersfeld, K., & Gull, K. 2004, "The small chromosomes of *Trypanosoma brucei* involved in antigenic variation are constructed around repetitive palindromes", *Genome Research*, vol. 14, no. 6, pp. 1014-1024.

Wirtz, E. & Clayton, C. 1995, "Inducible Gene-Expression in Trypanosomes Mediated by A Prokaryotic Repressor", *Science*, vol. 268, no. 5214, pp. 1179-1183.

Wirtz, E., Leal, S., Ochatt, C., & Cross, G. A. M. 1999, "A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*", *Molecular and Biochemical Parasitology*, vol. 99, no. 1, pp. 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*, vol. 290, no. 5500, p. 2309-+.

Wold, S., Crooke, E., & Skarstad, K. 1996, "The Escherichia coli Fis protein prevents initiation of DNA replication from oriC in vitro", *Nucleic Acids Research*, vol. 24, no. 18, pp. 3527-3532.

Woodward, R. & Gull, K. 1990, "Timing of Nuclear and Kinetoplast Dna-Replication and Early Morphological Events in the Cell-Cycle of Trypanosoma-Brucei", *Journal of Cell Science*, vol. 95, pp. 49-57.

Worthen, C., Jensen, B. C., & Parsons, M. 2010, "Diverse Effects on Mitochondrial and Nuclear Functions Elicited by Drugs and Genetic Knockdowns in Bloodstream Stage *Trypanosoma brucei*", *Plos Neglected Tropical Diseases*, vol. 4, no. 5.

Wright, J. R., Siegel, T. N., & Cross, G. A. M. 2010, "Histone H3 trimethylated at lysine 4 is enriched at probable transcription start sites in *Trypanosoma brucei*", *Molecular and Biochemical Parasitology*, vol. 172, no. 2, pp. 141-144.

Wyrick, J. J., Aparicio, J. G., Chen, T., Barnett, J. D., Jennings, E. G., Young, R. A., Bell, S. P., & Aparicio, O. M. 2001, "Genome-wide distribution of ORC and MCM proteins in S-cerevisiae: High-resolution mapping of replication origins", *Science*, vol. 294, no. 5550, pp. 2357-2360.

Xu, W. H., Aparicio, J. G., Aparicio, O. M., & Tavare, S. 2006, "Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S-cerevisiae", *Bmc Genomics*, vol. 7.

Yang, X. F., Figueiredo, L. M., Espinal, A., Okubo, E., & Li, B. 2009, "RAP1 Is Essential for Silencing Telomeric Variant Surface Glycoprotein Genes in *Trypanosoma brucei*", *Cell*, vol. 137, no. 1, pp. 99-109.

Yasuda, S. & Hirota, Y. 1977, "Cloning and Mapping of Replication Origin of Escherichia-Coli", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 12, pp. 5458-5462.

Yoshikawa, H. & Ogasawara, N. 1991, "Structure and Function of Dnaa and the Dnaa-Box in Eubacteria - Evolutionary Relationships of Bacterial Replication Origins", *Molecular Microbiology*, vol. 5, no. 11, pp. 2589-2597.

Yu, Z. L., Feng, D. R., & Liang, C. 2004, "Pairwise interactions of the six human MCM protein subunits", *Journal of Molecular Biology*, vol. 340, no. 5, pp. 1197-1206.

Zappulla, D. C., Sternglanz, R., & Leatherwood, J. 2002, "Control of replication timing by a transcriptional silencer", *Current Biology*, vol. 12, no. 11, pp. 869-875.

Zhao, Z. S., Granucci, F., Yeh, L., Schaffer, P. A., & Cantor, H. 1998, "Molecular mimicry by herpes simplex virus type 1: autoimmune disease after viral infection", *Science*, vol. 279, no. 5355, pp. 1344-1347.

Zyskind, J. W., Cleary, J. M., Brusilow, W. S. A., Harding, N. E., & Smith, D. W. 1983, "Chromosomal Replication Origin from the Marine Bacterium Vibrio-Harveyi Functions in *Escherichia-Coli - Ori*C Consensus Sequence", *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, vol. 80, no. 5, pp. 1164-1168.